Doctoral Dissertation

Urban landscape analysis methodology based on street view images using deep learning

（深層学習を用いたストリートビュー画像に基づく都市景観分析方法論）

Xia Yixi

January 2022

Graduate School of Engineering,

Osaka University

Abstract

Street View image has interested urban planners and researchers for decades and has gradually ascended as a data source for geospatial analysis and urban analysis, which deriving urban environment improvement and urban design optimization. The data we can obtain from the street view images not only includes the visual information of the urban physical environment but also reflects information about urban functions, human activities, and urban climate. However, the traditional digital image processing methods are limited, and cannot extract effective information accurately and efficiently from street view images. Despite extensive research into environmental assessment factors that reflect urban environment quality such as urban greenery, street openness, plant shades, it is still mostly done manually based on big data. In recent years, with the continuous improvement of artificial intelligence technology, breakthroughs have been made in the use of machine learning and deep learning to extract semantic information from images. Image semantic segmentation and instance segmentation technologies based on deep learning provide strong support for extracting key information from street view images and analyzing the quality of urban street environments. In this process, a large number of new methods and new perspectives have emerged, providing new research ideas for urban environment research, spatial data mining, and human activity analysis based on big data.

This research proposes a three-step work to develop a method for using street view images to evaluate the greenery of urban street-level and the openness of the built environment by improving the accuracy of image semantic segmentation.

The first step is to compare the deep learning models of each image semantic segmentation to find an algorithm suitable for the semantic segmentation of urban street view images. Next is developing a method based on semantic segmentation processing of street view images to calculate the Green View Index of urban streets. For this phase, the Panoramic View Green View Index (PVGVI) is proposed for measuring the visible street-level greenery. Then, this method extends to automatically extracting Sky View Factor from street view images to measure the openness of the street-built environment. Finally, the Green View Index and the Sky View Factor which separately represents the greenery and openness of the outdoor urban environment have been visualized on the street map, which can more intuitively show the characteristics of the urban environment, and provide strong support for the decision-making of urban planners and managers.

The outcome of this research is beneficial to both urban planners and urban managers. It helps to combine large-scale street view image datasets with image recognition technology and helps city planners to obtain urban environmental data for evaluation. With the support of artificial

intelligence technology and the help of street view images, it is possible to further study the spatial characteristics, laws, and evolution process of the city. The mentioned outcome helps researchers to propose better urban upgrading strategies to create a better urban environment.

Preface

This dissertation is the original work by Xia Yixi under the supervision of Professor Nobuyoshi Yabuki. Two journal articles and one international conference paper that relate to this dissertation have been published and are listed below.

Journal articles:

1. 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.
2. Xia, Y., Yabuki, N., & Fukuda, T. (2021). Sky view factor estimation from street view images based on semantic segmentation. *Urban Climate*, *40*, 100999.

International conference proceedings:

1. Xia, Y., Yabuki, N., & Fukuda, T. (2020). Development of an urban greenery evaluation system based on deep learning and Google Street View, in: 25th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2020, pp. 783–792.

Acknowledgment

This doctoral dissertation would not have been possible without the guidance, support, and assistance of the following people.

First of all, I would like to extend my sincere gratitude to my Ph.D. supervisor, Professor Nobuyoshi Yabuki, for his constant encouragement and guidance. Without his consistent and illuminating instruction, this research could not have reached its present form. Prof. Yabuki continuously provided support and was always willing and enthusiastic to assist in any way he could throughout the research project. I would also like to thank him for giving me the freedom to carry out various research projects and providing me the opportunity to participate in the international conference, which was my most valuable experience during my Ph.D. study. I would also like to express my heartfelt gratitude to my co-supervisor, Associate Professor Tomohiro Fukuda, for his patient instruction, insightful criticism, and expert guidance of my research. Dr. Fukuda can always point out valuable comments for all of my papers, which made me gain so much.

Next, my thanks would do to my beloved parents, Zhang Meijian and Xia Zhigao, and my family for their loving considerations and great confidence in me all through these years. I also owe my sincere gratitude to my best friend, Liu Yumeng and Wan Xuan who gave me their time and help in listening to me and assisting me to work out my confuses during the difficult course of this research.

I would also like to thank all of my friends in Yabuki's Lab, for example, Mr. Zhu, Ms. Manuel, Mr. Chavanont, Mr. Natthapol, Mr. Kido for their impressive kindness and their help during my stay in Japan. Thank are also due to my Chinese friends at Osaka University, whose help is indispensable to me. Without them, this period of living here would not be so unforgettable.

I also owe a special debt of gratitude to the dissertation committees for spending their time reading this dissertation and providing me with valuable feedback and recommendations. I also would like to thank Professor Masanori Sawaki for being the vice referee for my Ph.D. examination.

Last but not least, I shall extend my thanks to Ms. Zeng, Mr. Wang, and Ms. Gao, without their support, I would not have had the opportunity to study abroad and become a Ph.D. graduate. I

would also like to thank all of the colleagues in the Chengdu Institute of Planning and Design for all the warm and kind support they have given me for all these years.

I will firmly remember every day of these three years as the most precious meeting in my life.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Background and problem statements

The city is essentially a huge and complex system (Batty, 2008; Isalgue et al., 2007). For decades, city planners and managers have been working to find a way to accurately and effectively evaluate the urban environment. They hope that the results of these evaluations can be fed back into the urban design and management process, and can be further applied in the strategies to improve the quality of the urban environment. Most of the human activities occur in the urban street environment (Madanipour, 1996). They are places where human activities intensively occur in cities. Urban street space is one of the most important elements that constitute the urban environment, and it is also the main interface between humans and the city. High-quality street space not only helps to enhance the vitality of the city but also increases the frequency of social interaction and outdoor activities (Handy et al., 2002). The perception of spatial quality based on urban streets is regarded as an important public product. Many cities have also successively proposed many street designing measures and urban renewal policies, such as the *Urban Street Design Guide* (National Association of City Transportation Officials, 2012). As shown in these guidelines, the focus has gradually changed from "transport-centric" to "people-centric". At the same time, some evaluation platforms have also emerged, such as Walk Score and Bike Score, to assist in accurate quality evaluation and design interventions. This series of

actions shows that researchers are paying more and more attention to the quality of streets. From Jacobs (1961) to Montgomery (1998), these urbanists have published a lot of research about the elements which affect the quality of the urban street environment. However, they generally lack the discussion of these urban environment elements and the perceptual attributes related to the urban visual perceptions. In recent years, some quantitative analysis on urban spatial quality has been gradually introduced into the urban study field. They reveal some relationships between visual quality, spatial perception, and corresponding urban design elements. However, these studies are usually based on small-scale urban data sets collected manually, which cannot support their expansion to other areas of the city.

Street view imagery is a new type of big geographic dataset that perceives the physical environment of the city. This kind of high-density image data covering the urban street network depicts the urban environment in detail from a human perspective, thereby effectively supporting the quantitative research of the urban material environment. As a big data source, street view images not only include the visible environment of the urban material space, such as buildings, roads, plants, etc., but also cover some information hidden under the material space, including urban functions, human activities, and society economic information. With the development of technology, the achievements of computer vision and deep learning technology in image recognition (LeCun et al., 2015) have provided a new research path for understanding cities through images (Reichstein et al., 2019). Continuously updated computer vision algorithms enable researchers to solve and predict urban problems in a more precise and efficient manner.

Urban street-level greenery has long been recognized as one of the most prime landscape design elements in the urban ecological system (Wolf, 2005). It provides multiple benefits to urban environments, such as urban trees that can modify environment temperatures by providing shade and cooling, helping to significantly reduce the risk of heat-related illnesses (Mavrogianni et al., 2014). The street-side green spaces bring huge benefits to cities (Bain et al., 2012), it provides opportunities for community residents to engage in physical exercise, thereby reducing obesity and mental stress (Giles-Corti et al., 2003). It also provides more opportunities to be in green spaces, which is conducive to improved mental health, especially in reducing the risk of attention deficit disorder in childhood (Louv, 2008). Urban street-level greenery also makes an important contribution to the attractiveness and walkability of residential streets (Schroeder and Cannon, 1983; Bain et al., 2012). On the other hand, it provides a welcoming environment for people who have a certain impact on the occurrence of

various mental illnesses (Coutts, 2008; Lee and Maheswaran, 2011; Leslie et al., 2010). Therefore, ensuring widespread access to street-level green space is a key factor in providing these environments and benefits to residents (Landry and Chakraborty, 2009).

The openness of urban streets is another one of the main indicators for evaluating environmental spatial quality. In Jan Gehl's theory of public space design, human activities are regarded as the starting point and end of all design. In addition to the necessity and importance of the activity itself, people's feelings are also one of the important factors that affect the activity. And a good outdoor environment can prolong people's staying time outdoors, thereby inducing more public activities (Gehl, 1987). In most cases, open space is always more popular than closed space, and it is easier to make people feel comfortable. Sky View Factor (SVF) represents the ratio of the space point between the visible sky and the hemisphere centered on the analysis position in the urban street space, with a value between 0 and 1, where 1 indicates an open area without any obstructions and 0 indicates a completely blocked space (Brown et al., 2001; He et al., 2015; Scarano and Mancini, 2017).

In recent studies, analytical methods on urban environmental quality have mostly relied on manual extraction and analysis of urban big data. In addition, the source of analysis data is limited by the scale and technique, and it is often difficult to be applied for large-scale urban analysis. The complexity of urban environmental factors also increases the risk of perceived errors. Numerous studies have tried to facilitate this problem by proposing automated techniques, among which one of the most important methods is to use deep learning systems to automate the process. Therefore, urban landscape analysis based on street view images using deep learning would be introduced and proposed in this research.

1.2 Research objectives

The information extraction and classification of street view images through computer vision technology can reduce the huge workload of urban researchers. Due to the wide range of areas and abundant information in urban landscape analysis, the data collection process is usually challenging. Moreover, due to the different attributes of urban scenes, the developed analysis method must be applied to all urban street scene images, and the classification result can be used as a training dataset to improve the detection accuracy of the semantic segmentation model in the future. Therefore, the objectives of this research are as follows:

(1) To evaluate the performance of a deep learning model for street view image semantic segmentation.
(2) To propose a method that can automatically detect and classify the vegetation and sky area based on street view images.
(3) To develop a method that can estimate and map the street-level distribution of visible urban greenery and proposed the Panoramic View Green View Index (PVGVI) to calculate it.
(4) To develop a method that can accurately and efficiently estimate the sky areas from fisheye images and calculate the SVF.

1.3 Research significance

This research will provide an understanding of the computer vision from the urban street-level such as deep learning-based urban landscape elements detection issues. The proposed methods in this research are beneficial for both urban analysis research and practical management. For the research community, the proposed methods can improve the data as the basis to evaluate the urban environment in an all-around way, and rely on a wealth of design methods to achieve the effect of improving the urban environment quality. For professional practice, accurate urban landscape analysis elements can be achieved from street view images. The working time, cost, and error of handling various city data to obtain accurate details of the urban environment can be reduced.

1.4 Research scope

The focus of this research is using street view images to analyze the urban landscape environment based on deep learning. This research chooses the visible greenery of the urban street-level and the openness of the urban built environment as two starting points of this research. Since, in the urban street space, these are two important indicators that affect the quality of street space and people's visual perception of the built environment. At the same time, because plants are full of unique shapes and feature-rich textures, and the built environment on both sides of the street is updated at any time, it is more challenging to measure the visible greenery and sky visibility.#

1.5 Overview of the dissertation

This dissertation consists of six chapters as follows:

Chapter 1 Introduction:

This chapter presents the background information, problem statements, research objectives, research significance, research scope, and overview of all chapters in this dissertation.

Chapter 2 Literature review:

This chapter presents the concept, issues, and related works to this study. It is divided into four main parts. The first part explains the elements of urban landscape analysis. It consists of three parts, namely the natural environment, the built environment, and human interactions. The second part introduces computer vision information related to this research. The third part explains the deep learning models related to this research. The fourth part briefly reviews the applications of street view images in urban landscape analysis. The last part summarizes this chapter.

Chapter 3: Detecting the urban design elements from street view images for urban landscape analysis

In this chapter, the accuracy of the proposed semantic segmentation algorithm is examined through a comparative study. Subsequently, the Green View Index (GVI) and Sky View Factor (SVF) are proposed, which are recognized as the factors to evaluate the greenery of urban street-level and the openness of the built environment respectively. The results are discussed and summarized.

Chapter 4: Assessing the quality of urban greenery by estimating the Green View Index

This chapter proposes a method based on semantic segmentation processing of street view images to calculate the Green View Index of urban streets, and the Panoramic View Green View Index (PVGVI) is proposed for measuring the visible street-level greenery. Subsequently, a method for improving the accuracy and speed of segmentation of street view images by using a pre-trained semantic segmentation model was proposed. The proposal of PVGVI and the evaluation system is described. Validation through a case study is conducted. The results are discussed and summarized.

Chapter 5: Estimating the Sky View Factor to visualize the built environment openness

This chapter extends the proposed method from Chapter 4 to automatically detect the sky area from street view images. The proposed approach can then be used to estimate the accurate SVF of the urban built environment. The development of the detection and evaluation system is

described. Validation through a case study is conducted. The results are discussed and summarized.

Chapter 6: Conclusion

This chapter concludes the entire research by giving a summary, contributions of the research, and limitations.

Chapter 2

Literature review

This chapter presents the literature that is related to the work in this dissertation. It is separated into five sections. The first section focuses on urban landscape analysis elements, which can be divided into other three subsections: 1) built environment, 2) natural environment, 3) human perception. The second section guides computer vision, which includes two parts: 1) CNN, 2) FCN. The third section introduces the deep learning and some representative deep learning models, such as SegNet, PSPNet, DeepLab. The fourth section introduce the street view images and the mainly providers of open source street view image data. The fifth section presents the applications of street view images in the urban environment analysis. This section contains two subsections, which are using green view index to assess the quality of environment greenery and using sky view factor to evaluate the openness of the urban built environment based on street view image. The last section summarizes all the detail in this chapter.

2.1 The urban landscape analysis elements

In the mid-1940s, Saarinen (Saarinen, 1948) proposed the concept of urban design, which began to be widely accepted in the 1960s. Urban design is a discipline that pays attention to urban planning layout, urban appearance, urban functions, and especially urban public space. The complex process of urban design is to focus on the interrelationship between the elemental arrangement of the city and the social and psychological health of the residents. Through the

treatment of material space and landscape signs, a material environment is created, which can not only make residents happy but also inspire community behavior.

For example, New York vigorously promoted urban design in 1964 as a new policy to improve the urban environment. In the past ten years, many countries have begun to emphasize urban design to enhance the characteristic image of the city, improve the quality of the urban environment, and promote the coordinated development of people, the city, and the environment. Urban design is modeling design, but it is not an individual architectural model, but an orderly arrangement of various elements of the city. The so-called urban design is to establish an urban order that conforms to the lives of people in modern society. The goal of urban design is to create a comfortable, convenient, hygienic, and beautiful space environment for people. That is, through the comprehensive design of various material elements in a certain area, the city can achieve the coordination and coordination of various facilities and functions, as well as space. The form is unified and perfect, and the overall benefits are optimized. The elements of urban design include many aspects. In 1960, Kevin Lynch introduced the field of psychology into urban research. In the book "The image of the city" (Kevin, 1964), people's impressions of the city were summarized into five elements, which are path, edge, district, node, and landmark. The path is the trajectory of people's movement, which can be streets, trails, transportation lines, rivers or railways, etc. People move along the roads while observing the city, and rely on these roads to organize and connect the rest of the environmental factors. Edges are linear elements that are not regarded as roads by people: they are usually the dividing line between two areas. They play the role of side notes. The district is a medium-scale or large-scale unit in a city. The node is the identification point, an important strategic point that people can enter in the city, and the focal point of arrival and departure during the journey. Landmarks are another type of reference point. People are outside of them without entering them. They are usually objects that are simply defined: buildings, signs, shops, or mountain peaks.

These elements include street greening rate, sky visibility, continuity of building façade, pedestrian-friendly degree, degree of motorization. They are interdependent in the process of urban design. These elements together constitute an analysis system for evaluating urban environment quality. Urban scientists also try to use some methods to quantitatively analyze these elements. This research mainly focused on the urban design elements of the built environment, the natural environment, and the human interaction, which can be quantified by the visible greenery, the visible sky area percentage, etc.

2.1.1 The built environment

The built environment refers to various buildings and places that are constructed and renovated by humans, especially those environments that can be changed by urban design and human behavior, including central space, public space, green space, etc. There are many factors to measure the built environment. Cervero and Kockelman summarized the built environment into three important dimensions (3D), namely density, diversity, and design (Cervero et al., 1997); Handy et al. put forward six built environment characteristics to describe the building Environment, including density, mixed-use of land, street connectivity, block size, aesthetics, and regional structure (Handy et al., 2002); Picola and others discussed the characteristics of the built environment from four aspects: function, safety, beauty, and destination (Pikora et al., 2003).#

2.1.2 The natural environment

The natural environment is an important entry point for us to understand the city, which is composed of plants, animals, air conditions, climate conditions, etc. It not only affects our overall perception of the urban environment but also stimulates or inhibits the occurrence of interpersonal communication activities. More importantly, the microclimate environment of the city depends largely on the natural environment. Urban researchers are aware of its importance and published many aspects of research results. These studies include mapping greenery in cities, identifying the plant species (Krause et al., 2018; Sun et al., 2017), studying the interaction between the natural environment and wildlife (Mohanty et al., 2016).

2.1.3 Human perception

In the field of urban design, human behavior interacts with the surrounding environment, and human perception can directly feedback the quality of the environment and space. In different places and material spaces, people's sense of place, and activity status are different. The types, styles, and colors of buildings, the shape, vitality of streets, and the types and areas of vegetation all can affect the human perception of the urban environment (Liu et al., 2015). This information is closely related to the intensity of human activities, activity types, and urban functions, and is an important aspect of analyzing human social perception.

2.2 Computer vision

Computer vision is a discipline about studying machine vision capabilities or a discipline that enables machines to visually analyze the environment and the stimuli in it. Computer vision usually involves the evaluation of images or videos. It was defined as "the automatic extraction, analysis, and understanding of useful information from a single image or a series of images" by the British Machine Vision Association (BMVA). In the field of computer vision, the main tasks are image classification, image location, target detection, target tracking, semantic segmentation, and instance segmentation (see **Figure 2.1**).

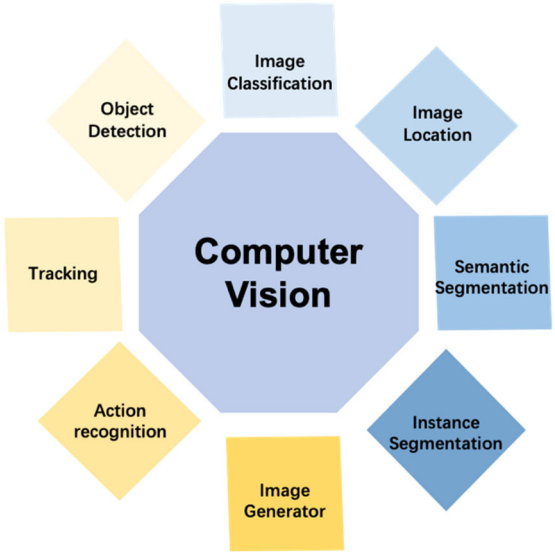


Figure 2.1 Computer vision tasks.

2.2.1 CNN Image classification

Convolutional Neural Network (CNN) is a type of feedforward neural network that contains convolution calculations and has a deep structure. It is a representative model of deep learning. Its iconic model AlexNet (Krizhevsky et al., 2012) won the first place in the ImageNet 1000 object picture recognition competition (Russakovsky et al., 2015) in 2012, with a score of less than 15%. The error rate far exceeds the model using traditional methods (26% error rate). However, there are some limitations of CNN, such as the storage overhead being too large, the calculation efficiency being low, and the pixel block size limiting the size of the sensing area, resulting in low classification performance.

2.2.2 FCN Image semantic segmentation

To resolve the issues mentioned before, FCN was released in 2015 (Long et al., 2015) and is the pioneering work of fully convolutional networks in the field of semantic segmentation. The

main idea is to improve the image classification network into a semantic segmentation network and restore the size of the feature map by rotating the classifier (fully connected layer) into an up-sampling layer for end-to-end training. It opens up a new world for semantic segmentation.

2.2.2.1 GoogLeNet

GoogLeNet was developed by Christian et al. (2015). Such as the AlexNet, VGG, and other models be mentioned before all achieve better training effects by increasing the depth (number of layers) of the network. However, as the number of layers increases, it also brings a lot of negative effects, such as overfitting, gradient disappearance, and gradient explosion. Christian and his team proposed a structure called inception, hoping to improve the training results from another angle. This structure can use computing resources more efficiently and can extract more features with the same amount of calculation.

2.2.2.2 ResNet

ResNet was developed by He et al. (2016a). It solves the problem of the disappearance of gradient backhaul by introducing cross-layer links. The model won the 2016 Imagenet competition with 96.4% accuracy. ResNet effectively solves the problem that deep neural networks are difficult to train, and can train up to 1000 layers of convolutional networks.

2.3 Deep learning

Deep learning is an algorithm that uses multi-layer artificial neural networks as the basic architecture to perform characterization learning on data. It is a branch of machine learning. The emergence of deep learning allows computers to process more image problems more accurately and efficiently (He et al., 2016a; LeCun et al., 2015). According to different task types and model principles, deep learning can be divided into Generative Adversarial Neural Network (GAN), Recurrent Neural Network (RNN), Auto Encoder, Deep Convolutional Neural Network (DCNN), etc. The DCNN is mainly applied in the image data analysis area. Since then, deep learning has entered the public eye, which has been widely applied in image recognition, speech recognition, text analysis, unmanned driving, game competition, etc. It has made remarkable progress in these areas. There are numerous deep learning models, which have their advantages and disadvantages. Among these deep learning algorithms, SegNet which is a deep convolutional encoder-decoder architecture for image segmentation (Badrinarayanan et al., 2017), Pyramid Scene Parsing Network (PSPNet) (Zhao et al., 2017),

and DeepLab (Chen et al., 2017a) are the representative deep learning models. The details of these algorithms are described in the following subsections.

