Mapping the Urban Landscape: A Remote Sensing and Deep Learning Approach to Identifying Forests and Land Cover Features in a Desert City

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Geography

in the

College of Graduate Studies

University of Idaho

by

Zhe Wang

Approved by:

Major Professor: Haifeng Liao, Ph.D.

Committee Members: Karen Humes, Ph.D.; Xian Min, Ph.D.; Chao Fan, Ph.D.

Department Administrator: Alistair M.S. Smith, Ph.D.

May 2023

Abstract

Urban forests are essential components of the urban environment, providing a wide range of benefits to society, such as air pollution reduction, carbon sequestration, and improved mental health. Accurately mapping and monitoring urban forests is crucial for better urban forest management and planning. In this dissertation, I explore the application of deep learning techniques to extract urban tree canopies and classify land cover features in urban areas. Specifically, Chapter 2 investigates the effectiveness of the U-net in urban tree canopy mapping using high-resolution aerial photos. Our results show that the U-net can achieve exceptional accuracy scores in fine-scale land cover mapping and outperforms other widely used methods. However, the lack of publicly available training datasets and very high-resolution data remains a challenge for scaling up the U-net to large areas.

Based on the findings in Chapter 2, I proposed a cross-scale transfer learning framework that utilizes pre-trained models at various scales to classify land cover features in Phoenix, Arizona. The results demonstrate that our framework achieves high accuracy in classifying impervious surfaces, buildings, low vegetation, and tree classes, outperforming other commonly used methods. Moreover, the robustness and feasibility of our proposed framework are evaluated by testing it over time and in different geographic areas, demonstrating its generalization ability. Overall, our study highlights the potential of deep learning techniques and transfer learning in improving the accuracy and efficiency of urban land cover classification, particularly in complex urban environments with varying spatial resolutions.

In addition to mapping urban forests and land cover, Chapter 4 explores the complex relationships between urban forest characteristics and socio-economic indicators in Phoenix,

Arizona, using advanced statistical models. The study finds that minority and low-income populations have limited access to natural amenities, which is consistent with earlier research demonstrating that trees are disproportionately distributed in neighborhoods with low-income or people of color. However, the study also finds that communities with more trees and shrubs are associated with higher educational performance, highlighting the potential positive effects of urban forests on public health and well-being, including educational outcomes. The results of the study suggest that targeted and localized approaches to urban forest management are necessary to promote equitable access to green spaces and improve public health and well-being in diverse metropolitan areas like Phoenix.

This dissertation makes significant contributions to the field of land cover mapping and urban remote sensing. The study proposes novel deep learning-based frameworks that utilize transfer learning and multi-scale segmentation techniques to achieve accurate and efficient urban land cover classification. Specifically, the study demonstrates the effectiveness of the U-net architecture for urban tree canopy mapping and the cross-scale transfer learning framework for land cover classification in urban areas. These findings provide valuable insights into the potential of deep learning and transfer learning for improving the accuracy and efficiency of urban land cover mapping, particularly in complex urban environments with varying spatial resolutions. Moreover, the study sheds light on the complex relationships between urban forest characteristics and socio-economic indicators, highlighting the need for targeted and localized approaches to urban forest management to promote equitable access to green spaces and improve public health and well-being in diverse metropolitan areas. Overall, the contributions of this dissertation provide valuable insights into the application of advanced remote sensing and machine learning techniques for urban land cover mapping and

management, with potential applications in urban planning, environmental management, and public health research.

Acknowledgments

I would like to express my appreciation to my two advisors during my doctoral studies. The first one is Dr. Chao Fan. I came to the University of Idaho because of her, and I am grateful for her decision to choose me as her student. During the three years under her guidance, my academic ability has improved significantly. She is rigorous and strict, occasionally harsh with me, but it is because of these traits that I am serious and confident in my scientific research today. I also want to thank Dr. Haifeng Liao. After Dr. Fan left the geography department, he took over as my advisor. He gave me a lot of freedom and understanding, allowing me to work remotely and supporting my scientific research activities. I am grateful for his guidance and support during the last two years. I would also like to thank Dr. Min Xian of the computer science department. His image processing course opened the door for me to deep learning. Starting from my second year of the doctoral program, as a professor in the computer science department, he has been generous in guiding me in computer science. Under his guidance, I also obtained a master's degree in computer science. At the same time, the application of deep learning to remote sensing images has become my new field of research. I would also like to thank Dr. Xiaogang Ma of the computer science department. During the last two years of my doctoral program, I joined his project team and received his support for tuition fees. During these two years, I met many new friends from different schools academically and learned many advanced ideas. At the same time, with his help, I got many opportunities to practice. I am grateful to have met him during the last two years of my doctoral program. Finally, I would like to thank Professor Karen Humes of the geography department. She is a very kind teacher. I am happy to have been her teaching assistant for

two years. She provided me with great support and help in the geography department and was also very patient with me.