2.3.1 SegNet

SegNet was developed by Badrinarayanan et al., (2017). It was developed based on FCN and was a semantic segmentation model obtained by modifying the VGG-16 network. The novelty of SegNet lies in the way the decoder upsamples its low-resolution input feature maps.

2.3.2 Pyramid Scene Parsing Network (PSPNet)

Pyramid Scene Parsing Network (PSPNet) was developed by Zhao et al. (2017), It introduces more context information based on the FCN algorithm through global average pooling and feature fusion, so the features are in a pyramid structure.

2.3.3 DeepLab

The DeepLab series was proposed by Liang et al (2014), mainly using DCNNs and probabilistic graph models (conditional random fields) to achieve image pixel-level classification (semantic segmentation tasks). It has been updated four versions so far, including DeepLab V1(Chen et al., 2017a), DeepLab V2 (Chen et al., 2017a), DeepLab V3 (Chen et al., 2017b), and DeepLab V3+ (Chen et al., 2018).

2.4 Street View Image

Street view images include street view pictures and social media photos in a broad sense. Street view images refer to pictures collected by map service providers such as Google Maps, Tencent Maps, and Baidu Maps. These pictures are collected by street view vehicles crossing the city road network. It also includes street view images provided by crowdsourcing platforms such as Mapillary under certain standards. Such kinds of images are generally stored in the form of a panorama, which contains the 360° panoramic visual information of the shooting location. In the actual acquisition and use, the visual environment of each location can be expressed by multiple street view pictures facing different directions and natural perspectives. Social media photos refer to photos of indoor and outdoor urban landscapes shared by users on social media platforms. Such platforms include mainstream social media such as Twitter and Facebook, as well as sharing platforms for photography enthusiasts and travel enthusiasts such as Flickr and Panoramio. Street view images are strictly distributed in accordance with the road network, and social media photos are distributed in various public spaces in the city, which can be used

as a supplement to the description of the spaces within the block. Street view images have the characteristics of wide coverage, high density, and high acquisition efficiency. The resolution of street view images has also been gradually improved with the advancement of image technology and photographic equipment, and the coverage has also become wider. For example, Google Street View images currently cover most cities in 195 countries around the world. At the same time, street view images have covered all levels of the city's road network with high density. The images formed between adjacent sampling points can be seamlessly connected and can fully express the built environment.

2.5 The applications of street view images in the urban environment analysis

The application of street view images in urban environment analysis includes two parts: 1) see cities from above; 2) see cities from a street level. In the traditional urban research process, researchers analyze the urban environment based on geographic information systems by identifying remote sensing images. Such methods have been widely applied in land use analysis, air pollution analysis, ecological environment analysis, and urban heat island effect analysis. It was found, they are very suitable for macro observation of large urban-scale areas. However, limited by the shooting angle of the image, it is not suitable for observing the microscopic built environment of the city. For example, the indicators used for environmental quality assessment, the urban green vision, sky openness, street valley index, etc., are difficult to be obtained from the remote sensing images. The emergence of street view images fills up this shortcoming, because it is acquired through ground-based photography equipment, which expresses the urban material environment from a human perspective, and has more detailed visual content (Gong, 2019). Some studies have pointed out that the visual indicators obtained from street view images are highly correlated with street feasibility and psychological conditions. On the other hand, because the shooting angle of street view images is similar to that of pedestrians, it can help researchers better understand the urban environment from the perspective of people, and it is usually used in the research of analyzing social perception based on big-data (Wang et al., 2018; Lu et al., 2019; Kang et al., 2020). The building types, colors, plants, and city traffic information, which are contained in street view images are all closely related to urban land use, urban functions, and the intensity of human activities, thereby helping researchers understand the impact of urban design on human activities (Zhang et al., 2019). Therefore, more and more researchers use machine learning and deep learning techniques to extract city

analysis data from street view pictures to study and evaluate the quality of the city's environment (see Table 2.1).

Table 2.1 Urban environment analysis based on street view image

Category	Evaluation elements	Method	Representative research
Physical environment	Green plants, pedestrian safety, pedestrians, road facilities, motorized transportation, construction, traffic signs, etc.	Correlation analysis Poisson regression Machine learning	Rundle et al., 2011; Kronkvist et al., 2014; Mooney et al., 2016
Social environment	Cars, sidewalks, pedestrians, buildings, sky, etc.	Machine learning Deep learning	Yin et al., 2015; Porzi et al., 2015;
Economic environment	Green vegetation, ground, buildings, tree, sky	Image analysis based on pixel Machine learning	Arietta et al., 2014; Glaeser et al., 2018;
Aesthetic environment	Street trees, green vegetation, buildings	Machine learning Deep learning Image analysis	Berland & Lange, 2017; Liu et al., 2017

2.5.1 Using Green View Index (GVI) to assess the quality of environment greenery based on street view image

Greenery is always cited as an essential factor in the study of urban environmental quality. Since the 1950s, Olmsted focuses on urban park renovation and street design that combines the natural environment and living space (Beveridge and Rocheleau,1995). In the later period of the 1980s, urban planners planned a large-scale green network to attract residents into the open space of the city. Another important contribution of green plants is to reduce the impact of air pollution and urban heat island effect to a certain extent. Therefore, most researchers pay more attention to the functional characteristics of plants but lack research on visual impact or aesthetics. It was found from some research by environmental psychologists (Tzoulas et al., 2007) that people's psychological feelings are closely related to the amount of green in the environment. Some research proved that 80% of people's perception of the surrounding environment comes from visual perception (Biocca and Delaney, 1995). Aoki (Yoji Aoki, 1987), who is a researcher at the National Institute of Environmental Studies of Japan, proposed a quantitative statistical analysis method to identify basic stimuli that affect specific

psychological changes and environmental green spaces that can produce positive subjective feelings. Later, the researcher of the institute officially proposed the concept of "green visible value" (Yoji Aoki, 1987), pointing out that the green visible value is the percentage of green in the human field of vision. Yoji Aoki (2006) summarized the research on visual greenery conducted in Japan since 1974, which confirmed the connection between the green quantity of the street and the psychological activities of people. This physical quantity will be a landscape evaluation factor for environmental greening. After that, many researchers tried to measure green visible values by different methods to quantitatively evaluate the urban green environment. To date, it has been challenging to estimate the size and location of urban green spaces. Traditional approaches rely on manual data collected by trained surveyors and community-based crowdsourcing (Seiferling et al., 2017; Wales, 2016). However, in some large-scale urban survey programs, volunteers are hired to assist in collecting data in the preliminary survey stage. Their usual lack of professional technical knowledge can cause sampling errors and even repeated sampling. To address these issues, means of calculating green space based on remote-sensing satellites and aerial imagery have been developed (Barbierato et al., 2019), and the combination of traditional remote sensing and proximity sensing seems to be a good choice. Obtaining geographic information by traditional remote sensing and mapping tree canopies based on high-resolution light detection and ranging (LiDAR) data has proven well suited for assessing urban green environments (MacFaden et al., 2012). However, software limitations and the high cost of acquiring high-quality data hinder the implementation of this approach on a large scale. The key limitation is that although satellite and aerial imagery quantify large-scale greenery relatively accurately, it is not good at showing street-level greenery (Yang et al., 2009). For this reason, the aforementioned approaches are useful for classifying large expanses of urban greenery, such as urban parks, urban forests, and gardens. Due to the lack of ground details, it is difficult to detect the contours and features of ground plants. Therefore, the assessment of street-level urban greenery remains a problem to be solved. With the development of urban planning, increasing emphasis has been placed on humanized urban space. The assessment of various elements of cities from the perspective of people has gradually become the focus of urban planning research. Therefore, an important aspect of assessing the level of urban greening is to evaluate residents' perceptions and experiences of the urban vegetation landscape from street level. There is a large difference in perspective between a street-level view and remote sensing from above. A street-level perspective can more intuitively reflect residents' actual perception of the surrounding environment. Fortunately, accessible data sources with geo-tagged data are becoming

increasingly common; for example, street view services on the web (e.g., Google, Tencent, Baidu) allow researchers to navigate virtually through urban spaces in the form of geo-tagged street-level images (Rundle et al., 2011). Some researchers began to use panoramic street view images to assess street-level greenery. Such images can fully display the surrounding environment as seen by pedestrians, creating what feels like a virtual tour of urban streets and giving people the feeling of "being there." These images are quite similar to what people see when driving, cycling, or walking on the street. Based on this, many researchers have tried different methods to quantify the visibility of greenery to evaluate the urban green environment. Meanwhile, with advances in computing technology, computer vision algorithms have been developed for processing street-level imagery to measure perceived urban safety (Naik et al., 2014), urban change (Naik et al., 2015), wealth (Glaeser et al., 2018), infrastructure (Zhang et al., 2018), and demographics (Gebru et al., 2017) and to classify building types (Kang et al., 2018). Yang et al. (2009) developed a method for evaluating the visibility of urban greenery by combining field surveys and manual photographs, and they developed the Green View Index (GVI) to represent pedestrians' view of greenery. However, their method relies heavily on manual input, which is laborious and prone to error. Li et al. (2015b) proposed a method for estimating the GVI by analyzing landscape images acquired from Google Street View (GSV). Their method involves using pixel-based color recognition in Abode Photoshop to recognize the green area in the images, but it still requires considerable effort. Li et al. (2015a) also proposed a method for calculating the GVI based on image-recognition technology and GSV images; this improved the work efficiency greatly and combined the latest computer technology with urban planning. To summarize the aims of those previous studies, they were focused on three key areas: (i) estimating the percentage of urban-level tree cover (Cai et al., 2018; Li et al., 2015b; Seiferling et al., 2017; Yang et al., 2009); (ii) calculating the number of urban trees (Branson et al., 2018; Wegner et al., 2016); and (iii) quantifying the sky view factor from street-level imagery to assess the effect of plant numbers on urban temperature (Li and Ratti, 2018, 2019; Li et al., 2017).

2.5.2 Using Sky View Factor (SVF) to evaluate the openness of the urban built environment based on street view image

In recent decades, there has been some substantial progress in research on improving the accuracy and efficiency of SVF estimation methods (Matzarakis et al., 2016; Zeng et al., 2018; Matzarakis and Matuschek, 2011). There are several representative estimation methods for calculating SVF in urban environments, which can be divided into the following types:

geometric-based methods (Watson and Johnson, 1987), GPS signal-based methods (Chapman and Thornes, 2004), simulation method-based 3D models (Cheung et al., 2016; Li et al., 2018), photographic methods (Grimmond et al., 2001; Chen et al., 2012), and SVI-based big data method (Gong et al., 2019; Middel et al., 2018).

2.5.2.1 Geometric methods

The geometric-based method was originally developed to measure SVF by calculating the ratio of the street width to the height of the buildings. In 1981, Oke first developed a formula to estimate SVF by calculating the ratio of the height of the buildings on both sides of the urban canyon to the width of the ground midpoint (Oke, 1981). In 1984, Johnson and Watson proposed a formula for calculating the SVF of a single building in an asymmetrical canyon of finite length (Johnson and Watson, 1984). In 2003, Botyan and Unger improved Oke's formula. They used a theodolite with a height of 1.5-m to measure two angles perpendicular to the axis of the street and calculated the SVF on both sides of the street (Botty´an and Unger, 2003). Geometric methods provide a simple theoretical basis for calculating the SVF in street canyons and provide a basis for accuracy and parameter analysis for future SVF estimation methods (Chen et al., 2012).

2.5.2.2 Global Positioning System methods

The GPS signal-based method estimates SVF values by obtaining GPS signal information through a GPS receiver (Chapman et al., 2002). Chapman and Thornes further developed GPS proxy technology to provide faster SVF calculations to achieve shorter processing times (Chapman and Thornes, 2004). This method is mostly used in the research of urban environment, but because the diversity of tree types, heights, and morphology increase the difficulty of measurement, so it is poorly applied in areas with high plant coverage. The main limitation is that the GPS signal-based method is an indirect modeled estimate. For example, the regression model coefficients need to be adjusted according to different research areas and land-use types, and the differences in local vegetation should also be considered. This method has high requirements in terms of manual operation and on-site measurement.

2.5.2.3 Fish-eye photographic methods

Developments in computer power have seen the use of digital mapping techniques such as 3DSky View, Arc View SVF, and digital surface models (DSMs) for estimating SVF values in artificial environments. These methods are based on urban morphology modeling and computer geometric technology and can quickly and effectively measure continuous SVF over large

areas (Gal et al., 2009). However, these methods also have obvious limitations; for example, it is difficult to use in some areas lacking 3D model information or DSMs. Furthermore, in actual urban environments, plants are one of the main characteristics of space. However, 3D models usually lack plant information, so estimated SVF values based on 3D city models only are usually high because they ignore the obstacles of plants to solar radiation. Recently, many researchers have focused on using open SVIs such as Google Street View (GSV) images, Baidu Street View images, and Tencent Street View images to represent the urban street environment and calculate the SVF, green view index (GVI), and other street environment evaluation indicators. Carrasco-Hernandez et al. also confirmed the reliability of estimating SVF values based on GSV images (Carrasco-Hernandez et al., 2015). They proposed using open-source panoramas to generate fisheye images and calculate SVF values. Although this method saves substantial time for field surveys and photography, it remains time-consuming and laborious in terms of performing manual image processing using large-scale city data. Lindberg et al. proposed a software named sky view factor calculator which can compute the SVF on hemispherical photographs using a Graphical User interface (GUI) (Lindberg and Holmer, 2012). However, it needs manual correction of non-sky pixels. Usually, light areas such as windows and white walls are classified as the sky and this has to be corrected to obtain correct SVF values. With the development of computer technology, several researchers have attempted to use the high efficiency of deep learning and image recognition technology to extract information from a large number of street scene pictures and perform relevant analysis. Dong et al. used image segmentation methods to identify plants from Tencent Street View images and calculate the amount of visual greening to evaluate the level of visual greening of the city (Dong et al., 2018). Li et al. proposed a semantic segmentation method based on U-Net to extract architectural footprints from high-resolution multispectral satellite images. U-Net is a popular deep convolutional neural network architecture for semantic segmentation. it is the most commonly used and simplest segmentation model in the field of semantic segmentation. Its original intention is to solve the problem of medical image segmentation, using an encoder-decoder structure (Li et al., 2019a). Liang et al. used an open-source deep convolutional encoder-decoder architecture called SegNet to estimate SVF values from GSV images (Liang et al., 2017). Zeng et al. developed an automatic method to estimate SVF values by stitching SVIs into panoramic SVIs to detect sky areas based on OpenCV (Zeng et al., 2018).

2.5.2.4 Simulation methods

The simulation method refers to the method of calculating SVF by simulating the urban environment by using digital application technology and computer computing power (Chen et al., 2012). This method provides a fast way to simulate and calculate the continuous SVF of a large area of the city based on city-based morphological modeling and computational geometry technology. Extension programs such as the 3DSky view program and ArcView SVF use GIS-based 3D building data to reconstruct the city structure through a computer (An et al., 2014). SVF Engine is a computational framework, which calculates SVF by generating virtual fisheye images from a 3D city model. These techniques are suitable for urban analysis that needs to quantify the characteristics of various buildings and natural environments (Liang et al., 2017).

2.6 Summary

Street view images focus on recording city street-level scenes from a human perspective. The data covers a wide range and the cost of collecting data is low. It is an important new data source in current urban application research. The continuous development of artificial intelligence and computer image technology, and the combination of deep learning and high-performance computing have solved the problem of data processing and information extraction of a large number of street scene images in a short period and developed a large-scale urban analysis based on artificial intelligence technology. At the same time, the new method of urban physical feature extraction and graphical expression has promoted the research of urban social environment and economic environment evaluation.

In the past, due to the lack of effective technology, the use of street view images was mostly small-scale, artificial comparative analysis. With the development of the computer field and the introduction of many deep convolutional neural network algorithms, it has become possible to use the deep learning algorithm to identify and classify the sky, sidewalks, lanes, buildings, and vegetations from street view images accurately. The combination of machine learning and street view images has changed the situation that in the past it was difficult to obtain basic street data and that street view images were difficult to use efficiently. The application of related algorithms of machine learning technology can not only provide refined basic data for spatial quality research but also can quickly process large-scale data while ensuring refinement.

Chapter 3

Deep learning based urban landscape element detection

The traditional urban landscape analysis methods were mostly relying on manual data collection and analysis, which is time-consuming, error-prone, and cost-intensive. Therefore, the objective of this chapter is to develop an approach that can accurately and efficiently detect and classify the urban landscape elements from street view images, then be used to analyze the urban environment quality.

In this chapter, the DeepLab V3+ model is been introduced to semantic segment the street view images. Then the accuracy verifying of the proposed method and the accuracy improvement is described in detail. The proposed method pre-train the deep learning algorithm with manually labeled 300 street view images combined with the Cityscape dataset. The proposed deep learning model is verified and the accuracy is validated through case studies. The proposed approach for detecting and estimating the urban landscape analysis factor is extended in Chapter 4 and Chapter 5.

3.1 Semantic segmentation of street view images based on deep learning

With the development of artificial intelligence technology, deep learning has outstanding performance in image recognition tasks. Li et al., (2018) proposed a method to identify the green and blue pixels from street view images using a sematic segmentation technique. Therefore, this research proposed to use the DeepLab V3 model for semantic segmentation to automatically segment the street view images and classify them into common urban landscape elements (e.g., vegetation, sky). The research proposed to use the collection of annotated images from the CityScapes Dataset to pretrain the network, which contains a diverse set of stereo video sequences recorded in street scenes from 50 different cities, with high-quality pixel-level annotations of 5,000 frames in addition to a more extensive collection of 20,000 weakly annotated frames. By entering a Street View image into the pre-trained network, the semantically segmented images can be obtained, which include vegetation, sky, buildings, and so on, that have been identified (see Figure 3.1).

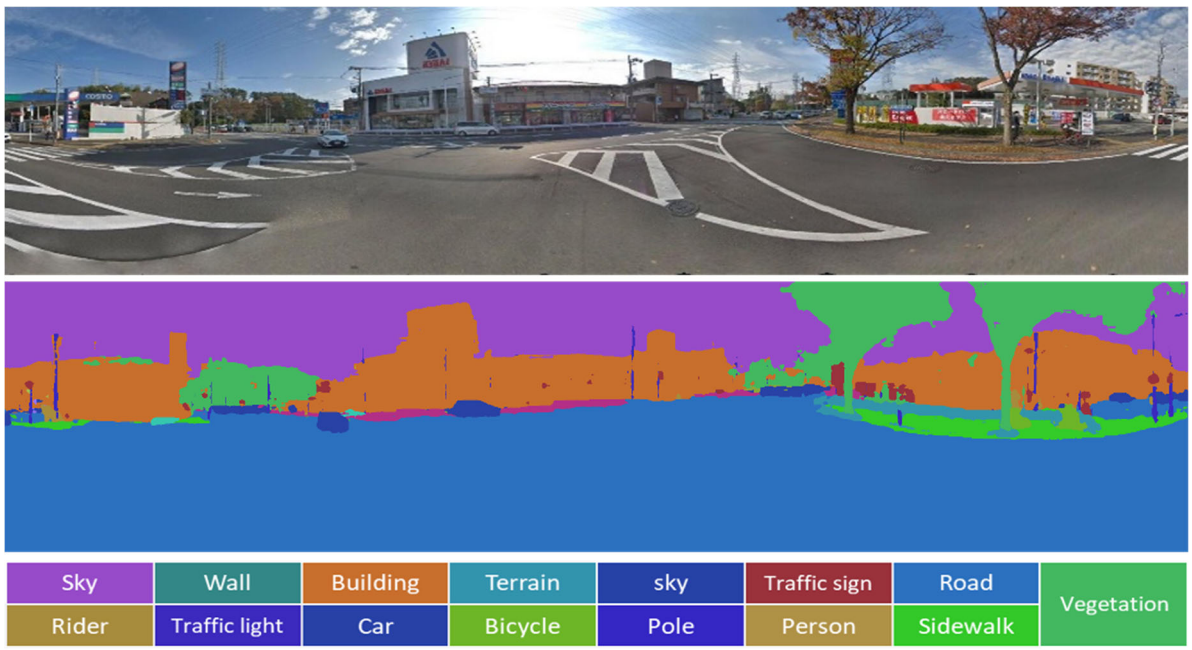


Figure 3.1 The result of street view image semantic segmentation using DeepLabV3

3.2 The verification of street view image detecting accuracy using deep learning

100 street view images were randomly selected to verify the accuracy of semantic segmentation. By manually marking with Adobe Photoshop and automatically detecting using the proposed method to obtain the classification results of the 100 images (see Figure 3.2). This research created reference images by marking the vegetation using Adobe Photoshop and filled it with orange. The areas identified as vegetation by the proposed semantic algorithm are filled with green, and the remaining areas are filled with black. After combining two images into one using multiply-blend mode of Adobe Photoshop, the region where the orange color of the manually marked image (see Figure 3.2b) overlaps with the green color of the segmented image (see Figure 3.2c) is the correct region, which is filled with dark green (see Figure 3.2d). In the composite images, the definition of different color regions is as follows in Table 3.1**.** Then, the rate of pixels of each part in the combined images was calculated to perform verification analysis, the calculation method is as follows in Table 3.2. As shown in the table, the accuracy rate is about 0.94, however, the inaccuracy rate is still about 0.06, there is some part like lawns or poles that are hard to be detected or false detect.

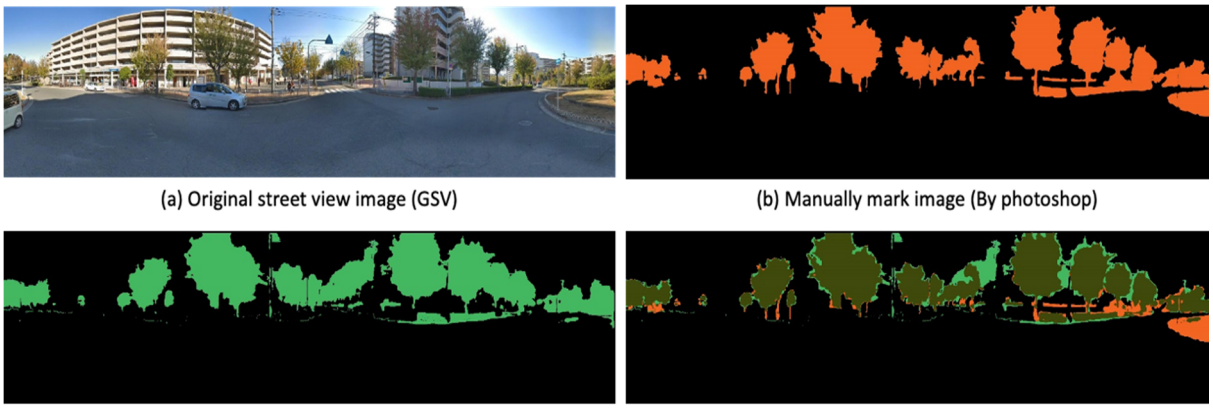


Figure 3.2 The accuracy verification of the proposed semantic segmentation method

Table 3.1 The definition of each color in the composite image

Color	Definition
Olive Green (OG)	Accurately extracted pixels
Black (B)	Accurately unextracted pixels
Green (G)	Over-extracted pixels
Orange (O)	Unextracted area pixels

Table 3.2 The calculation formula of accuracy rate and inaccuracy rate

Accuracy rate	Calculation formula	Result
Extract accuracy rate [%]	$\frac{OG}{X} \times 100$	0.15
Unextracted accuracy rate [%]	$\frac{B}{X} \times 100$	0.78
Accuracy rate [%]	$(\frac{OG}{X} + \frac{B}{X}) \times 100$	0.94
Over extracted rate [%]	$\frac{G}{X} \times 100$	0.03
Unextracted inaccuracy rate [%]	$\frac{O}{X} \times 100$	0.03
Inaccuracy rate [%]	$(\frac{G}{X} + \frac{O}{X}) \times 100$	0.06

3.3 The improvement of the street view image semantic segmentation accuracy based on DeepLab V3+ model

Based on the results mentioned in the previous section, to improve the accuracy of semantic segmentation of street view images using the DeepLab V3 model. 300 street view images were labled manually and were combined with the Cityscape dataset to train the deep learning algorithm. This research chooses to rely on pre-trained weights provided by the official to train the DeepLab V3+ model on the PC (NVIDIA GeForce RTX 2080Ti and Intel(R) Core (TM)

i7-8086K CPU @ 4.00 GHz). Since the architecture has already been successfully trained for a similar purpose, and it can be leveraged by using the final weights. The training result as shown below (see **Figure 3.3**), after 300,000 steps the mIoU reached 0.7792. 100 street view images were randomly selected to evaluate the performance of the semantic segmentation model after training.

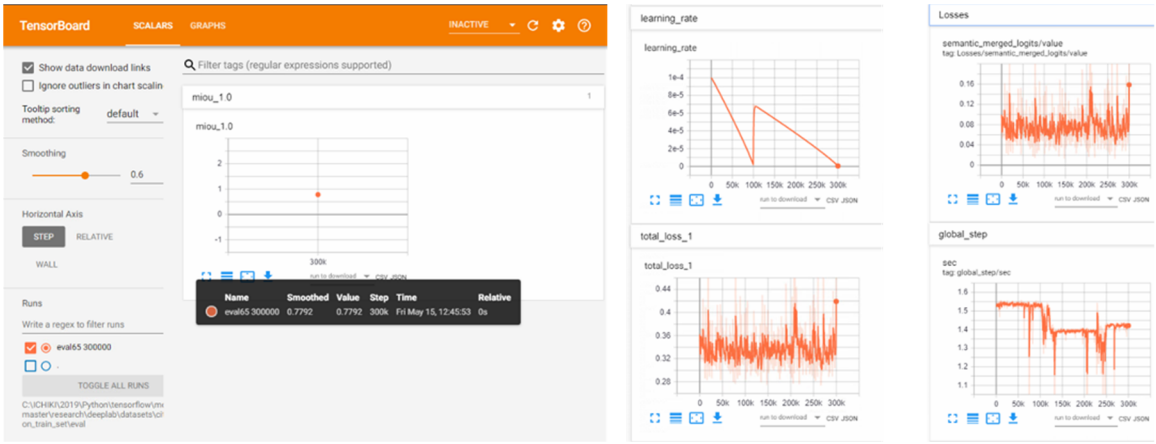


Figure 3.3 The result of training DeepLabV3+ with proposed dataset

The scatter plot shows the relationship between the vegetation percentage of the image detected using proposed method and the corresponding values based on the reference data delineated manually using Adobe Photoshop (see **Figure 3.4**). The vegetation percentages are distributed near the 1:1 line, and the regression coefficient is up to 0.9809, which indicates that the vegetation pixels classified using the two different methods were quite similar. Additionally, the root means square error (RMSE) is 0.018, which means that the vegetation classified automatically by the proposed system is of high accuracy. This verification indicates that the proposed method can be used to classify the urban design elements from street view images for urban landscape analysis.

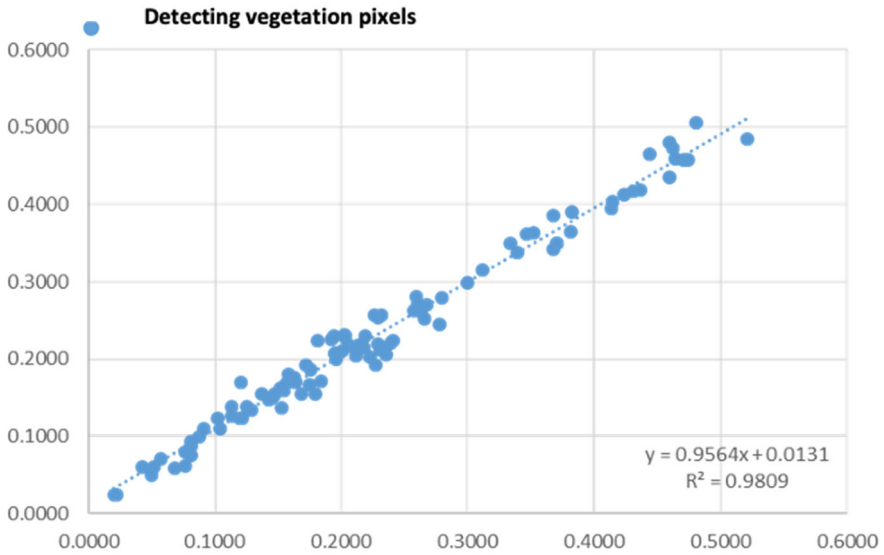


Figure 3.4 A scatter plot of vegetation pixels classified using the suggested deep learning-based method vs the reference data delineated manually using Photoshop.

3.4 Summary

The application of street view images in urban environments analysis using deep learning outperforms traditional methods, but it still has limitations. Various studies have attempted to improve the efficiency and accuracy of semantic segmentation based on street view images. However, the urban environment is a very complex system, the urban scene contains various elements, which are intertwined with each other. The dataset used to train the deep learning model is very limited, which does not cover all possible urban scenes. At the same time, due to the different backgrounds, cultures, and development history of various cities, there will be differences in their street space, architectural styles, and plant species. This will also reduce the segmentation accuracy of the existing deep learning model for different elements classification in the street view image. Although some studies proposed improved methods, including using various semantic segmentation algorithms, to improve the detecting accuracy from street view images, no research has focused on using the manually labeled dataset to train the deep learning model. Therefore, this research focuses on improving the accuracy of deep learning-based vegetation and sky area detection from the street view images, using the pre-trained semantic segmentation model. The urban landscape elements in this study are street-level greenery and the openness of the built environment because they constitute most of the visual perception spaces at the street level. The proposed methods, the validations, the discussions, and the conclusion will be presented in the following chapters. #

Chapter 4

Assessing the quality of urban greenery by estimating the Green View Index

Street greenery has long played a vital role in the quality of urban landscapes and is closely related to people's physical and mental health. Also, the level of street greenery is a key factor of urban landscape quality analysis. However, despite extensive research into environmental assessment methods for urban greenery, plant identification and greenery index calculations are still mostly done manually. This research developed a method based on semantic segmentation processing of street view images to calculate the Green View Index of urban streets, and the Panoramic View Green View Index (PVGVI) is proposed for measuring the visible street-level greenery. The research validated the results by comparison with those of manual inspection and the Pyramid Scene Parsing Network method. The vegetation detection rate of the proposed method is very close to the ground truth value, which means it can distinguish almost all of the vegetation information from the street view images, and the calculated PVGVI is reliable. In addition, this research conducted a case study of street-level greenery using the PVGVI and confirmed that this method can better visualize urban street-level greenery. The proposed method is scalable and automatable, and it contributes to the growing trend of integrating large freely available street view image datasets with semantic segmentation to inform urban planners.

4.1 Introduction

As the global urban population expands, global environmental problems are becoming increasingly serious, and the human living environment is continuously threatened by extreme severe climate, urban heat island effects, and air and water pollution (Blanco et al., 2009). Urban green spaces, including parks, street trees, community gardens, and green roofs, provide numerous ecosystem services on a local scale and represent a potential adaptation strategy for offsetting the growing impact of human activities on the urban environment. Thus, there is an urgent need to assess whether the urban green environment can help mitigate the impact of climate change on human health and to increase the number of urban public green spaces.

Urban street-level greenery has long been recognized as one of the most prime landscape design elements in the urban ecological system (Wolf, 2005). It provides multiple benefits to urban environments, such as urban trees that can adjust environment temperatures by providing shade and cooling, helping to significantly reduce the risk of heat-related illnesses (Mavrogianni et al., 2014). The street-side green spaces bring huge benefits to cities (Bain et al., 2012), it provides opportunities for community residents to engage in physical exercise, thereby reducing obesity and mental stress (Giles-Corti et al., 2003). It also provides more opportunities to be in green spaces, which is conducive to improved mental health, especially in reducing the risk of attention deficit disorder in childhood (Louv, 2008). Urban street-level greenery makes an important contribution to the attractiveness and walkability of residential streets (Schroeder and Cannon, 1983; Bain et al., 2012). At the same time, it also provides a welcoming environment for people who have a certain impact on the occurrence of various mental illnesses (Coutts, 2008; Lee and Maheswaran, 2011; Leslie et al., 2010). Therefore, ensuring widespread access to street-level green space is a key factor in providing these environments and benefits to residents (Landry and Chakraborty, 2009).

4.2 Study area and data collection

4.2.1 Study area

This research was conducted in the city of Suita in northern Osaka, Japan. Founded on April 1, 1940, Suita hosted Expo ’70 (a world’s fair held in 1970) and is home to Osaka University. The study area is a 3.07 km^2 residential area near the Suita campus of Osaka University, the location and road map of which are shown in Figure 4.1. The study area has an extensive green urban fabric, with parks and gardens covering around 20 % of the total area, and it offers mature

plant morphology with which to measure and understand the greenery quality. The variety of greenery (e.g., street greening, courtyard greening, and private greening) provides a wealth of research objects, and the data obtained from surveying this area will provide a useful framework for assessing the proposed method.

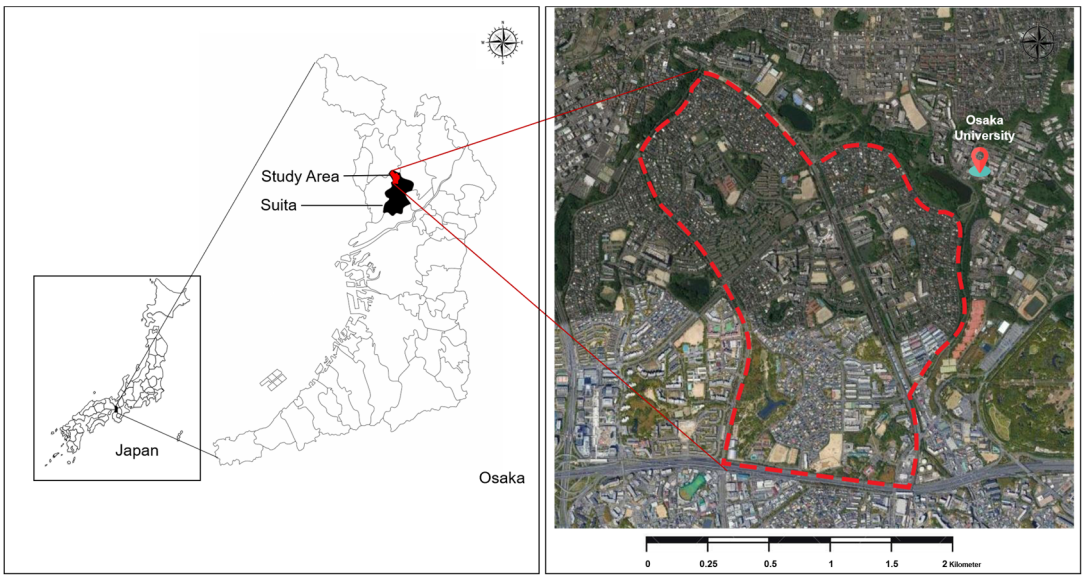
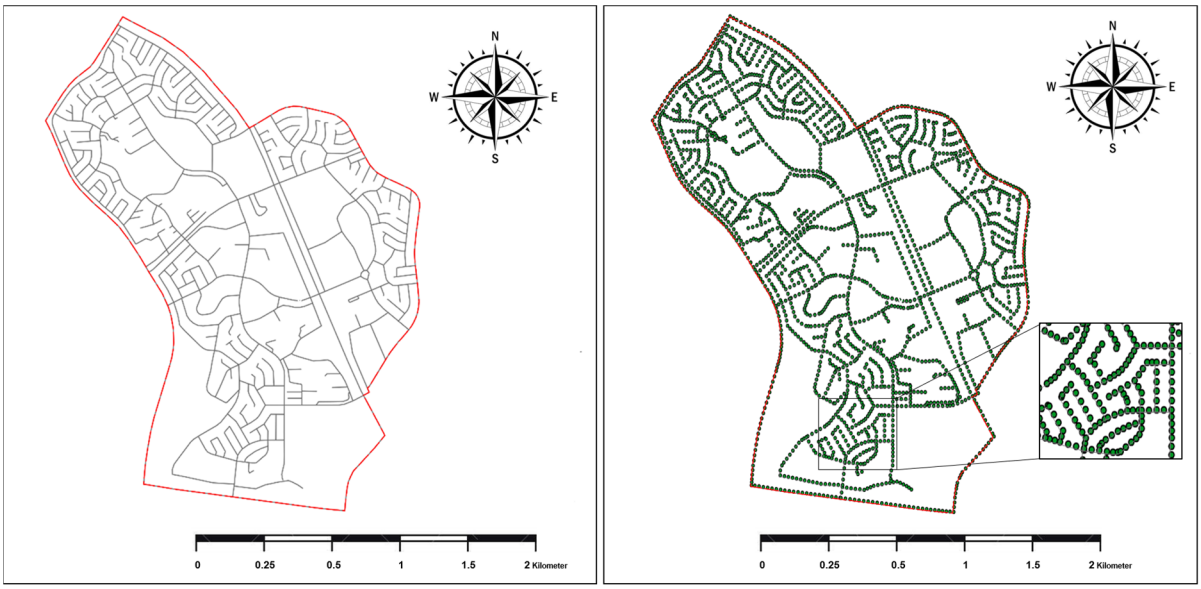


Figure 4.1 Location of the study area.

4.2.2 Data collection

This research used two main data sources to test the proposed image segmentation algorithm and estimate the PVGVI values; they are Cityscapes dataset (Marius et al., 2016) and GSV images. Cityscapes dataset was used to benchmark the performance of the proposed method. It is a new large-scale urban landscape data set and a benchmark for urban scene image segmentation. It contains a diverse set of stereo video sequences recorded in street scenes from 50 different cities, especially with high-quality pixel-level annotations of 5000 frames in addition to a larger set of 20,000 weakly annotated frames which is very useful in urban research. The study used it as a dataset to train the semantic segmentation model. The GSV images were used to apply the proposed method to the study area, the GSV images are street-level imagery data that provide extensive geographical coverage, and standardized, geocoded, and high-resolution images of the urban environment. It is a city image resource that can be quickly obtained in large quantities. Note that although GSV images are used here to reflect the complete view of pedestrians at street level to calculate the PVGVI value, however, the proposed method is not dependent on GSV data and can be applied to arbitrary images captured at the street level from multiple sources.

In urban planning, researchers are increasingly using street view images to audit the urban environment and various environmental elements. Because such images reflect cities from the actual perspective of pedestrians, they can give the feeling of being there. In addition to street-view map services for user browsing, most providers have released application programming interfaces (APIs) to allow the development of customized web applications. The research obtained the street network from the Open Street Map website (see Figure 4.1). Then, the following steps were used to process the street network data to meet the requirements: (i) Applying vector clipping to the study area, this research reduced bidirectional parallel vectors of the same street to a single vector and connected discontinuous streets. (ii) Because the roads amounted to a total length of 574,285 m, and the research sampled points at 30 m intervals along the street lines shown in Figure 4.2; in total, the study area contained 2942 sampling points with coordinates. (iii) This research downloaded street view images taken along each street, although some sampling points had no data.



(a) (b)

Figure 4.2 (a) Road map and (b) sampling points of the study area.

The GSV Static API was used to download the street view images. By specifying different parameters in the API, users can download GSV images with different fields of view and heading and pitch angles. The required API parameters are listed in Table 4.1 and include the street view image size, the location or location ID, the horizontal and vertical angles, and the developer's key. In this research, by setting the acquisition requirements given in Table 4.1, where LAT and LON are the latitude and longitude, respectively, Fov set as 60 to simulate the

horizontal field of view of the pedestrians, Heading (0 by default) indicates the compass heading of the camera, Pitch (0 by default) specifies the up or down angle of the camera relative to the street view vehicle, here it was set as 0. Finally, by entering the coordinates of the sampling points into a Python script, 24,920 GSV images (1000 ×1000-pixels) were downloaded and stored for green vegetation classification and PVGVI calculation. Figure 4.3 shows an example of the downloaded GSV image and the generated panoramic image. The requested GSV images were captured in 2015–2019 in various seasons except for winter; most were captured in 2018–2019, with the older images tending to be of quiet side streets or the corners of parks.

Table 4.1 API parameters for crawling GSV.

Parameter	Description	Example
Size	The output size of the image in pixels	size = 400×400 returns an image that is 400 pixels wide and 400 pixels high
Location	Coordinates of GSV location	location = 34.80932445, 135.5066877
Heading	Compass heading of camera; accepted values are 0–360	North: heading = 0 (360) South: heading = 180 East: heading = 90 West: heading = 270
Fov	Horizontal field of view of the image	fov = 60
Pitch	Up or down angle of the camera relative to GSV vehicle	pitch = 0
Key	Developer's key (retrieved through online application)	key = your API key

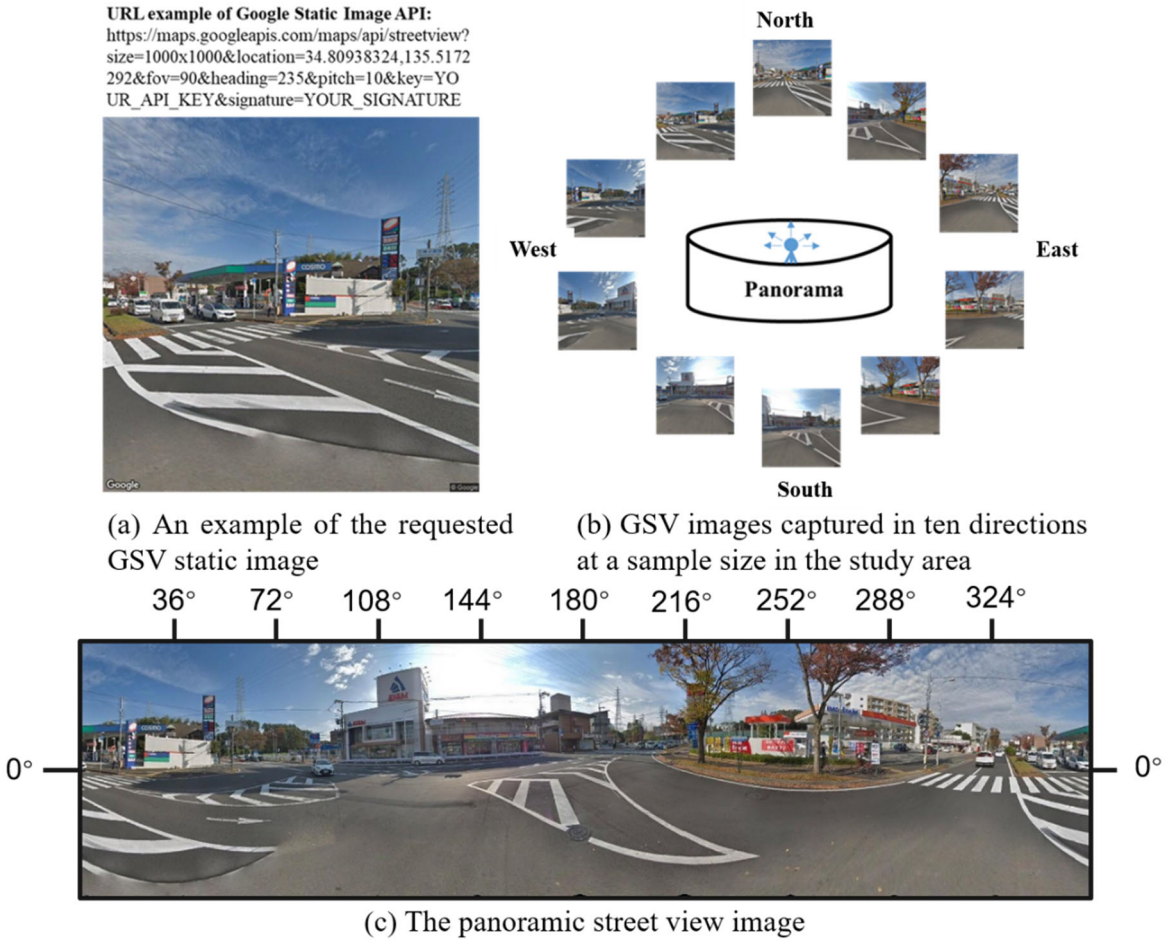


Figure 4.3 Example of panoramic street view image using GSV API.