Dedication

I would like to thank the friends I met during my time in the United States. Everyone's emotions during their doctoral studies are up and down. Because of the presence of friends, I felt less lonely. I would also like to thank my parents. It is because of their support that I was able to begin my academic career in the United States, and it is because of their understanding that I was able to continue to pursue my academic dreams. Over the years, they have given me financial and emotional support. I am grateful to have such happy parents.

Finally, I would like to thank my husband. We got married during my second year of doctoral studies. He always provides me with unconditional support and companionship when I am most helpless. When I am under pressure, he is my spiritual mentor. When I encounter academic difficulties, he is my Zhuge Liang. He supports all of my decisions and is my strongest support. I am grateful to have met him and to have married him.

I would like to thank all the people I have met in these five years, who have made up my colorful doctoral life.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Research Background

Urban remote sensing involves using satellite and aerial imagery to analyze and monitor urban environments. Urban areas need constant monitoring and management to ensure sustainability and livability because they are growing and changing quickly. Various urban features and processes, such as land use, land cover, urban heat islands, air and water quality, and urban sprawl, can be studied using remote sensing data (Fan & Wang, 2020; Tran et al., 2017). To gain insights into urban dynamics and aid in decision-making, the data can be analyzed using various methods, including image processing, classification, and machine learning (Reichstein et al., 2019).

Getting data with a high spatial resolution is one of the critical issues with urban remote sensing. Urban areas have a high degree of heterogeneity and various features and structures, all of which call for accurate and detailed representation (Thapa & Murayama, 2009). High-resolution satellite and aerial imagery are needed to achieve this level of detail (Xu et al., 2014). The cost of these data can be high, and they might not be widely accessible or available everywhere.

The requirement for accurate and trustworthy data processing and analysis presents another challenge for urban remote sensing. This entails creating complex algorithms and techniques to extract pertinent information from the remote sensing data and analyzing the data to produce insights into urban dynamics and processes. Data errors and inaccuracies can influence decision-making and result in incorrect conclusions (Congalton, 1991). Deep learning has emerged as a powerful tool for analyzing remote sensing data in urban areas to address these challenges (Yuan et al., 2020).

Deep learning is a subset of machine learning that utilizes neural networks to learn patterns and features from large amounts of data. Deep learning is used in urban remote sensing to evaluate and interpret satellite and aircraft data (X. X. Zhu et al., 2017). Deep learning applications in urban remote sensing mainly fall into three categories: image classification (G. Cheng et al., 2020), object recognition (Z. Li et al., 2022), and image restoration (Rasti et al., 2021). In this dissertation, I focused on the task of image classification in urban areas.

In urban remote sensing, deep learning can classify and segment different urban features, such as buildings, roads, and vegetation. This can contribute to a more accurate and detailed depiction of the urban environment, which is especially important in areas with high levels of urbanization and complex land use patterns (Kuras et al., 2021). Another application of deep learning in urban remote sensing is in change detection. Urban areas are developing rapidly, with new buildings, infrastructure, and land use patterns emerging regularly. Deep learning may be used to automatically detect and classify these changes, providing significant insights into urban dynamics and patterns across time.

Additional to the land cover mapping in urban areas, my other focus in this dissertation is extracting the urban forest information from land cover maps in urban areas. Urban forests play a vital role in maintaining urban areas' ecological, economic, and social well-being. However, studying and managing urban forests can be challenging, particularly in regions with high levels of urbanization and complex land use patterns (Fan et al., 2019a). The application of deep learning techniques in urban forest study has the potential to provide more accurate and detailed information on the structure and function of urban forests.

One of the key benefits of using deep learning in urban forest study is the improved accuracy and efficiency in analyzing remote sensing data. Deep learning models can provide more

accurate and efficient remote sensing data analyses, reducing the time and resources required for data processing and analysis (Ma et al., 2019). Additionally, deep learning can improve the spatial resolution of remote sensing data, enabling more detailed and accurate mapping and monitoring of urban forests. Another benefit of using deep learning in urban forest study is the improved interpretation of remote sensing data. Deep learning techniques can help identify patterns and features in remote sensing data that may be difficult to detect using traditional image processing and analysis methods (L. Zhang et al., 2016). Furthermore, deep learning in remote sensing can significantly improve image analysis's speed, accuracy, and scalability, making it possible to analyze and extract useful information from large amounts of data in near real-time (Sedona et al., 2019). This can provide valuable insights into the structure and function of urban forests, helping to guide management and conservation strategies.