4.3 Proposed methodology

4.3.1 Image semantic segmentation

Image semantic segmentation refers to use machine learning to understand and segment image content semantically. Based on semantic segmentation, scene parsing is a fundamental topic in computer vision, the goal being to assign a category label to each pixel in the image. Both scene and image segmentation can be viewed as extensions of object detection, their purpose being to locate specific objects in an image and classify the image pixels into a class from a series of discrete categories describing the image. Consequently, scene analysis has gained increasing recognition and an important correlation role in applications such as autonomous driving in recent years. Furthermore, semantic segmentation allows the automatic detection of elements such as plants, buildings, sky, and roads in images, thereby facilitating the calculation of the greening quality, space openness, and building closure of urban streets. Deep learning is the

current state of the art in scene parsing (LeCun et al., 2015) and is typified by deep convolutional neural networks (Krizhevsky et al., 2012; LeCun et al., 1998). Recent deep learning research has involved various sophisticated architectures, such as fully convolutional neural networks (Long et al., 2015), PSPNet (Zhao et al., 2017), the semantic segmentation network SegNet (Badrinarayanan et al., 2017), Deep Labelling for Semantic Network (DeepLabV3; Chen et al., 2017b), and DeepLabV3+, a more recent iteration of DeepLabV3 (Chen et al., 2018). These models both build on the earlier work in the related field of object detection and classification, where very deep convolutional networks (e.g., VGG16; Simonyan and Zisserman, 2015), deep residual learning image recognition (e.g., ResNet; He et al., 2016b) and the Inception networks (e.g., GoogleLeNet; Christian et al., 2015) have notably improved the ability to detect and classify objects in general terms. This research took a scene-parsing approach based on DeepLabV3+, using a model provided by Chen et al. (2018) for the following two main reasons. First, DeepLabV3+ was specifically designed to parse urban scenes and so is a good choice for identifying vegetation in street-level images. Second, DeepLabV3+ has outperformed several of the most popular deep-learning algorithms (e.g., ResNet, PSPNet) in major performance evaluation competitions such as the 2012 Pa VOC benchmark and the Cityscapes benchmark. DeepLab v3+ uses ASPP (Atrous Spatial Pyramid Pooling), using multiple effective fields-of-view and upsampling to achieve multi-scale feature extraction. At the same time, deep separable convolution is used to reduce the number of parameters and improve calculation efficiency. Figure 4.4 shows the workflow for semantic segmentation using DeepLabV3 +. The downloaded and combined panoramic street view image was inputted into the model, which passes through the backbone network (backbone, which is the part marked as DCNN Atrous Conv in the figure) to get two outputs: one is a low-level feature (output = 4x output); the other is advanced features (output = 16x output), using for ASPP output. In the Encoder part, the advanced features get 5 outputs through 5 different operations of ASPP, and the output stride = 16x is obtained after concatenating and 1 × 1 convolution. In the Decoder part, the low-level feature adjusts the dimension through 1 × 1 convolution, the Encoder output is up-sampled 4 times, and then concatenates the two 4x features, through some 3 × 3 convolutions and up-sampling 4 times to obtain the Dense Prediction. Finally, the per-pixel prediction and produce a pixel-wise classified street view image with semantic categories can be obtained. The output image has the same size as the input image includes 19 classifications and features of the vegetations were extracted and calculated in this study. This study relied on pre-trained weights to train the DeepLabV3+ model on the experimental computer. Due to the fact that this architecture had already been

trained successfully for similar purposes, something that this research leveraged by using those final weights. The contribution of this research in this context is to compare the predictive performance of a pre-trained DeepLabV3+ model on the Cityscapes benchmark and that of the PSPNet model trained on the same benchmark data (Cordts et al., 2016). Both of these sets of weights specifically identify the urban scene.

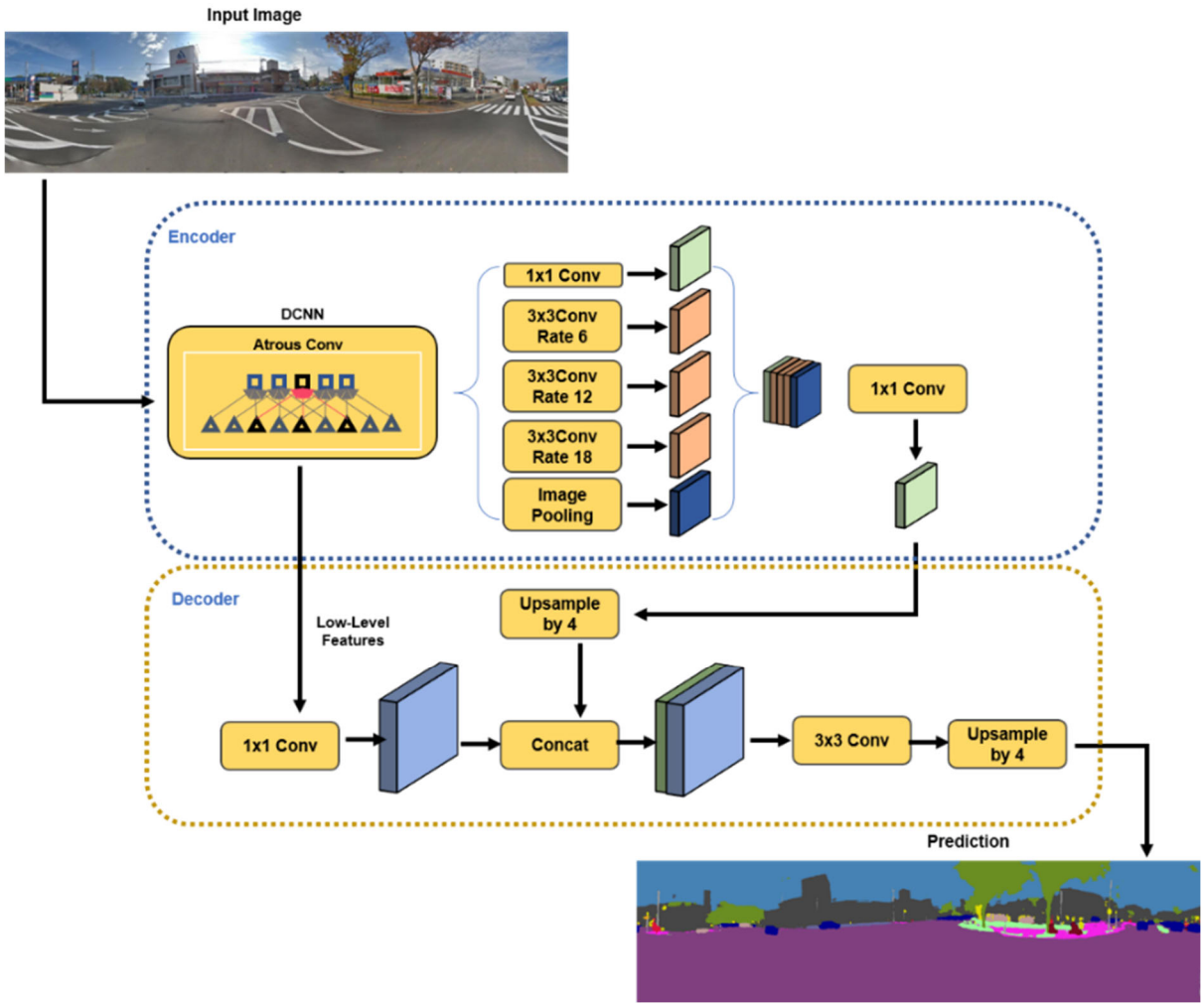


Figure 4.4 DeepLabV3+ architecture.

4.3.2 PVGVI calculation

In 1981, the National Institute of Environmental Studies of Japan proposed quantitative statistical analysis methods for identifying the essential sources of stimulation that affect specific psychological changes and to identify environmental green areas that can generate positive subjective feelings. Later, researchers of the Institute officially proposed the concept of a “green visible value” (Aoki et al., 1985) as the percentage of green in the human field of vision. This physical quantity can be used as a landscape evaluation factor for environmental

greening. After that, Yang et al. (2009) defined "the GVI to evaluate the visibility of urban greenery, this being the ratio of the total green area from four pictures taken at a street intersection to the total area of the four pictures, as given by the Equation (1):

$$\mathrm{GVI} = \frac{\sum_{i=1}^{4} Area_{g\_i}}{\sum_{i=1}^{4} Area_{t\_i}} \times 100\% \quad (1)$$

where $Area_{g\_i}$ is the total number of green pixels in the image taken horizontally in the direction i (= 1-4) for one intersection, and $Area_{t\_i}$ is the total number of pixels in that image.

To represent the view of pedestrians more effectively, Li, Zhang, Li, Ricard, et al. (2015) proposed the modified GVI, which used six images covering the 360-degree horizontal surroundings and three different vertical view angles was considered at each direction for every street sample site to calculate the GVI. The modified GVI is calculated using Equation (2):

$$\mathrm{GVI} = \frac{\sum_{i=1}^{6} \sum_{v=1}^{3} Area_{g\_iv}}{\sum_{i=1}^{6} \sum_{v=1}^{3} Area_{t\_iv}} \times 100\% \quad (2)$$

where $Area_{g\_iv}$ is the number of green pixels in one of these images captured in six directions with three vertical viewing angles for each sampling point, and $Area_{t\_iv}$ is the total number of pixels in one sampling point for the 18 GSV images (6 directions × 3 vertical viewing angles).

However, although the modified GVI considers the horizontal and vertical directions, some vegetation information is still omitted. Therefore, this research used a 360° panoramic street view image assembled using the PTGui software as a single image that represents the entire view of the visual environment at a specific location to evaluate the PVGVI around the street, as described by Equation (3):

$$\mathrm{PVGVI} = \frac{Area_{g\_i}}{Area_{t\_i}} \times 100\% \quad (3)$$

where $Area_{g\_i}$ is the total number of green pixels in the panoramic image along with direction i, and $Area_{t\text{-}i}$ is the total number of pixels in the image.

4.4 Experiments and results

This section first presented the segmentation results given by the pre-trained DeepLabV3+ model, and the research compared the proposed approach against the manual approach and the PSPNet methods described in Section 4.3. Then, this research assessed the relative performance

of these methods for segmenting image pixels into vegetation and non-vegetation classes with some evaluation metrics. And the computational performance of these methods was also assessed. Lastly, GVI in the study area is visualized to help us better understand the greenery quality.

4.4.1 Green vegetation extraction result

Figure 4.5 shows a visualization result of segmentation via the semantic algorithm discussed in chapter 3. The elements in the street view image are classified into different classes and marked with corresponding colors. Then, from the segmentation results, a CSV (comma-separated values) file is generated automatically that includes the image proportions of the 19 categories in Table 4.2, including roads, buildings, vegetation, sky, cars, and pedestrians.

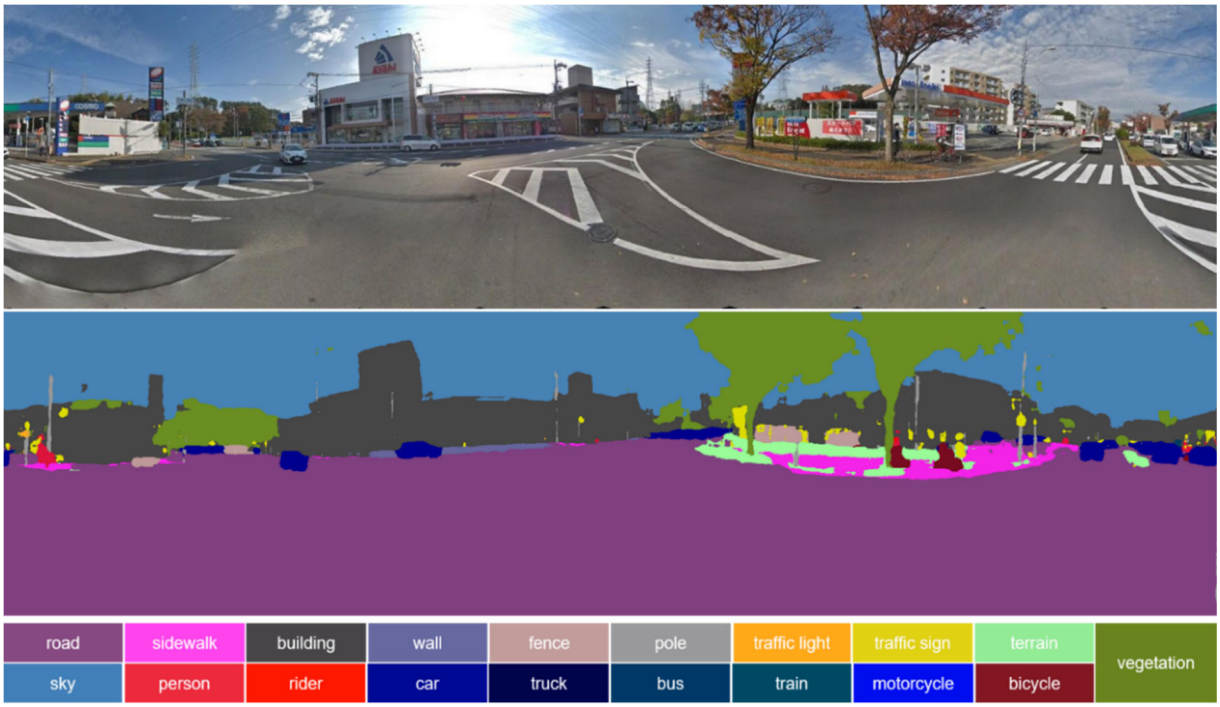


Figure 4.5 Visualization result of segmentation by the proposed method.

Table 4.2 Categories of semantic segmentation.

Number	Sort	Code	Number	Sort	Code
1	Road	0	11	Sky	10
2	Sidewalk	1	12	Person	11
3	Building	2	13	Rider	12
4	Wall	3	14	Car	13
5	Fence	4	15	Truck	14
6	Pole	5	16	Bus	15
7	Traffic Light	6	17	Train	16
8	Traffic Sign	7	18	Motorcycle	17
9	Vegetation	8	19	Bicycle	18
10	Terrain	9			

A part of the CSV file is given in Table 4.3, showing some of the segmentation results for five sampling points. Such data could be used in future research to assess urban design factors other than the GVI, such as sky view factor, street wall continuity, cross-sectional proportion, and street accessibility.

Table 4.3 Example of the percentage of pixel values for each type.

Point X	Point Y	Road	Sidewalk	Building	Sky	Vegetation	…
34.80932445	135.5066877	0.319069	0.013988	0.1311152	0.091107	0.262751	…
34.80933038	135.5091766	0.269873	0.010119	0.119538	0.149191	0.348329	…
34.8093325	135.5100636	0.267972	0.004118	0.206583	0.157556	0.208733	…
34.80933361	135.5105305	0.29529	0.003258	0.347979	0.155143	0.080301	…
34.80933467	135.5109766	0.259571	0.024141	0.246661	0.107456	0.195494	…

Figure 4.6 shows the semantic segmentation results for four randomly selected sample points in the study area. The first column of the picture matrix shows the original GSV images; the second column shows the vegetation and sky extracted manually using Adobe Photoshop CC 2019 as references to validate the automatically unsupervised classification results; the third

column shows the results extracted by the pre-trained DeepLabV3+ model discussed in Section 4.3.1, almost all the vegetation area of these images has been detected and marked; the final column shows the results extracted by the PSPNet method, compared with the proposed approach, some vegetation areas are not detected or are detected as other categories.

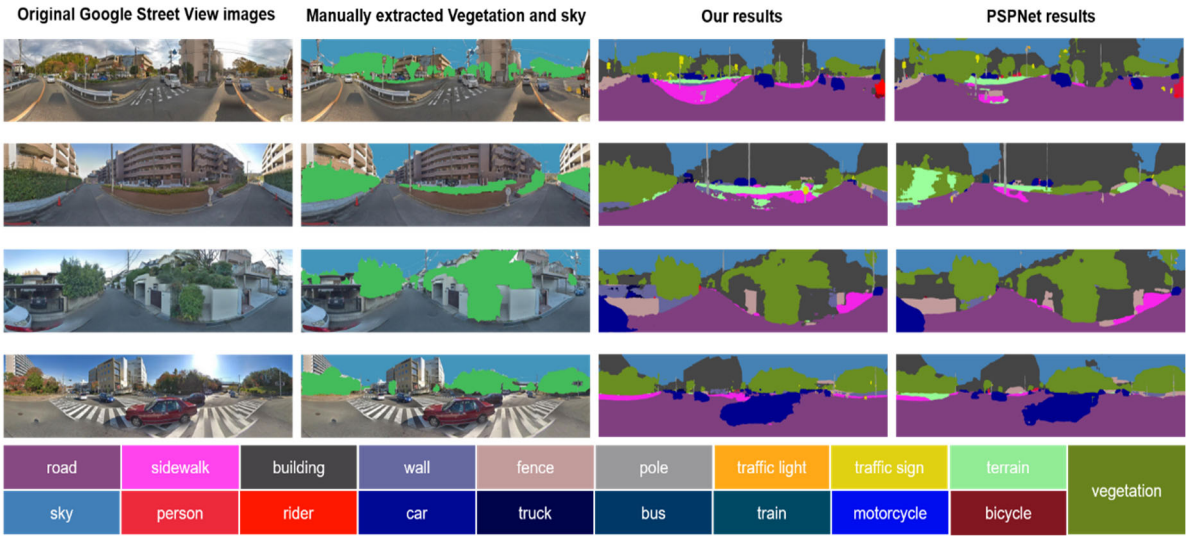


Figure 4.6 Green vegetation extraction results and references.

4.4.2 Comparative assessment

To verify further the accuracy of the segmentation results, this study selected 300 points randomly from the database and estimated their PVGVI values manually and with PSPNet methods. All the semantic segmentation analyses were performed on a Windows machine with an NVIDIA GeForce RTX 2080Ti and Intel (R) Core (TM) i7-8086K CPU @ 4.00 GHz.

Here, to compare the accuracy of vegetation extraction, the research used the following three main evaluation metrics that are commonly used to assess the accuracy of classification outcomes: (i) mean intersection over union (mIoU), which is used to measure the accuracy of the location of vegetation pixels; (ii) root-mean-square error (RMSE), which is used to measure the accuracy of estimating the overall PVGVI values; (iii) mean absolute error (MAE), which is used to measure the average over the test sample of the absolute differences between prediction and actual observation. This research calculated the mIoU based on Equation (4):

$$\overline{IoU_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{TV_i}{TV_i + FV_i + FN_i} \tag{4}$$

where n is the number of images, TV_i is the number of true-positive identified vegetation pixels in image i, FV_i is the number of false-positive identified vegetation pixels in image i, and FN_i is the number of false-negative rejected vegetation pixels in image i. The RMSE was calculated by Equation (5):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(TV_i - PV_i)^2} \tag{5}$$

where N is the number of pixels in a single image, TV_i is the number of true positive vegetation pixels identified in image i, and PV_i is the number of vegetation pixels predicted in image i. The MAE was calculated by Equation (6)

$$\overline{MAE_i} = \frac{1}{m}\sum_{i=1}^{n}\left|\frac{1}{N}\sum_{i=1}^{n}PV_i - TV_i\right| \tag{6}$$

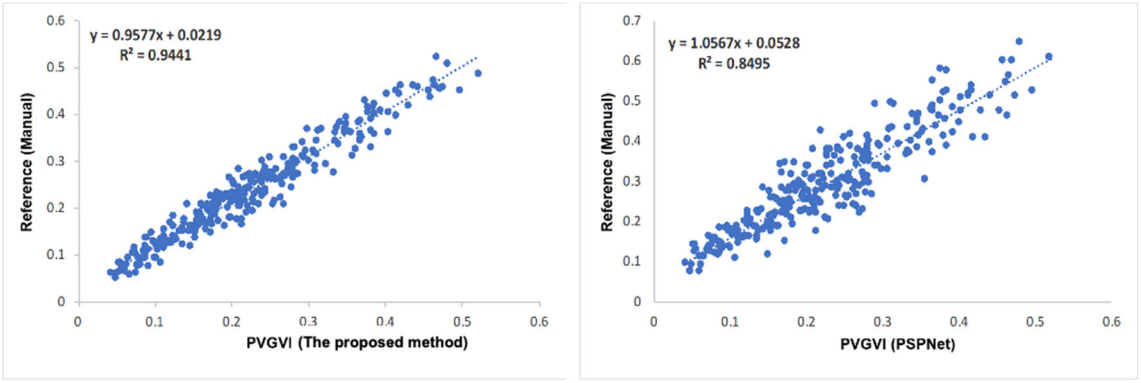
where m is the number of images, N is the number of pixels in a single image, TV_i is the number of true positive vegetation pixels identified in image i, and PV_i is the number of vegetation pixels predicted in image i.

This research presented the relative performance of these methods based on Cityscapes datasets to classify image pixels into vegetation and non-vegetation classes is assessed. Performance results from applying DeepLabV3+ are available in Table 4.4. The various performance metrics consistently show a significant improvement in prediction quality compared to the PSPNet method. For example, the higher mIoU and the lower RMSE of the proposed method indicate that it offers much better prediction accuracy and quality compared with the PSPNet. The R^2 is over 1.1 times higher than that of the PSPNet based method (see Figure 4.7). This research also assessed the computational performance of the two algorithms depending on the same implementation. It was found that the proposed method is significantly more efficient in terms of processing time, it took about 12 minutes to finish the 2,492 panoramic images, roughly 0.3 seconds per panorama image. It was demonstrating that the proposed method is more flexible and efficient.

Table 4.4 Results of evaluation metrics for the proposed method and PSPNet in the Cityscapes dataset.

Method	mIoU [%]	RMSE [%]	MAE [%]	Image per Second (s)
The proposed method	78.37	2.75	2.28	0.29
PSPNet	77.23	8.04	6.85	0.42

Next, the research used linear regression to compare the results of proposed method with those obtained manually and using PSPNet. Figure 4.7 shows the relationships among these results, with the fitting lines shown as dotted blue lines. Figure 4.7 (a) shows the fitting line between the PVGVI values calculated manually and by the proposed method; the regression line is very close to the 1:1 line and the correlation coefficient is 0.9716 with $R^2 = 0.9441$, which indicates that the PVGVI values calculated using the two different methods are quite similar. Figure 4.7 (b) shows the fitting line between the PVGVI values calculated manually and by the PSPNet method; the correlation coefficient is 0.9217 with $R^2 = 0.8495$, and the correlation coefficient is less than that with the proposed method. This shows that the PVGVI values calculated by the proposed approach are more consistent with the results of manual extraction than those calculated by the PSPNet method, which means that the proposed approach can be used to evaluate the quality of urban greenery.



(a) (b)

Figure 4.7 Accuracy assessment for the proposed method: "The proposed method" = PVGVI calculated with pre-trained DeepLabV3+ model; "Reference" = PVGVI calculated manually using Adobe Photoshop; "PSPNet" = PVGVI calculated with PSPNet method.

4.4.3 Comparison of PVGVI and modified GVI values

In this section, 300 points were randomly selected from the sampling points to compare the PVGVI and the modified GVI. The purpose is to determine whether the present method of measuring the PVGVI using stitched panoramic images could replace the measurement method proposed by Li, Zhang, Li, Ricard, et al. (2015).

The GVI values calculated using the modified GVI equation [i.e., Eq. (2)] and the present PVGVI equation [i.e., Eq. (3)] are seemingly different and the difference was calculated as the modified GVI value minus the PVGVI value (see Figure 4.8). As shown, most of the points have negative values, which means that the modified GVI is less than the PVGVI.

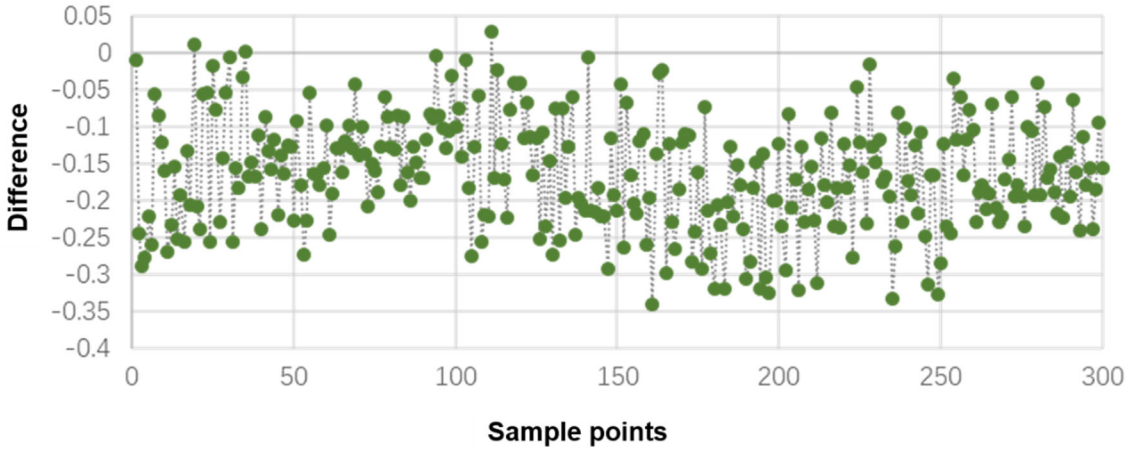


Figure 4.8 Value differences between PVGVI and modified GVI (value difference = modified GVI − PVGVI).

100 points were randomly selected from the 2,492 sampling points to compare the PVGVI and modified GVI values with manually marked ground truth values through linear regression, and the results are shown in Figure 4.9. The regression coefficient of the PVGVI values concerning the ground truth values is 0.9770 with R^2 = 0.9454, which is higher than the regression coefficient 0.7183 with R^2 = 0.875 of the modified GVI values concerning the ground truth values. The result indicates that the PVGVI values calculated by the proposed approach are more accurate than those calculated by the modified GVI calculation method.

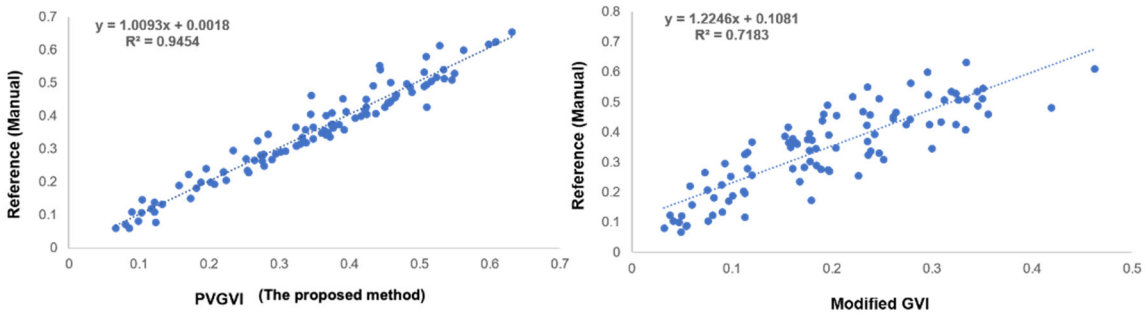


Figure 4.9 Accuracy inspection of the proposed method: "The proposed method" = values calculated with pre-trained DeepLabV3+ model; "Reference" = PVGVI calculated manually using Adobe Photoshop; "Modified GVI" = GVI calculated with a modified method.

It is easy to understand the differences between the modified GVI values and PVGVI values. Unlike the proposed PVGVI, the modified GVI is based on images captured at six horizontal angles and three vertical angles. Figure 4.10 shows the GSV images for a sampling point where the modified GVI is much less than the PVGVI. Figure 4.10a shows the panoramic GSV image that was used to calculate the PVGVI value. Figure 4.10b shows the GSV images at the three different vertical viewing angles toward the west; all these images are 640×640 pixels in size and were used to calculate the modified GVI value. The areas of greenness in the GSV images at the different vertical viewing angles are very different. The panoramic image shows clearly that all the vegetation there in is within the viewing angle range of a pedestrian looking horizontal. There are relatively few shrubs or tall trees, which results in the images with high and low viewing angles containing almost no vegetation (see Figure 4.10b and see Figure 4.10d). Therefore, the modified GVI value covers much sky and bare ground, making it much lower than the PVGVI value. By contrast, when there many kinds of vegetation are present (shrubs, tall trees, etc.), the panoramic image covers the whole view of pedestrians in a single image with no overlapping areas, however, the modified GVI values are obtained using images with different vertical angles that will contain the same area of vegetation, which will lead to overestimation of the real GVI values.

Figure 4.10 GSV images for a sampling point with different vertical viewing angles (i.e., pitches): (a) panoramic street view image, fov = 360; GSV images with (b) pitch = 45, fov = 60, heading = 60; (c) pitch = 0, fov = 60, heading = 60; (d) pitch = −45, fov = 60, heading = 60.

The comparison results described above indicate that the proposed method, which use DeepLabV3+ equipped with Cityscapes pre-trained weights, has higher recognition accuracy and can identify and classify plant pixels from street images. The comparison of computing performance also indicates that the proposed method has greater advantages in computing speed when processing large batches of image data. Lastly, the method proposed to measure GVI using panoramic street view images is closer to the ground truth GVI value than the modified GVI method, which has high accuracy and strong reliability.

4.4.4 Distribution of PVGVI in the study area

Figure 4.11 shows the PVGVI values at all the sampling sites in the study area. The PVGVI values range from 0.019 to 0.729 with a mean of 0.25. To show the PVGVI values of the study area more clearly, this reseaerch used the natural break method (Chen, Yang, Li, Zhang, & Lv, 2013) to separate them into five intervals: 0.067058–0.204220, 0.204221–0.291200, 0.291201–0.375902, 0.375903–0.474319, and 0.474320–0.802120. The change from red to green as shown in the legend of Figure 4.11 represents the change of the GVI value in the study area. From the figure, it was found that the greenery around the streets is not “green” enough, based on the view proposed by Aoki (1991) that most people have a favorable impression of a street landscape when the GVI is at least 0.3 and they will feel the greenery quality is high

enough. However, the figure shows most of the streets with red color which means the PVGVI value of most streets is in the range of 0.067058 – 0.204220, which is lower than the mean value of 0.25, thus it means with better green visible planting needed to be improved on these streets.

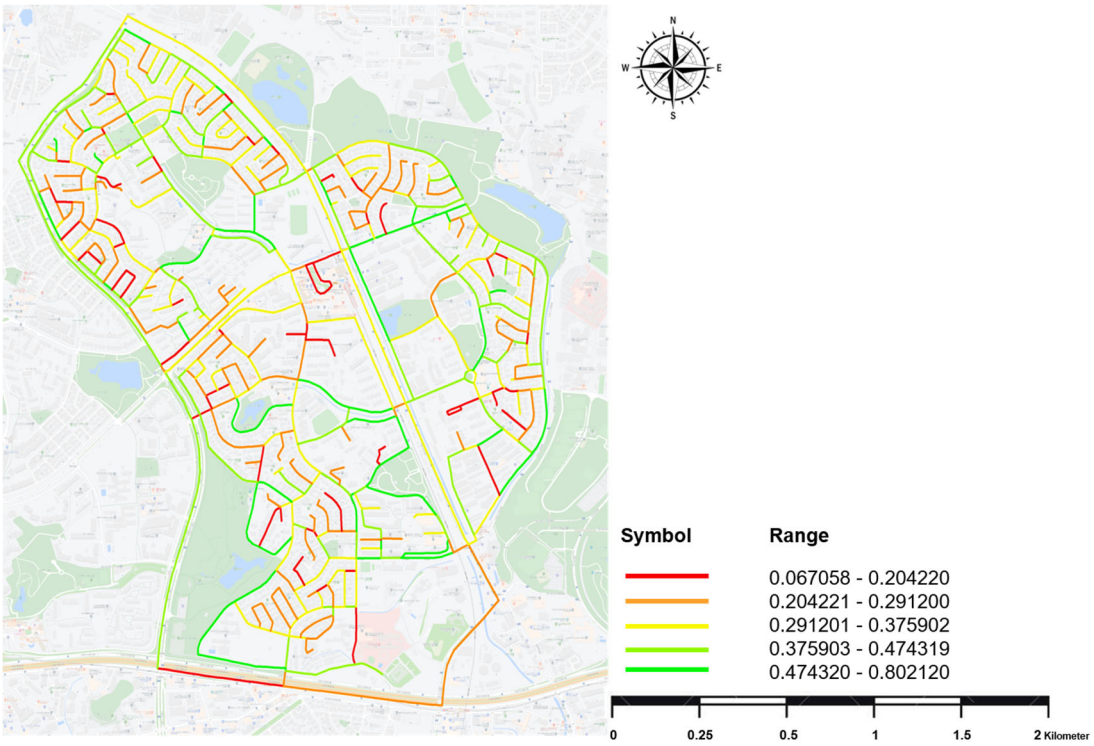


Figure 4.11 Panoramic View Green View index (PVGVI) results calculated using Eq. (3) at sampling points in the study area, overlaid on the road map.

Figure 4.12 shows more directly the frequency distribution of the PVGVI values. As shown, the proportions of the five intervals are 16.5 %, 18.4 %, 21.8 %, 22.2 %, and 21.1 %, respectively. By analyzing the results, it was found that most of the street greenery has PVGVI values greater than 0.3, primarily for the following reasons: more sidewalk trees are planted along main streets, there are many urban parks along the road, and the distant mountains provide a green background that improves the PVGVI values. Meanwhile, the residential area in the north of the study area has lower PVGVI values, mainly because the greenery there is mostly in the form of courtyards surrounded by walls or fences to protect privacy, thereby hiding most of the vegetation and resulting in lower PVGVI values. To increase the visible greenery, landscape designers could use hedgerows and vines instead of walls.

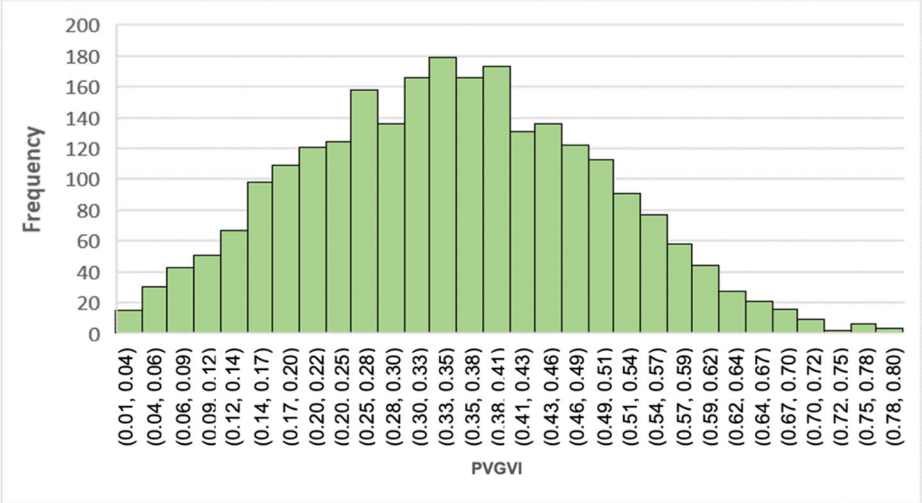


Figure 4.12 PVGVI results based on segmentation.

4.5 Discussion

This research proposed to use street view images to assess street-level greenery in an urban area. The panoramic street view images were taken on the ground and have view angles similar to those of pedestrians, were used for assessing the abundance of street greenery. The research shows that the semantic segmentation model has higher detection accuracy, which can detect almost all vegetation from the street view images to make sure the reliability of the final PVGVI calculation. The proposed PVGVI calculation method should be more suitable for representing the greenery that pedestrians can see on the ground, which can prevent the overestimation of green vegetation results with modified GVI by Li et al. (2015) to make it more rational. The proposed method is a flexible and efficient method, in which many processes can be done automatically. For example, this research does not need to manually take pictures instead of download the street view images by parsing the URL using GSV image API. Then use a semantic segmentation model to automatically extract the greenery and calculate the PVGVI. The method can be used for green space assessment for any place where the street view image is available. As GSV services are extended to more countries and regions, it will be possible to apply this method to more areas to assess visible street greenery. The proposed method for measuring the PVGVI based on GSV images is understandable and easy to use, especially in a large area of a city.

There are a number of areas with urban greenery research that the proposed method can be applied in the future. First of all, the accurate and efficient computing power of this method

has greatly expanded the research area. Furthermore, researchers can apply the PVGVI calculation

in all areas of the research city to get a more comprehensive understanding of the urban street greenery and use it to compare the level of greening between different cities. Second, this research provides a new perspective of using urban street view images to interpret the urban street environment. In future research, the objectives of the research will be expanded to sky visibility, architecture façade, and many other aspects. By extracting information from street view images, city planners can better study the colors of streets, the spatial scale, even the social activities that occur in urban street spaces, to propose strategies to improve and enhance the spatial quality. Third, street view images also continue to stock continuously that will be useful for analysis of changes over time. Therefore, it can be used as a monitoring tool for analysis of gain or loss in urban street greenery and targeted designated greening improvement measures. Thus, it seems to be a useful tool for urban planners and urban environment managers, rather than a simple gadget for users.

However, although the PVGVI values can be calculated based on GSV images and deep learning instead of manual measurements, some urgent concerns remain to be resolved. First, the semantic image segmentation could be more accurate. For example, autumn vegetation tends to lack leaves and exists instead as branches, thereby making it unrecognizable as vegetation. The future work, can include equipping the training dataset with more labeled images and then using it to train the existing model to obtain a more powerful subdivision framework and optimizing the segmentation method.

Second, this research set the distance between adjacent sampling points to 30 m to ensure the availability of GSV images of every street in the study area. However, that distance was 50 or 80 m in other studies (Li, Zhang, Li, Ricard, et al., 2015; Yu, Zhao, Chang, Yuan, & Heng, 2019), therefore future topics to be considered are (i) how that distance affects the greenery assessment and (ii) what that distance should be.

Third, GSV images indicate the urban greenery around the streets at only the time when taken, but plants change with time and look different from one calendar season to the next and between the growing and non-growing seasons. Therefore, in the present work, the time of acquisition is a key factor that impacts the PVGVI calculation accuracy, and one must ensure temporal uniformity of the entire data acquisition and that the PVGVI measurement is not affected by the GSV shooting time. Also, the inability to obtain quickly many GSV images

taken at different locations and times is currently limiting the range of applications in the dynamic monitoring of urban greenery. The good news is that Google and Baidu have both added a new feature known as “Time Machine” (Shet, 2014), which allows users to view images going back to different points in time the Street View feature. This updated version also allows users to view streetscapes at night. This function makes it possible to detect and monitor the dynamic changes of urban street greening, assess the visual impact of urban greening measures, and record the visibility of urban greening. However, such data are currently cannot be accessed via the user API. In the future, the time-series problem in image acquisition may be solved by giving access to the times when street view images were taken. Also, as unmanned aerial vehicle (UAV) technology develops and is gradually applied in urban research, the research might one day incorporate a UAV system to detect and monitor the dynamic changes of street greening for research into urban landscape planning. For some areas that are difficult to obtain street view images, it would be possible to use a panoramic camera combined with a UAV to manually shoot, as a supplement to the street view image dataset.

Lastly, the PVGVI model assessment be proposed in this research is only used as one of the methods for evaluating the street-level greenery visibility. Based on the analysis meant on these data, the urban planners can check the visual impact of some urban forest management practices and document the visibility of urban greenery in cities. In the urban greening program, it is more necessary for urban planners to use these data as the basis to evaluate the urban environment in an all-around way, and rely on a wealth of design methods to achieve the effect of improving the greenery quality.

4.6 Conclusion of this chapter

Visible street greenery is associated with multiple positive health outcomes but is difficult to measure across large expanses. The purpose of this chapter is to develop an approach for estimating and mapping the street-level profile of visible urban greenery and propose the Panoramic View Green View Index (PVGVI) to calculate it. The approach is based on open-access panoramic street view images with view angles similar to those of pedestrians, and a semantic segmentation technique for street feature extraction and is verified by comparing the results of the proposed method with manual statistics and results from the Pyramid Scene Parsing Network (PSPNet), verified the significantly accuracy and efficiency of the proposed approach. The developed approach represents a pedestrian perspective of the visibility of urban street-level greenery that covers complicated urban contexts. Furthermore, this research took

an area near the Suita campus of Osaka University in Japan as a case study, and assess the quantity of greenery and its spatial distribution at the street level. Section 4.4 indicated that the PVGVI is well suited for evaluating street-level greenery. This chapter makes the following contributions. First, compared to the previous GVI calculation methods using multiple images from different directions at a sampling point, the proposed PVGVI calculation method is simplified. Using one panoramic street view image can cover the 360-degree view which was similar to that of a pedestrian. Second, the proposed PVGVI equation is easier to understand and operate, and the results are closer to the visible urban greenery as seen from the ground. Third, the proposed framework based on a pre-trained DeepLabV3+ model has a high accuracy of vegetation detection in urban scenes and can be applied to a large number of cities, which allowed us to find universal laws of street visible greenery. Therefore, urban planners, decision-makers, and sociologists could use PVGVI data as analytical data to better direct urban development and how the urban environment is improved. Future work will focus on improving the accuracy of the semantic segmentation method proposed herein and solving the time-series problem in capturing street view images. Because the method has been used only in GVI calculation and limited case studies to date, it is hoped that future research will involve more evaluation of urban design elements and the development of a system for assessing the urban environment.

Chapter 5

Estimating the Sky View Factor to visualize the built environment openness

The previous chapter introduced an approach that can automatically estimate the PVGVI based on street view images to evaluate the urban greenery quality. However, greening is only one of the indicators used to evaluate the pros and cons of the urban landscape environment. This chapter proposes a method to measure street-level Sky View Factor (SVF) based on semantic segmentation processing to extract sky area data from street view images and estimate the fisheye photographic-based sky view factor (SVF_f). The sky view factor (SVF) has been recognized as an indicator to evaluate the openness of streets in the field of urban planning. It represents the ratio of the visible sky area to the total sky area at one point in space. However, due to the time-consuming and laborious acquisition of data and manual detection in traditional measurement methods, the SVF measurement in large-scale space has been greatly restricted. With the development of street view images (SVIs), some SVI services provide panorama data of the urban street level that can be used to estimate the SVF. This subsection shows the reliability and efficiency of the proposed SVF value estimated method by comparing it with the previous research. The generated street-level SVF_f maps based on estimation results can be served as a design base for creating more comfortable pedestrian street spaces. The method proposed in this research provides a more accurate and convenient approach to evaluate the

openness of the built-up environment of the urban landscape. The following section first outlines the development system, then describes the methodology in detail, and selects an urban residential block as a case to verify the evaluation system. Finally, the experimental results and the limitations of this method are discussed, and the conclusions are summarized.

5.1 Introduction

With the global urbanization process, the urban heat island effect has appeared in major cities in the world, and the corresponding urban climate and air quality have also shown signs of deterioration. How to alleviate the urban heat island effect is the research focus of many scientists and urban planners, and it is also a climate and environmental issue that all countries are concerned about. Thermal comfort is a key indicator for evaluating the quality of urban space, human comfort, and micro-ecology at the urban street level (Huang et al., 2015; Yang et al., 2015a). There have been many studies on the relationship between urban thermal environment and urban form. Some studies have further proved that urban spatial morphology will impact the urban microclimate environment (Algeciras et al., 2016; Martinelli and Matzarakis, 2017). In 1981, Oke first proposed the concept of the Sky View Factor (SVF) to evaluate the heat island effect in cities (Oke, 1981). It was then denoted as one of the main topics for discussing the urban microclimate, air pollution, and urban thermal comfort (Venhari et al., 2019; He et al., 2015; Johansson, 2006; Bourbia and Boucheriba, 2010). In these studies, SVF represents the proportion of the radiation received by the plane from the sky in the total environmental radiation (Krüger et al., 2011; Watson and Johnson, 1987).

This research used street view images (SVIs) based on deep learning to estimate sky area and to calculate SVF values. The sky areas were extracted from panoramic SVIs based on the deep learning model and generated fisheye (hemispherical) images, then calculated SVF values automatically. Because there are abundant SVI data, this research can quantify street-level SVF values in urban areas using the proposed method.