Despite the promise of deep learning in urban remote sensing, applying deep learning techniques in urban forest study also faces several challenges. Access to high-quality and high-resolution remote sensing data is critical for accurate and detailed analyses of urban features and urban forests. Obtaining such data can be expensive and may be limited in some regions (Z. Wang et al., 2021). Deep learning models can be complex, making interpreting how they arrive at their results difficult. The availability of labeled data can also be limited, making it challenging to train deep-learning models effectively. Additionally, the interpretability of deep learning models is still an ongoing challenge, as it can be difficult to understand how the neural networks are making their predictions (Stiglic et al., 2020).

In summary, the use of deep learning techniques has the potential to provide more precise and detailed information on the structure and function of urban landscape and urban forests.

While it has many advantages, it also has some issues that must be solved to fully fulfill its potential in urban planning and urban forest management. Decision-makers and managers can build more effective strategies for conserving and managing urban ecosystems by using the potential of deep learning and maximizing the ecological, economic, and social benefits for urban communities (Dubey et al., 2016; Z. Wang et al., 2021).

Despite the challenges in applying deep learning techniques to urban forest research, such as a lack of high-quality remote sensing data and the complexity of deep learning models, my dissertation usually integrated remote sensing and deep learning techniques to overcome these obstacles. I developed a cross-scale transfer learning framework for deep learning-based land cover classification tasks utilizing open-access data that was temporally and spatially assessed. The proposed approach has a lot of potential for use in other cities, and it can assist in enhancing the accuracy and efficiency of urban forest monitoring and management.

1.2 Literature review

1.2.1 Urban remote sensing

Urban remote sensing is a rapidly evolving interdisciplinary field of research that utilizes an array of imaging technologies, including satellite and aerial imagery, to investigate the physical, social, and economic characteristics of urban areas (Fugate et al., 2010). This field has become increasingly pivotal in providing a comprehensive understanding of the complexities of contemporary cities and their development.

The multifarious applications of urban remote sensing encompass a broad range of areas. Remote sensing data offers an indispensable tool to chart the physical structure and functional layout of cities, including the distribution of green spaces, land use patterns, and

transportation networks. The analysis of such information is of paramount importance to urban planners and decision-makers as it illuminates the existing conditions and constraints of a city while also providing critical insight into future urban development trajectories.

Furthermore, urban remote sensing has become a cornerstone of environmental monitoring, providing a reliable means to monitor the impacts of urbanization on the natural environment. Remote sensing data can be utilized to track changes in land cover and land use, assess the loss of natural habitats, and quantify the degradation of air and water quality(Ahmed & Akter, 2017; K.-S. Cheng et al., 2008; Hegazy & Kaloop, 2015) It is also an indispensable tool for tracking and monitoring the distribution and extent of environmental hazards such as floods, wildfires, and oil spills (Joyce et al., 2009).

In addition to physical characteristics, remote sensing has also proven to be a powerful tool in analyzing the socio-economic features of urban areas. It can help to discern patterns of wealth, income, and poverty, and to track changes in these patterns over time. Such information is of critical importance in elucidating the social and economic conditions of urban residents and in addressing issues related to inequality and poverty (Chan & Hopkins, 2017; Fan et al., 2019a; Wilkerson et al., 2018).

Urban remote sensing is also reliant on a suite of analytical methods and techniques, including image processing and analysis, Geographic Information Systems (GIS), and statistical methodologies to extract meaningful information from remote sensing data. Recent advances in technology have revolutionized urban remote sensing, including the increasing availability of high-resolution satellite imagery, which has improved the ability to discern fine-scale features and changes in urban areas(Kuras et al., 2021; Pashaei et al., 2020; Y. Zhang & Billie, 2020) Additionally, machine learning and deep learning techniques have

been implemented to enhance the accuracy and efficiency of land use and land cover classification, building detection, and other key applications. The usage of Unmanned aerial vehicles (UAVs) and Light Detection and Ranging (LiDAR) has also enhanced the spatial and temporal resolution of remote sensing data for urban areas, thus enabling more detailed observations and analyses(Alonzo et al., 2016; De Luca et al., 2019).