5.2 Proposed methodology

This research proposed a workflow following the four steps below to estimate SVF based on the fisheye photographic method (see Figure 5.1). The first step was to search through SVIs and generate panoramic SVIs. Then, semantic segmentation was applied to the panoramas to classify the sky areas. The third step was to generate classified fisheye images through

hemispherical transformation. Finally, SVF_f was calculated based on the classified fisheye sky images.

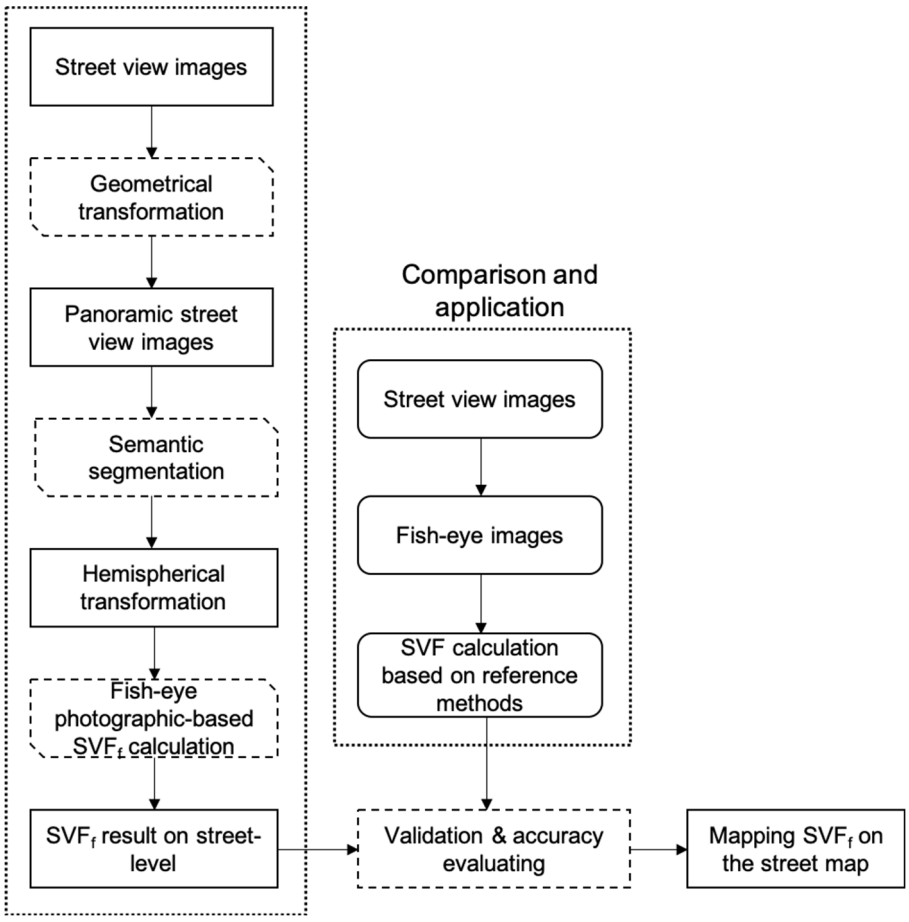
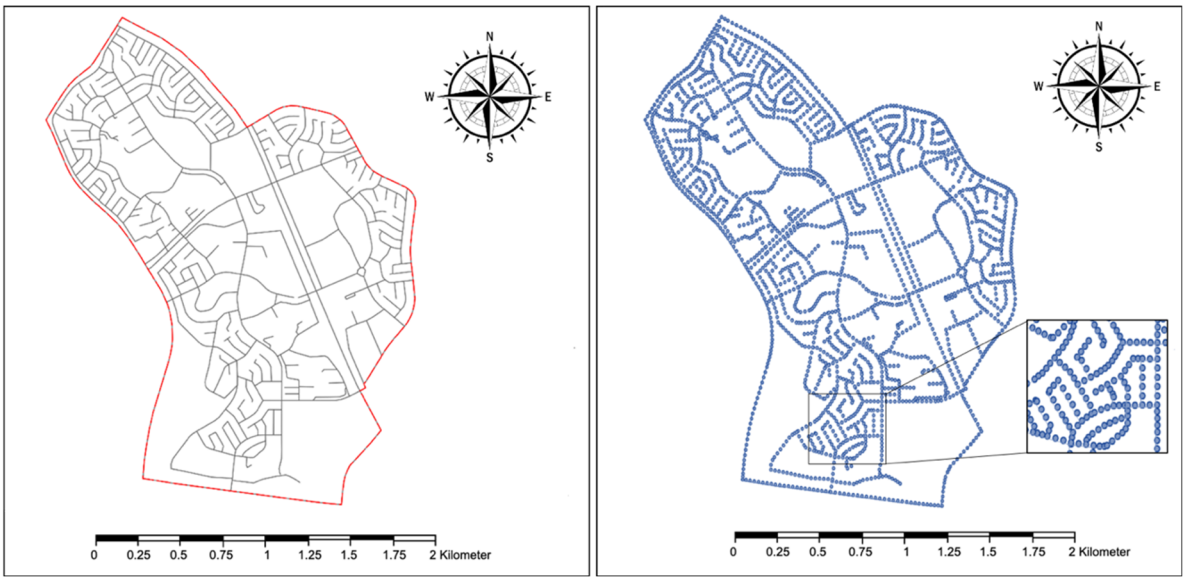


Figure 5.1 Workflow chart of the methodology proposed in this study.

5.2.1 Panoramic SVI collection

This research obtained the street network of the study area from the Open Street Map (OSM) website (Open Street Map (OSM), 2021) and set sampling points along each road at 30-meter intervals as shown in Figure 5.2b. In total, 2,492 sampling points with coordinates were determined within the study area and the SVIs of these sample points were downloaded.



(a) (b)

Figure 5.2 (a) The street map and (b) sampling points distribution map of the study area.

Google (Google Street View (GSV), 2021), Baidu (Baidu Street View (BSV), 2021), and Tencent (Tencent Street View (TSV), 2021) have provided public application programming interface (API) services that enable users to request and download static SVIs by using the corresponding APIs. This research chooses to use the GSV static API to search through crawling SVIs. Usage examples of the API parameters for crawling GSV are shown in Table 5.1. The required data includes the image size, sampling point location information, pitch and heading angles, and the API key. Since the objective of this research is to measure the visibility of the sky area, the SVIs must contain the whole sky area, so the pitch value was set as 45 to ensure the complete sky area can be obtained. Most of the images were captured in the spring and summer of 2018–2019, which effectively avoided the lack of leaves in autumn and winter that affect the calculation accuracy of SVF.

Table 5.1 API parameters for crawling GSV.

Parameter	Description	Example
Size	The output size of the image in pixels	size = 500×500 returns an image that is 500 pixels wide and 500 pixels high
Location	Coordinates of GSV location	location = 34.80932445, 135.5066877
Heading	Compass heading of camera; accepted values are 0–360	North: heading = 0 (360) South: heading = 180 East: heading = 90 West: heading = 270
Fov	Horizontal field of view of the image	fov = 90
Pitch	Up or down angle of the camera relative to GSV vehicle	pitch = 45
Key	Developer's key (retrieved through online application)	key = your API key

The previous research used tools such as Hugin (d'Angelo, 2007), PTGui, or coordination transformation to stitch the panoramic SVIs. Figure 5.3 shows a generated example of a fisheye image using in this research to measure SVF_f, which involved searching through GSV images, generating panoramic GSV images, and generating fisheye images based on hemispherical transformation.

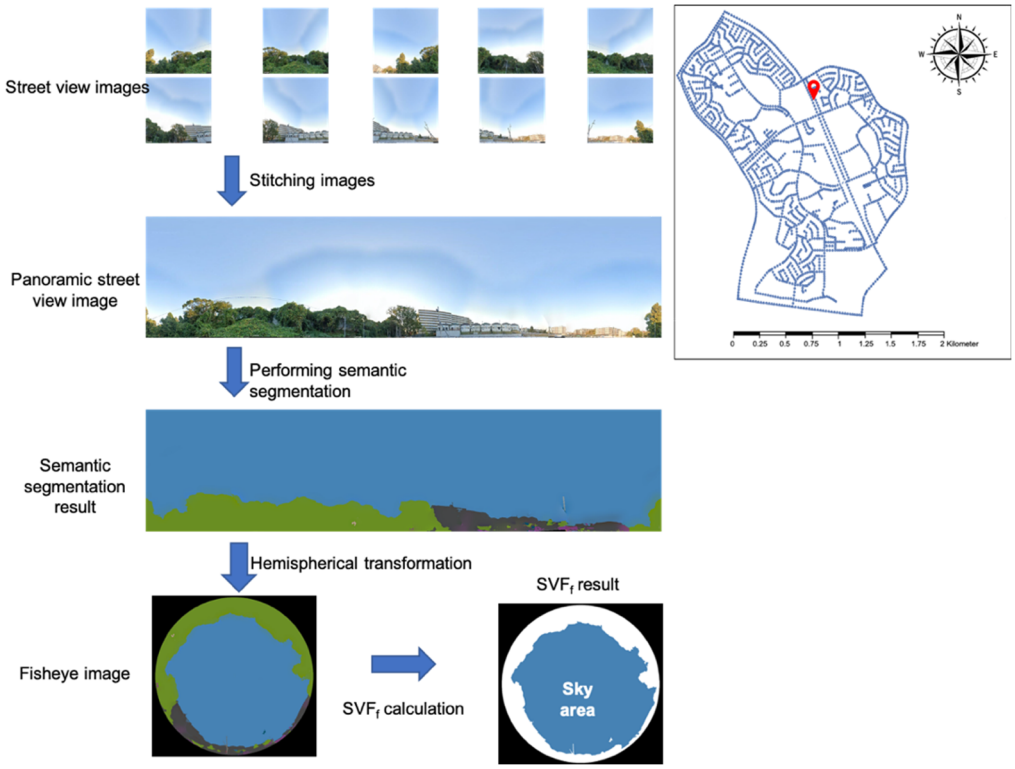


Figure 5.3 The location of the sample point and the example of extracting SVFf from SVIs.

5.2.2 Fisheye images generated based on a hemispherical transformation

As mentioned in the literature review, fisheye images were required when calculating SVF using the standard photography method. In previous studies, fisheye lenses are usually used to manually capture fisheye images. This research automatically generated fisheye images from panoramic street view images through the conversion of cylindrical projection and azimuth projection. Through the geometric model shown in Figure 5.4, the cylindrical projection of the panoramic image is converted into an azimuthal projection (Li et al., 2018). This projection is achieved by establishing a corresponding relationship between the pixels (xa, ya) on the fisheye image and the (xb, yb) on the panoramic street view image, as shown in Equations (7) and (8):

$$x_b = \begin{cases} (\pi/2 + \tan^{-1}[(y_a - f_y)/(x_a - f_x)]) \times W_b/2\pi, & x_a < f_x \\ (3\pi/2 + \tan^{-1}[(y_a - f_y)/(x_a - f_x)]) \times W_b/2\pi, & x_a > f_x \end{cases} \quad (7)$$

$$y_b = \left(\sqrt{(x_a - f_x)^2 + (y_a - f_y)^2} \Big/ r_0 \right) \times H_b \quad (8)$$

where H_b and W_b are the height and width of the panoramic street view image, r_0 is the radius of the fisheye image, and (f_x, f_y) are the coordinates of the center pixel on the fisheye image; the corresponding relationship is shown in Figure 5.4.

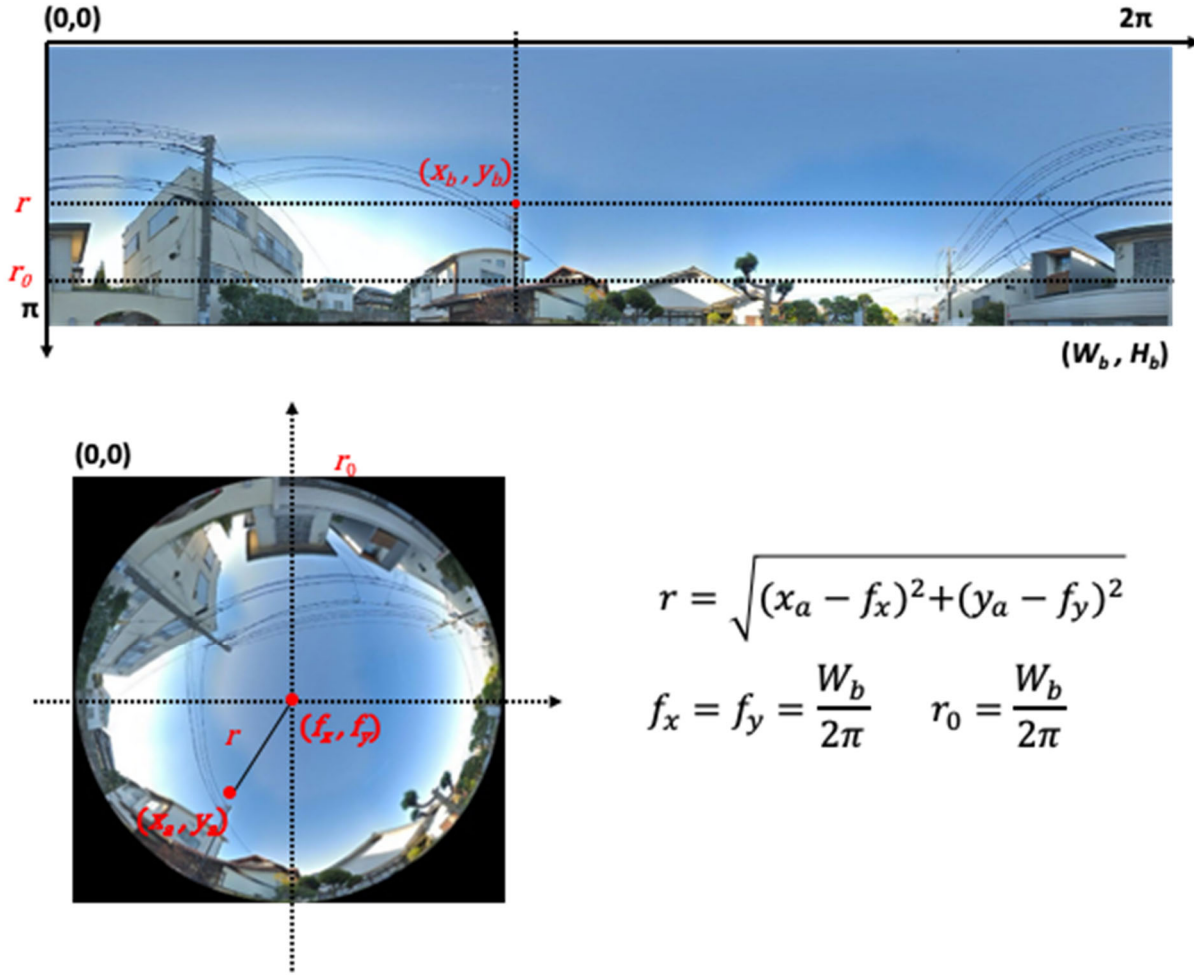


Figure 5.4 Geometric model for transformation of a panoramic street view image to a fisheye image.

By using Equations (7) and (8), each pixel on the panoramic street view image and the fisheye image can be connected. Repeating the process for each pixel, fisheye images can be generated from panoramic street view images.

5.2.3 Sky area extraction based on deep learning

This research proposed to use semantic segmentation to automatically extract sky regions from street view images. Semantic segmentation is a very important field in computer vision. It uses machine learning to understand and segment image content semantically and predict an image in form of pixel-level, each pixel is classified into a specific category. The semantic segmentation model used in this paper is DeepLabV3+, which was developed by Chen et al. (Chen et al., 2018). It extends DeepLabV3 by introducing an encoder-decoder structure which is commonly used in semantic segmentation. And unlike most encoder-decoder architecture

designs, DeepLabV3 provides a distinctive semantic segmentation method. DeepLabV3 proposes an architecture for controlling signal extraction and learning multi-scale context features. The two main reasons that this research chose to use this model were (1) that DeepLabV3+ has an excellent performance in the analysis of urban scenes, and has high accuracy in identifying the sky, vegetation and buildings in SVIs, and (2) in some major performance evaluation competitions, DeepLabV3+ has demonstrated a better performance than other semantic segmentation networks (e.g., U-Net, SegNet) as described in the previous paper (Xia et al., 2021).

5.2.4 SVF_f calculation

The definition of SVF in (Johnson and Watson, 1984) is "the ratio of the radiation received (or emitted) by the planar surface from the sky to the radiation received (or emitted) by the entire hemispheric environment", as shown in Figure 5.5a. The original SVF is defined by Steyn (Steyn, 1980), who proposed an equation that can be used to obtain a geometrically corrected fish-eye image, as shown in Equation (9):

$$\mathrm{SVF} = \frac{1}{\pi {r_1}^2} \int_S dS \qquad (9)$$

where S is the area of the circular sky area projected on the ground, and r_1 is the radius of the hemispheric radiating environment. The dS can be defined using Equation (10) below:

$$dS = \frac{\pi r_3}{2} \sin\left(\frac{\pi r_2}{2r_3}\right) \cos\left(\frac{\pi r_2}{2r_3}\right) dr_2 d\alpha \qquad (10)$$

where (r_2, α) are polar coordinates defining dS on the equiangular projection, and r_3 is the radius of the horizontal image on the print.

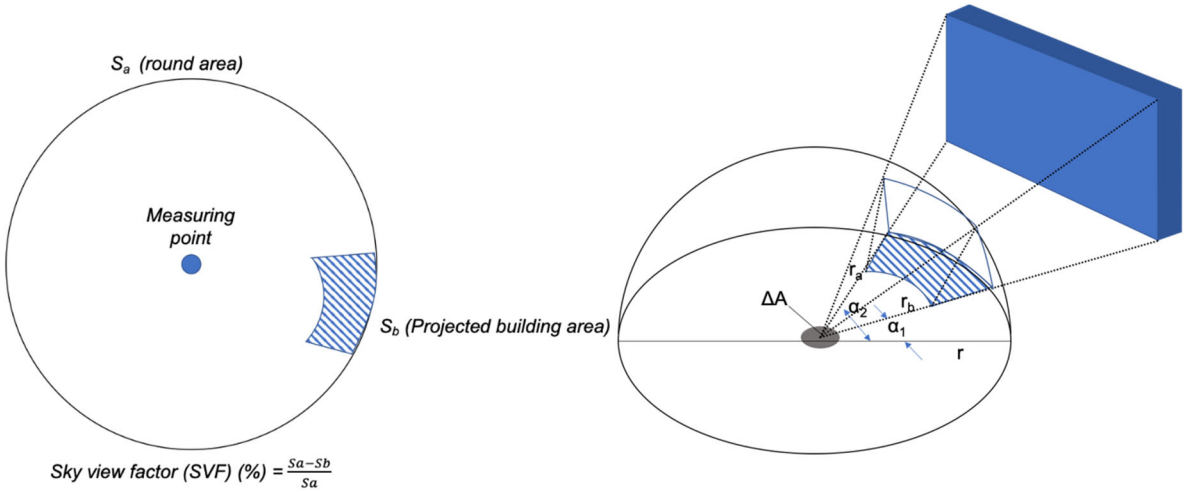
Steyn (1980) introduced the fish-eye photographic method to urban climatology research. He used two steps to calculate the SVF, through dividing the fish-eye image into m concentric annulus of equal width and summing up all annulus sections representing the visible sky, as shown in Equation (11):

$$\mathrm{SVF} = \frac{1}{2m} \sum_{m=1}^{m} \sin\left(\frac{\pi(i-1/2)}{2m}\right) \cos\left(\frac{\pi(i-1/2)}{2m}\right) \alpha_i \qquad (11)$$

This method was further improved by Johnson and Watson (Miao et al., 2020) (see Figure 5.5b), as shown in Equation (12):

$$\mathrm{SVF} = \frac{1}{2m}\sin\left(\frac{\pi}{2m}\right)\sum_{i=1}^{m}\sin\left(\frac{\pi(2i-1)}{2m}\right)\alpha_i \qquad (12)$$

where m is the number of annuli, i is the annulus index, and α_i is the angular width of the ith annulus (Figure 5.5).



(a) Definition of sky view factor

(b) Projection of the wall on the circular plate through a fisheye lens

Figure 5.5 Definition of SVF and its application in previous urban climate research.

SVF also has many applications in architecture. A conventional SVF estimation method for the built environment is to analyze the fisheye image taken from a specific viewpoint to use image processing software to identify the sky area in the photo next, and to calculate the SVF finally. The SVF result estimated by this method has higher accuracy, it is easier to operate and applied in urban research. According to the definition of SVF and the previous calculation method of SVF, the research proposed to apply the fisheye photographic-based method to calculate the $\mathrm{SVF_f}$, as shown in Equation (13) (see Figure 5.6), which has also been applied and proved effective by Cao et al. (Cao et al., 2019).

$$SVF_f = \frac{Area_{s\_i}}{Area_{t\_i}} \times \frac{4}{\pi} \qquad (13)$$

where $Area_{s\_i}$ refers to the sky area pixels in the image taken in the ith sampling point, and $Area_{t\_i}$ refers to the total pixels of the image taken in the ith sampling point.

Figure 5.6 Equation for calculating SVF_f.

5.3 Experiments and results

This section introduces the extraction results of the sky area and SVF_f estimation based on the semantic segmentation network. Then, the SVF_f estimation results are compared with those from two independent methods: one is to use the image processing software Photoshop to manually detected and marked the sky area in the hemispherical photographic image and calculate SVF (SVF_m) , another is using the U-Net scene parsing deep learning model to extract the sky area from the image and calculate SVF (SVF_u) to assess the accuracy. Through these two sets of comparisons, it can be verified that the proposed method has higher accuracy for sky area detection and SVF estimation. Furthermore, the estimated SVF values were mapped to the street map of the study area to reflect the sky visibility and spatial openness in this area more intuitively.

5.3.1 SVF_f estimation results based on deep learning

Figure 5.7 shows the visualization recognition result of semantic segmentation. The different color markers represent the different elements in SVIs that have been identified and classified.

Figure 5.7 Example of the semantic segmentation result.

5.3.2 Accuracy assessment of the SVIs-based SVF_f estimations

100 points was randomly selected from the study area to assess the accuracy of the SVIs-based SVF_f estimation. Comparing the estimated SVF_f value with the manually measured SVF value and the SVF value estimated based on the U-Net method can objectively verify the accuracy of the proposed SVF_f estimation method based on SVIs. At the same time, the correlation between these methods can also be reflected. This research used the U-Net-based SVF_u estimate method (Cao et al., 2019), which has been confirmed in previous studies to perform well in extracting sky areas, to verify the applicability of the proposed SVF_f estimate method.

Figure 5.8 shows three examples of the transformed fisheye images (a), (b), and (c), which show the proposed fisheye photographic-based SVF_f and U-Net-based SVF_u; (d) shows the SVF_m based on manual extraction for reference. The corresponding SVF values are also shown. The result of manual extraction is consistent with the fisheye SVI-based SVF_f, and the difference is within 0.03. However, there is a significant difference between the U-Net-based SVF_u and the SVF_m, which are mainly affected by the color of buildings in the environment. It was found that, by using the U-Net-based method to extract sky areas, lightly colored surrounding buildings were often mistaken as the sky, which will lead to higher SVF values than those of the ground truth.

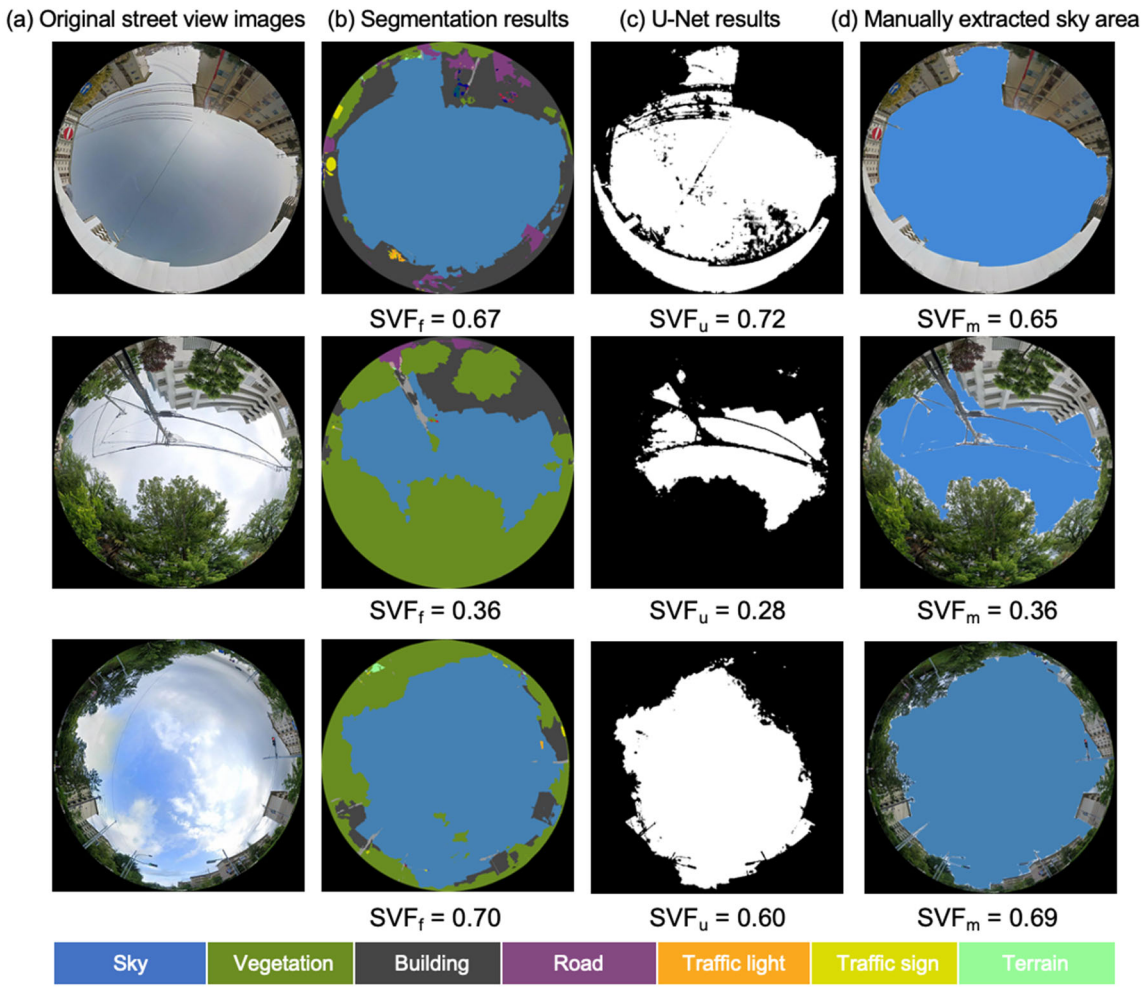
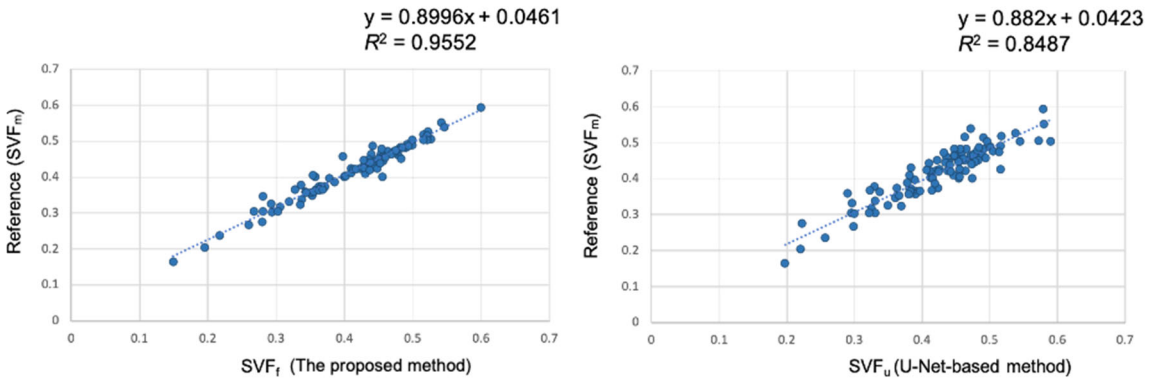


Figure 5.8 Sky extraction results and reference.

Figure 5.9 shows scatter plots of SVF_m reference data from manual extraction and the proposed fisheye SVI-based SVF_f and the U-Net-based SVF_u estimates. Figure 5.9a shows the correlation between reference data and the proposed fisheye SVIs-based method (R = 0.9777, R^2 = 0.9552), which indicates these two sets of data have a tight correlation. Figure 5.9b shows the correlation between the reference SVF_m and the U-Net-based SVF_u (R = 0.9212, R^2 = 0.8487), which indicates the correlation between these two sets of data is weaker than the previous two sets of data. This means that the proposed approach is more suitable for evaluating the visibility of sky areas in urban environments.



(a) (b)

Figure 5.9 Evaluation of the accuracy of the proposed method: "SVF_f" = SVF estimated using the pre-trained DeepLavV3+ model; "SVF_m" = SVF estimated manually using Adobe Photoshop; "SVF_u" = SVF estimated using the U-Net-based method.

Here, this research also used the following three main evaluation indicators to assess the accuracy of the semantic segmentation model used to detect the sky area in the SVIs. As Table 5.2 shows, intersection over union (IoU) was used to assess the accuracy of sky pixels position detection, root-mean-square error (RMSE) was used to assess the overall accuracy of the SVF value estimation, and mean absolute error (MAE) was used to evaluate the actual situation of the predicted value error. As the result shows, the proposed method had higher IoU and lower RMSE and MAE, which means that compared with the U-Net-based method, it can extract the sky area more accurately and obtain a more accurate SVF estimation.

Table 5.2 Comparison of the related evaluation metrics for the proposed method and U-Net-based method

Method	IoU [%]	RMSE [%]	MAE [%]
Fisheye photographic-based SVF_f estimation method	88.29	1.82	1.17
U-Net-based SVF_u estimation method	86.09	3.13	2.50

It can be seen from the above comparison results, the semantic segmentation method proposed in this research has significantly higher accuracy than the U-Net-based method in the detection of the sky area in the street view image.

5.3.3 Mapping SVF_f on the street map of the study area

Figure 5.10 shows the distributions result of SVI-based SVF_f estimation. The SVF_f values range from 0.000 to 0.669, and the mean is 0.365. This research used the natural break method (Chen et al., 2013) to divide them into five intervals: 0.000–0.274, 0.274–0.376, 0.376–0.458, 0.458–0.534, and 0.534–0.669. The five SVF_f value ranges are represented by different colors from yellow to dark blue, as shown in the legend. The SVF_f value ranges near 0 indicate that little sky area can be viewed, while the opposite is seen in the value ranges near 1.0 that indicating total sky openness. These SVF_f values are related to the height and density of buildings. Areas with higher building density have lower SVF_f values. Conversely, the lower the density, the higher the SVF_f value. It is clear that narrow street canyons formed by high-density buildings hinder the visibility of the sky, resulting in low SVF_f values.

It was found that in the southwest and northwest areas of the study area, SVF_f values are generally higher. Because the buildings in this area are mostly two- or three-story single-family detached homes, most of the plants are lawns and shrubs, forming an open urban canyon space. On the contrary, the height of the buildings in the center of the study area is generally high, and the sidewalk plants are mainly arbor, which forms the narrow street canyons and led to the low SVF_f.

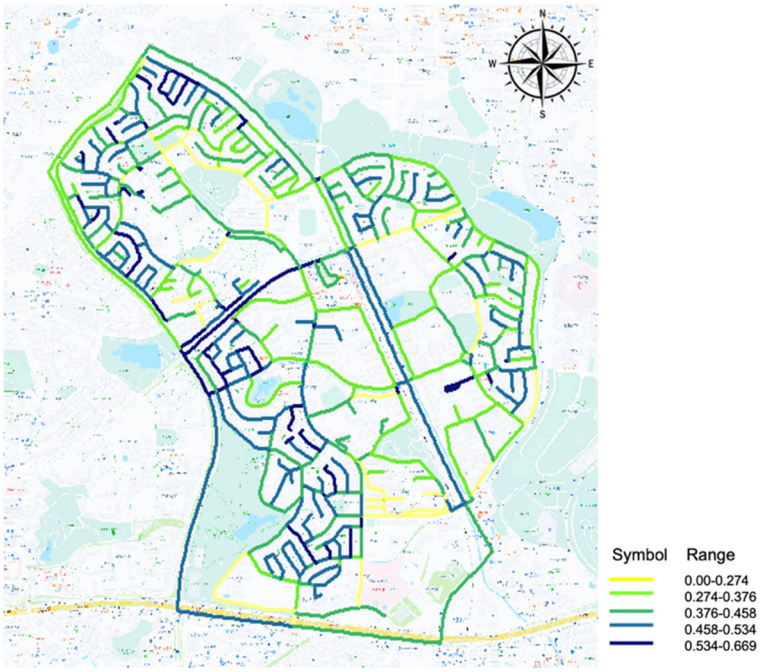


Figure 5.10 Mapping the SVF_f values estimated by Equation (12) on the road map.

Figure 5.11 shows the frequency of SVF_f value distribution in the study area. The proportions of the five intervals are 20.7%, 29.5%, 28.4%, 16.9%, and 4.5%. The analysis of the results shows that SVF_f is uniformly distributed in the two intervals of 0.274–0.376 and 0.376–0.458.

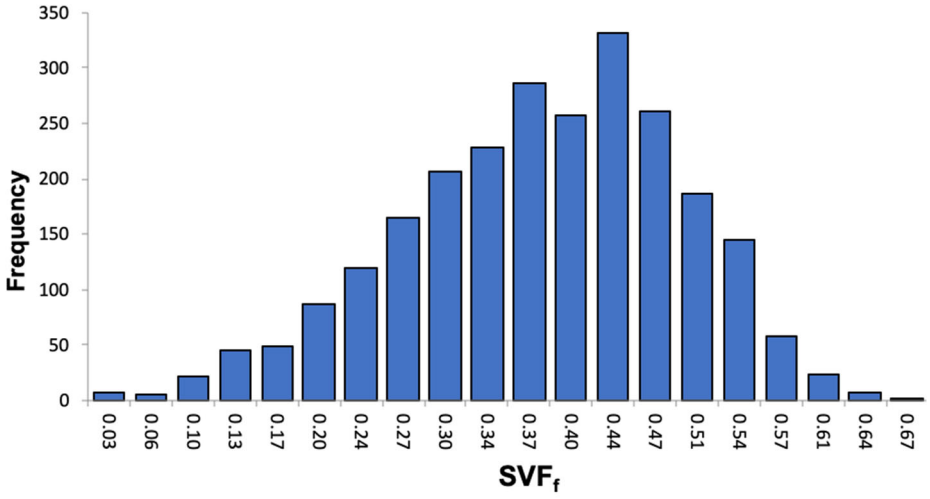


Figure 5.11 Frequency distribution histogram of SVI-based SVF_f estimates.

5.3.4 Computation performance

In large-scale SVF measurements of cities, the research needs to analyze big data, which means that computation performance is another important indicator in this study. Semantic segmentation classification of SVIs is the most time-consuming part of the proposed methodology. The analyses were all run on a Windows PC with an Intel® Core™ i7-8086K processer at 4.00 GHz and NVIDIA GeForce RTX 2080Ti. This research ran DeepLabV3+ in the GPU (CUDA) mode to perform semantic segmentation on the 2,492 panoramic SVIs, with each image consisting of 2,000 × 500 pixels. Semantic segmentation processing took about 12 min to complete and it almost spent about 0.3 s per image. In a previous study (Yin and Wang, 2016), researchers spent about 2 days classifying 3,592 panoramic SVIs (each image consisted of 416 × 254 pixels) based on a support vector machine (SVM) machine learning algorithm. Also, Liang et al. used SegNet to do the classification, and the processing time to classify the panorama SVIs was about 1 s per image (1,024 × 1,024 pixels) in GPU mode (Liang et al., 2017). These results show that the proposed method has higher accuracy in image information extraction capabilities and higher efficiency in processing big data, and it is more suitable for application in large-scale urban-related research.

5.4 Discussions

A method for automatically estimating SVF value by acquiring open-source SVIs and using the deep learning framework DeepLabV3+ to semantically segment and classify SVIs was present in this chapter. The proposed fish-eye photographic-based SVF_f estimation method saves a lot of time-consuming and laborious processes of on-site investigation and manual identification. This research proposed a framework to classify panoramic SVIs based on semantic segmentation to automatically extract sky area, then convert the images to fisheye images, and calculate SVF_f values. By comparing the verification results against reference data, the proposed method showed good agreement with the ground truth data. The results showed that this proposed SVF_f estimation method based on panoramic SVIs can realize fast and accurate SVF_f automatic calculation. For urban planners and designers, the proposed method is straightforward, feasible, and effective. By accessing public street view image data, fully automated SVF_f estimation is achievable. Compared with the traditional method based on manual fisheye image photography, the proposed method is easier to operate. Compared with simulation methods, the proposed method based on panoramic SVI can represent more realistic street conditions. It can avoid the inaccurate SVF_f estimation results caused by the lack of real street tree information and the uncertainty of model simulation in the simulation method.

More importantly, because the proposed method is based on computer processing, it can be applied to large-scale SVF_f estimation. As the area covered by SVIs grows larger, the proposed SVF_f estimation method based on SVIs can be used in any area with street view image data sources. The high accuracy of the SVF_f estimation method can also help us understand the urban thermal environment and strongly support global studies of the urban thermal environment.

However, the limitations of some existing methods still need to be addressed in future research. Firstly, this study is based on panoramic SVIs for calculating SVF_f, and all the SVIs were taken by the street view capture vehicle in the center of the road, which cannot represent the visible sky area of the pedestrian view. So, this research will be more suitable for the research that focuses on the visibility of sky areas at the street level. Secondly, owing to the continuous renovation of cities, plants and buildings in street spaces are dynamically changing, but the shooting time of the SVI is fixed. Because of the time difference between the acquisition time of the SVI and the actual situation, the estimation accuracy of the SVF_f value will be affected.

Based on the advantages and limitations discussed above, future studies should focus on the following: (1) applying the proposed method to different areas and cities to verify that the SVI-based SVF_f estimation method can be used as a general method and can accurately reflect urban street-level sky visibility and the openness of the street space; (2) while considering the phenology of vegetation and the surroundings of urban streets, estimate the average value of SVF_f by searching through SVIs during each season. However, the future research can obtain real-time SVI by using unmanned aerial vehicle (UAV) photography technology to estimate SVF_f more accurately; (3) considering that temperature is also an important indicator for evaluating environmental comfort, the proposed method can provide a data analysis basis for urban thermal environment improvement and urban design.

5.5 Conclusion of this chapter

This chapter proposes a method to automatically estimate SVF_f values based on panoramic SVIs. This method involves the use of a pre-trained DeepLabV3+ deep learning model to perform semantic segmentation of SVIs, then generates fisheye images to automatically calculate the SVF_f based on the result. By comparing the result of semantic segmentation recognition with the results of manually labeled sky area, the recognition rate of the sky area in SVI by the proposed method is as high as 98.62%. The comparison results of calculation performance also verify the reliability of this new method could reach 88.29% when compared with the U-Net-based method.

Besides, the proposed SVIs-based SVF_f method uses publicly accessible street view pictures as the data source to estimate the city street-level SVF. This research shows that with the development of wearable equipment and unmanned aerial vehicle photography technology, more urban areas will have street view image data, and more real-time street view image resources will also appear. Then, the future research will be able to use these big data resources to obtain more efficient and accurate assessment results of the urban environment. Furthermore, the future research can rely on these data to build a city information data platform to help planners and managers better urban planning. In future research on urban thermal environment, by combining the semantic segmentation of street view images with the Local Climate Zones (LCZ) classification system (Stewart and Oke, 2012), the urban heat island intensity can be calculated more accurately, making up for the insufficient distribution of the number of ground meteorological observation points, and providing a more comprehensive urban heat island intensity analyze data.

#

Chapter 6

Conclusions

This dissertation has presented the development of deep learning-based methods for urban landscape analysis using SVIs. Chapter 1 introduced the background information, problems statements, research objectives, research significance, and research scope was introduced in Chapter1. Chapter 2 reviewed the literature on urban landscape analysis elements, computer vision, deep learning, street view images, and the application of street view images in urban environment analysis. Chapter 3 present an image semantic segmentation model for urban elements detection. Subsequently, a comparative study to examine the performance of this semantic segmentation algorithm was conducted. Chapter 4 proposed an approach to automatically detect the vegetation pixels from SVIs to calculate the visible greenery of street-level. Chapter 5 extended the proposed method from Chapter 4 to estimate the openness of the built environment using SVF. This chapter described the summary, contributions, and limitations of this research. Furthermore, the future works are discussed at the end.

6.1 Summary

People-oriented has always been the core concern of urban planning, but in the past, due to the limitations of technology and data, people-oriented means that it must rely on manual analysis and subjective experience of experts, and it is difficult to apply it to practice on a large scale

and quickly. In response to this problem, this research proposes an efficient and fast spatial quality quantitative measurement framework based on open data.

Street view images have become a new data source for urban research. Extracting relevant data from street view images for analysis has also become a new trend in urban environmental research. However, the research on urban environment evaluation based on street view images is still in the early stage. On the one hand, the development of computer vision technology represented by deep learning has made it possible to recognize patterns that cannot be expressed explicitly. The fusion of urban analysis application research data represented by street view images, satellite remote sensing images, and geotagged social media images poses a challenge to the development of new theories and methods in urban environmental assessment research.

This research aims to develop a deep learning-based analysis system, which can automatically analyze the urban landscape environment using street view images. First, the captured city street view images are processed in batches using Python code to obtain 360-degree panoramic street view images and fish-eye images. Then, these street view images are semantically segmented through a pre-trained deep learning model. And the environmental quality evaluation indicators such as visible greenery and the built environment openness are estimated from the image semantic segmentation results. Finally, the corresponding relationship is formed by mapping the position of each street view image to the relevant index and visualizing these factors which quantify the urban spatial quality. The researhc choosed two key urban environment elements as the starting points of the research: the street-level visible greenery and the visible sky area of the built environment.

The first method is the urban street-level visible greenery evaluation method (PVGVI method) presented in Chapter 4. The PVGVI method can evaluate the street-level greenery using panoramic street view images. The calculation and evaluation system were developed, and the proposed method was validated with a case study. The PVGVI method effectively increases the accuracy of the vegetation detection from the street view images when compared to the PSPNet-based semantic segmentation method. The prototype system takes only a few seconds to execute the semantic segmentation on street view images. Therefore, this method can significantly improve efficiency and save time when performing large-scale urban environmental analysis. This method provides new ideas for evaluating other factors that affect

the urban environment. The future research can extract the other urban landscape elements from street scene pictures for analysis in the same way.

The second method is the urban built environment Sky View Factor estimation method (SVF_f method) presented in Chapter 5. A new perspective for evaluating the openness of the built environment is proposed, which is to estimate the sky view factor from the street view image. The SVF_f estimation system was developed, and a case study was used to verify the system. The SVF_f method calculated SVF by converting panoramic street view images into fisheye images. Furthermore, this research pre-trained the DeepLabV3+ model by manually marking 300 fisheye images of the sky area as the training set, which improved the recognition accuracy of fisheye images based deep learning. At the same time, the GPU-based semantic segmentation model also improved the processing efficiency of a single image to 0.1second. However, in the process of generating the fisheye image from the panoramic image, the distortion of the figure will be generated, which may affect the accuracy of the final measured SVF_f. Especially when the color of the building environment and the sky in the street view image are similar, the fisheye image generated by the proposed method is more likely to be deformed.

6.2 Research Contributions

Based on the street view images and deep learning algorithm, this study developed an automatic evaluation system for urban landscape analysis, showing the development direction of combining artificial intelligence with urban landscape research, which has inspired more aspects of urban research.

The research achieves a high-precision analysis of large-scale panoramic graphics with the support of open-source street view images and artificial intelligence technology. It can still ensure high-precision and high-efficiency processing capabilities when carrying out urban-scale spatial analysis, which can be used to reveal potential laws that were not discovered in the past due to analysis scales or analysis accuracy.

The PVGVI method presents a new approach for SVIs-based urban street-level greenery analysis. The SVF_f method presents an approach to automatically estimate the visible sky area in the urban built environment which can be used to evaluate the openness. All the proposed methods in this research can extract the urban landscape elements from the street visible environment in much less time than the manual site survey and manual calculation methods.

The time cost, error and money investment due to manual operation can be reduced. The proposed calculation and evaluation approaches can be further extended to analysis other urban landscape elements, such as the building style, architectural façade and the street walkability. However, the current street view image recognition and classification based on semantic segmentation technology is still very limited. With the continuous updating of deep learning models, the accuracy of image semantic segmentation is also continuously improving. At the same time, the improvement of computer performance also makes image processing more efficient. The two proposed methods in this research are both general methods that can be applied in any city.

6.3 Limitations and future research

The proposed methods still have room for improvement, and these limitations should be addressed in future research. The limitations of each method are highlighted as follows:

The limitations of the PVGVI method are that, firstly, the shooting time of street view images is not uniform. In some street view images taken in late autumn or winter, plants are easily confused with light poles or architectural structures in the surrounding environment due to the fallen leaves, which affects the accuracy of semantic segmentation results. Secondly, the developed methodology was tested on only the street view images of the study area, although the concept of utilizing semantic segmentation to segment the urban environment elements from street view images is very general and applicable to extract many kinds of urban environment elements. However, if the training data set for the deep learning algorithm is sufficiently rich, it should be further verified in other cities or regions to improve its reliability and generalization. The Green View Index is just one of the urban greenery evaluation factors and is not enough to assess the whole urban greenery quality. In future research, the quality of urban green space can be evaluated more comprehensively by measuring the following three greenery-related evaluation indicators. Urban green coverage ratio can be measured by analyzing satellite images, urban street plant diversity can be evaluated by using image detection techniques, and the visible green level of the street level can be obtained by applying the method proposed in this research.

The SVF_f method by generating fish-eye images from panoramic street view images will inevitably cause distortion of fish-eye images during the generation process, and then affect the accuracy of SVF evaluation.

Furthermore, the street view images are usually captured by street view collection vehicles on the roadway, however, the spaces where people stay are more in the public spaces and pedestrian spaces. Therefore, there is a deviation between the viewpoint of the obtained street view image and the viewpoint of an actual person. Future research should explore new paths to obtain street view images that can represent the perspective of pedestrians. For example, the street view image can be manually taken with wearable devices and miniature photography equipment or obtained from social platforms using various APIs. These images can be used as supplementary data to more realistically reflect the greening level of urban streets in the eyes of pedestrians.

Second, because the urban built environment is in a constantly changing state, the growth state of street plants and the construction sequence of the city will all affect the measurement results of sky visibility. Therefore, not only the evaluation of SVF but also the evaluation of the visible greenery should be a dynamic process, which is continuously updated following the construction of the urban environment.

Finally, limited by the existing street view image resolution, deep learning algorithms, computer computing power, etc., the proposed method cannot extract more analysis elements that affect urban environmental quality assessment from street view images, such as city skylines, architectural styles, and plant richness, etc. The future research should focus on improving the detection and classification accuracy of the deep learning model by creating a training set of street view images that contains more abundant city scene tags. And, the accuracy and efficiency of all methods could be improved by further case studies and various performance improvements of the deep learning model.

References

Algeciras, J. A. R., Consuegra, L. G., & Matzarakis, A. (2016). Spatial-temporal study on the effects of urban street configurations on human thermal comfort in the world heritage city of Camagüey-Cuba. *Building and Environment*, *101*, 85-101.

An, S. M., Kim, B. S., Lee, H. Y., Kim, C. H., Yi, C. Y., Eum, J. H., & Woo, J. H. (2014). Three-dimensional point cloud-based sky view factor analysis in complex urban settings. *International Journal of Climatology*, *34*(8), 2685-2701.

Aoki, Y., Yasuoka, Y., & Naito, M. (1985). Assessing the impression of street-side greenery. *Landscape Research*, *10*(1), 9-13.

Aoki, Y. (1987). Relationship between perceived greenery and width of visual fields. *J. Jpn. Inst. of Landscape Architects*, *51*(1), 1-10.

Aoki, Y. (1991). Evaluation methods for landscapes with greenery. *Landscape Research*, *16*(3), 3-6.

Aoki, Y. (2006). Trends of researches on visual greenery since 1974 in Japan. *Environmental Information Science (Japan)*.

Arietta, S. M., Efros, A. A., Ramamoorthi, R., & Agrawala, M. (2014). City forensics: Using visual elements to predict non-visual city attributes. *IEEE transactions on visualization and computer graphics*, *20*(12), 2624-2633.

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*(12), 2481-2495.

Baidu Street View (BSV), 2021. Available online: http://lbsyun.baidu.com/index.php?title=static (accessed on 26 December, 2021).

Bain, L., Gray, B., Rodgers, D., 2012. Living Streets: Strategies for Crafting Public Space. John Wiley & Sons.

Barbierato, E., Bernetti, I., Capecchi, I., & Saragosa, C. (2019). Remote sensing and urban metrics: An automatic classification of spatial configurations to support urban policies. *Earth Observation Advancements in a Changing World; Italian Society of Remote Sensing: Torino, Italy*, 187.

Batty, M. (2008). The size, scale, and shape of cities. *science*, *319*(5864), 769-771.

Berland, A., & Lange, D. A. (2017). Google Street View shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening*, *21*, 11-15.

Beveridge, C. E., & Rocheleau, P. (1995). *Frederick Law Olmsted*. Rizzoli International Publications.

Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. *Communication in the age of virtual reality,* 15(32), 10-5555.

Blanco, H., Alberti, M., Forsyth, A., Krizek, K. J., Rodriguez, D. A., Talen, E., & Ellis, C. (2009). Hot, congested, crowded and diverse: Emerging research agendas in planning. *Progress in Planning*, *71*(4), 153-205.

Bottyán, Z., & Unger, J. (2003). A multiple linear statistical model for estimating the mean maximum urban heat island. *Theoretical and applied climatology*, *75*(3), 233-243.

Bourbia, F., & Boucheriba, F. (2010). Impact of street design on urban microclimate for semiarid climate (Constantine). *Renewable Energy*, *35*(2), 343-347.

Branson, S., Wegner, J. D., Hall, D., Lang, N., Schindler, K., & Perona, P. (2018). From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, 13-30.

Brown, M. J., Grimmond, S., & Ratti, C. (2001). *Comparison of methodologies for computing sky view factor in urban environments* (No. LA-UR-01-4107). Los Alamos National Lab., NM (US).

Cai, B. Y., Li, X., Seiferling, I., & Ratti, C. (2018, July). Treepedia 2.0: applying deep learning for large-scale quantification of urban tree cover. In *2018 IEEE International Congress on Big Data (BigData Congress)* (pp. 49-56). IEEE.

Cao, R., Fukuda, T., & Yabuki, N. (2019). Quantifying visual environment by semantic segmentation using deep learning. *Paper presented at the Intelligent and Informed - Proceedings of the 24th International Conference on Computer-Aided Architectural Design Research in Asia*, CAADRIA 2019, 2, 623-632.

Carrasco-Hernandez, R., Smedley, A. R., & Webb, A. R. (2015). Using urban canyon geometries obtained from Google Street View for atmospheric studies: Potential applications in the calculation of street level total shortwave irradiances. *Energy and Buildings*, *86*, 340-348.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation research part D: Transport and environment*, *2*(3), 199-219.

Chapman, L., & Thornes, J. E. (2004). Real-time sky-view factor calculation and approximation. *Journal of Atmospheric and Oceanic Technology*, *21*(5), 730-741.

Chapman, L., Thornes, J. E., & Bradley, A. V. (2002). Sky-view factor approximation using GPS receivers. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *22*(5), 615-621.

Chen, J., Yang, S., Li, H., Zhang, B., & Lv, J. (2013). Research on geographical environment unit division based on the method of natural breaks (Jenks). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, *3*, 47-50.

Chen, L., Ng, E., An, X., Ren, C., Lee, M., Wang, U., & He, Z. (2012). Sky view factor analysis of street canyons and its implications for daytime intra-urban air temperature differentials in high-rise, high-density urban areas of Hong Kong: a GIS-based simulation approach. *International Journal of Climatology*, *32*(1), 121-136.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017a). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), 834-848.

Chen, L., Papandreou, G., Schroff, F., & Adam, H. (2017b). Rethinking Atrous Convolution for Semantic Image Segmentation. *ArXiv, abs/1706.05587*.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

Cheung, H. K. W., Coles, D., & Levermore, G. J. (2016). Urban heat island analysis of Greater

Manchester, UK using sky view factor analysis. *Building Services Engineering Research and Technology*, *37*(1), 5-17.

Christian, S., Wei, L., Yangqing, J., Pierre, S., Scott, R., Dragomir, A., Andrew, R., 2015. Going deeper with convolutions. Paper Presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213-3223).

Coutts, C. (2008). Greenway accessibility and physical-activity behavior. *Environment and Planning B: Planning and Design*, *35*(3), 552-563.

d'Angelo, P. (2007). Radiometric alignment and vignetting calibration. *Proc. Camera Calibration Methods for Computer Vision Systems*.

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*(7), 1367.

Gal, T., Lindberg, F., & Unger, J. (2009). Computing continuous sky view factors using 3D urban raster and vector databases: comparison and application to urban climate. *Theoretical and applied climatology*, *95*(1), 111-123.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, *114*(50), 13108-13113.

Gehl, J. (1987). *Life between buildings* (Vol. 23). New York: Van Nostrand Reinhold.

Giles-Corti, B., Macintyre, S., Clarkson, J. P., Pikora, T., & Donovan, R. J. (2003). Environmental and lifestyle factors associated with overweight and obesity in Perth, Australia. *American journal of health promotion*, *18*(1), 93-102.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, *56*(1), 114-137.

Gong, P. (2019). Towards more extensive and deeper application of remote sensing. *Yaogan Xuebao/Journal of Remote Sensing*.

Gong, F. Y., Zeng, Z. C., Ng, E., & Norford, L. K. (2019). Spatiotemporal patterns of street-level solar radiation estimated using Google Street View in a high-density urban environment. *Building and Environment*, *148*, 547-566.

Google Street View (GSV), 2021. Available online: https://developers.google.com/maps/documentation/streetview/ (accessed on 26 December 2021).

Grimmond, C. S. B., Potter, S. K., Zutter, H. N., & Souch, C. (2001). Rapid methods to estimate sky-view factors applied to urban areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(7), 903-913.

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: views from urban planning. *American journal of preventive medicine*, *23*(2), 64-73.

He, K., Zhang, X., Ren, S., & Sun, J. (2016a). Proceedings of the IEEE conference on computer vision and pattern recognition.

He, K., Zhang, X., Ren, S., & Sun, J. (2016b). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

He, X., Miao, S., Shen, S., Li, J., Zhang, B., Zhang, Z., & Chen, X. (2015). Influence of sky view factor on outdoor thermal environment and physiological equivalent temperature. *International journal of biometeorology*, *59*(3), 285-297.

Huang, K. T., Lin, T. P., & Lien, H. C. (2015). Investigating thermal comfort and user behaviors in outdoor spaces: A seasonal and spatial perspective. *Advances in Meteorology*, *2015*.

Isalgue, A., Coch, H., & Serra, R. (2007). Scaling laws and the modern city. *Physica A: Statistical Mechanics and its Applications*, *382*(2), 643-649.

Jacobs, J. (1961). Jane jacobs. *The Death and Life of Great American Cities*.

Johansson, E. (2006). Influence of urban geometry on outdoor thermal comfort in a hot dry climate: A study in Fez, Morocco. *Building and environment*, *41*(10), 1326-1338.

Johnson, G. T., & Watson, I. D. (1984). The determination of view-factors in urban canyons. *Journal of Applied Meteorology and Climatology*, *23*(2), 329-335.

Kang, J., Körner, M., Wang, Y., Taubenböck, H., & Zhu, X. X. (2018). Building instance

classification using street view images. *ISPRS journal of photogrammetry and remote sensing*, *145*, 44-59.

Kang, Y., Wang, J., Wang, Y., Angsuesser, S., & Fei, T. (2017). Mapping the Sensitivity of the Public Emotion to the Movement of Stock Market Value: A Case Study of Manhattan. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*.

Krause, J., Sugita, G., Baek, K., & Lim, L. (2018, June). WTPlant (What's That Plant?) A Deep Learning System for Identifying Plants in Natural Images. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval* (pp. 517-520).

Kronkvist, K. (2014). Virtual observations of urban neighborhood physical disorder using Google street view. In *The Stockholm Criminology Symposium, Stockholm, Sweden (2014)* (pp. 169-169). The Swedish National Council for Crime Prevention (BRÅ).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097-1105.

Krüger, E. L., Minella, F. O., & Rasia, F. (2011). Impact of urban geometry on outdoor thermal comfort and air quality from field measurements in Curitiba, Brazil. *Building and Environment*, *46*(3), 621-634.

Landry, S. M., & Chakraborty, J. (2009). Street trees and equity: evaluating the spatial distribution of an urban amenity. *Environment and Planning a*, *41*(11), 2651-2670.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Lee, A. C., & Maheswaran, R. (2011). The health benefits of urban green spaces: a review of the evidence. *Journal of public health*, *33*(2), 212-222.

Leslie, E., Sugiyama, T., Ierodiaconou, D., & Kremer, P. (2010). Perceived and objectively measured greenness of neighbourhoods: Are they measuring the same thing? *Landscape and urban planning*, *95*(1-2), 28-33.

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L. (2019a). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*, *11*(4), 403.

Li, X., Ratti, C., & Seiferling, I. (2018). Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landscape and Urban Planning*, *169*, 81-91.

Li, X., Zhang, C., Li, W., Kuzovkina, Y. A., & Weiner, D. (2015a). Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA. *Urban Forestry & Urban Greening*, *14*(4), 751-759.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015b). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, *14*(3), 675-685.

Liang, J., Gong, J., Sun, J., Zhou, J., Li, W., Li, Y., ... & Shen, S. (2017a). Automatic sky view factor estimation from street view photographs—A big data approach. *Remote Sensing*, *9*(5), 411.

Liang, J., Gong, J., Sun, J., & Liu, J. (2017b). A customizable framework for computing sky view factor from large-scale 3D city models. *Energy and Buildings*, *149*, 38-44.

Lindberg, F., Holmer, B., 2012. Sky View Factor Calculator [User manual - version 1.1]. Available online: https://cms.it.gu.se/infoglueDeliverWorking/digitalAssets/1377/1377754_skyviewfactorcalculator-user-manual.pdf (accessed on 26 December 2021).

Lynch, K. (1964). *The image of the city*. MIT press.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, *105*(3), 512-530.

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, environment and urban systems*, *65*, 113-125.

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).

Louv, R. (2008). *Last child in the woods: Saving our children from nature-deficit disorder*. Algonquin books.

Lu, Y., Yang, Y., Sun, G., & Gou, Z. (2019). Associations between overhead-view and eye-level urban greenness and cycling behaviors. *Cities*, *88*, 10-18.

MacFaden, S. W., O'Neil-Dunne, J. P., Royar, A. R., Lu, J. W., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, *6*(1), 063567.

Madanipour, A. (1996). Urban design and dilemmas of space. *Environment and planning D: Society and Space*, *14*(3), 331-355.

Martinelli, L., & Matzarakis, A. (2017). Influence of height/width proportions on the thermal comfort of courtyard typology for Italian climate zones. *Sustainable Cities and Society*, *29*, 97-106.

Matzarakis, A., & Matuschek, O. (2011). Sky view factor as a parameter in applied climatology-rapid estimation by the SkyHelios model. *Meteorologische Zeitschrift*, *20*(1), 39.

Matzarakis, A., Fröhlich, D., & Gangwisch, M. (2016). Effect of radiation and wind on thermal comfort in urban environments—Applications of the RayMan and SkyHelios model. In *4th International Conference on Countermeasures to Urban Heat Island, National University of Singapore, Singapore* (Vol. 36, pp. 323-334).

Mavrogianni, A., Davies, M., Taylor, J., Chalabi, Z., Biddulph, P., Oikonomou, E., ... & Jones, B. (2014). The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments. *Building and Environment*, *78*, 183-198.

Mehrotra, S., Bardhan, R., & Ramamritham, K. (2020). Diurnal thermal diversity in heterogeneous built area: Mumbai, India. *Urban Climate*, *32*, 100627.

Miao, C., Yu, S., Hu, Y., Zhang, H., He, X., & Chen, W. (2020). Review of methods used to estimate the sky view factor in urban street canyons. *Building and Environment*, *168*, 106497.

Middel, A., Lukasczyk, J., Maciejewski, R., Demuzere, M., & Roth, M. (2018). Sky View Factor footprints for urban climate modeling. *Urban climate*, *25*, 120-134.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, 1419.

Montgomery, J. (1998). Making a city: Urbanity, vitality and urban design. *Journal of urban*

design, *3*(1), 93-116.

Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D., Teitler, J. O., ... & Rundle, A. G. (2016). Use of Google Street View to assess environmental contributions to pedestrian injury. *American journal of public health*, *106*(3), 462-469.

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 779-785).

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2015). *Do people shape cities, or do cities shape people? The co-evolution of physical, social, and economic change in five major US cities* (No. w21620). National Bureau of Economic Research.

Oke, T. R. (1981). Canyon geometry and the nocturnal urban heat island: comparison of scale model and field observations. *Journal of climatology*, *1*(3), 237-254.

Oke, T. R. (1988). Street design and urban canopy layer climate. *Energy and buildings*, *11*(1-3), 103-113.

Open Street Map (OSM), 2021. Available online: https://www.openstreetmap.org/ (accessed on 26 December 2021).

Pikora, T., Giles-Corti, B., Bull, F., Jamrozik, K., & Donovan, R. (2003). Developing a framework for assessment of the environmental determinants of walking and cycling. *Social science & medicine*, *56*(8), 1693-1703.

Porzi, L., Rota Bulò, S., Lepri, B., & Ricci, E. (2015, October). Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 139-148).

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204.

Rundle, A. G., Bader, M. D., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. *American journal of preventive medicine*, *40*(1), 94-100.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211-252.

Saarinen, E. (1948). The search for form in art and architecture.

Sadik-Khan, J. (2012). Urban street design guide. New York: NACTO.

Scarano, M., & Mancini, F. (2017). Assessing the relationship between sky view factor and land surface temperature to the spatial resolution. *International Journal of Remote Sensing*, *38*(23), 6910-6929.

Schroeder, H. W., & Cannon, W. N. (1983). The esthetic contribution of trees to residential streets in Ohio towns. *Journal of Arboriculture*, *9*(9), 237-243.

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.

Shet, V. (2014). Go back in time with Street View. *Official Google Blog, Available at: https://googleblog. blogspot. com/2014/04/go-back-in-timewith-street-view. html, accessed Decmbert*, *26*(2021), 319-341.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR, abs/1409.1556*.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, *93*(12), 1879-1900.

Steyn, D. G. (1980). The calculation of view factors from fisheye-lens photographs: Research note.

Sun, Y., Liu, Y., Wang, G., & Zhang, H. (2017). Deep learning for plant identification in natural environment. *Computational intelligence and neuroscience*, *2017*.

Tencent Street View (TSV), 2021. Available online: http://lbs.qq.com/panostatic_v1/ (accessed on 26 December 2021).

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., & James, P. (2007). Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and urban planning*, *81*(3), 167-178.

Venhari, A. A., Tenpierik, M., & Taleghani, M. (2019). The role of sky view factor and urban street greenery in human thermal comfort and heat stress in a desert climate. *Journal of Arid Environments*, *166*, 68-76.

Wales, N.R., 2016. Tree Cover in Wales' _Towns and Cities. Retrieved from Cardiff, UK.

Wang, W., Yang, S., He, Z., Wang, M., Zhang, J., & Zhang, W. (2018, April). Urban perception

of commercial activeness from satellite images and streetscapes. In *Companion Proceedings of the The Web Conference 2018* (pp. 647-654).

Watson, I. D., & Johnson, G. T. (1987). Graphical estimation of sky view-factors in urban environments. *Journal of climatology*, *7*(2), 193-197.

Wegner, J. D., Branson, S., Hall, D., Schindler, K., & Perona, P. (2016). Cataloging public objects using aerial and street-level images-urban trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6014-6023).

Wolf, K. L. (2005). Business district streetscapes, trees, and consumer response. *Journal of Forestry*, *103*(8), 396-400.

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.

Yang, J., Wong, M. S., Menenti, M., & Nichol, J. (2015a). Study of the geometry effect on land surface temperature retrieval in urban environment. *ISPRS journal of photogrammetry and remote sensing*, *109*, 77-87.

Yang, J., Wong, M. S., Menenti, M., & Nichol, J. (2015b). Modeling the effective emissivity of the urban canopy using sky view factor. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 211-219.

Yang, J., Zhao, L., Mcbride, J., & Gong, P. (2009). Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, *91*(2), 97-104.

Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015). 'Big data'for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Applied Geography*, *63*, 337-345.

Yin, L., & Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Applied geography*, *76*, 147-153.

Yu, X., Zhao, G., Chang, C., Yuan, X., & Heng, F. (2019). Bgvi: A new index to estimate street-side greenery using baidu street view image. *Forests*, *10*(1), 3.

Zeng, L., Lu, J., Li, W., & Li, Y. (2018). A fast approach for large-scale Sky View Factor estimation using street view images. *Building and Environment*, *135*, 74-84.

Zhang, F., Wu, L., Zhu, D., & Liu, Y. (2019). Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. *ISPRS Journal of*

Photogrammetry and Remote Sensing, *153*, 48-58.

Zhang, W., Witharana, C., Li, W., Zhang, C., Li, X., & Parent, J. (2018). Using deep learning to identify utility poles with crossarms and estimate their locations from google street view images. *Sensors*, *18*(8), 2484.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881-2890).