In summary, urban remote sensing is an interdisciplinary field that plays a significant role in comprehending the complex dynamics of modern cities and the issues that influence urban life. It offers a powerful tool for urban planning, environmental management, disaster response, and socio-economic analysis, and it has critical implications for the future development of cities and their inhabitants. As such, it is vital for future research to continue to advance this field, in order to better understand and address the complex challenges posed by urbanization.

1.2.2 Urban Forest

The growing significance of urban forests in the United States is due to the rise of urban land. Over 141 million acres of American forests are located in cities and towns. Urban forests come in various forms, such as urban parks, street trees, landscaped boulevards, gardens, river and coastal promenades, greenways, river corridors, wetlands, nature preserves, shelter belts of trees, and working trees at former industrial sites. Urban forests, by connecting green spaces, create the green infrastructure that communities rely on. Green infrastructure functions at various levels, from the neighborhood to the metropolitan area to the regional landscape (Pauleit et al., 2017). The 2010 census revealed that almost 81% of Americans reside in urban areas, an increase from 79% a decade prior. Urban populations also grew by 12.1%, surpassing the national average growth rate of 9.7%(Bureau, n.d.). This trend

indicates that the United States is becoming more urbanized. Due to these population patterns, urban forests have become increasingly crucial- they are the trees in our immediate surroundings. They are dynamic ecosystems that offer essential advantages to both humans and wildlife (Coutts & Hahn, 2015).

An urban forest in a desert city refers to a collection of trees and other vegetation located within an urban or suburban area in a region that experiences dry, arid conditions. These trees and plants are important for a number of reasons, including providing shade, cleaning the air, reducing noise pollution, and improving the aesthetics of the city. They can also provide habitat for wildlife and help to mitigate the urban heat island effect, which is the phenomenon of cities being significantly warmer than surrounding rural areas due to the presence of concrete and other heat-absorbing materials (Epa, 2016).

There are many cities around the world that are located in desert regions, including Phoenix, Arizona and Las Vegas, Nevada in the United States; Dubai and Abu Dhabi in the United Arab Emirates; and Adelaide and Perth in Australia. In these cities, the urban forests play a critical role in providing a green space for residents and helping to mitigate the harsh desert climate. These forests are typically made up of drought-tolerant trees and plants that are able to withstand the dry conditions and high temperatures of the desert.

In this dissertation, the main study area is Phoenix. More than 94,000 trees have roots on city-owned land in Phoenix, from the iconic towering palms of Encanto Park to native mesquite and paloverdes that line many street medians and rights-of-way (Gardiner, n.d.). Almost 1,000 trees are lost every year to the powerful gusts of monsoon storms or accidents. The city used to replace them, but it nearly eliminated funding for tree replacement during the recession. Some residents and civic groups say the loss of trees in public spaces has

become apparent after years of budget cuts. And the city recently agreed to spend more money on trees (Gardiner, n.d.). Trees in Phoenix are more important since they provide more shade, which can help reduce the urban hear island effect and lower energy consumption (*Official Website of the City of Phoenix, Arizona*, n.d.).

Urban forests are essential for mitigating the urban heat island effect in Phoenix, where temperatures can be significantly higher in urban areas compared to rural areas (Hillel & Rosenzweig, 2010). Disparities in tree canopy coverage exist among different socio-economic groups, with low-income and minority populations having limited access to natural amenities such as parks and urban forests (R. Wang et al., 2017). In general, the urban forests in Phoenix face a number of challenges, including limited water availability, high temperatures, and exposure to strong winds. To overcome these challenges, these forests often rely on irrigation systems and other innovative water-saving technologies, as well as careful planning and management to ensure that the right trees and plants are selected for the local climate (Nikolaou et al., 2020). Despite these challenges, the urban forests in desert cities can be beautiful and thriving places, providing a vital green space for the community and helping to make the city a more livable place. Therefore, there are increasing studies to explore the relationships between the green space and socioeconomic conditions in cities.

There is evidence of associations between indicators of urban forests and socioeconomic factors. Urban forests can provide a range of benefits to communities, including increased property values, improved air and water quality, and reduced energy costs. However, these benefits are not equally distributed across different socioeconomic groups.

Research has shown that low-income communities and communities of color are often less likely to have access to urban forests and green spaces, and that these communities also

experience higher levels of environmental pollution and stress. This can contribute to health disparities and a lower quality of life for residents in these communities. Additionally, urban development patterns and zoning regulations can also play a role in shaping access to urban forests and green spaces. For example, higher-income communities may be more likely to have access to private parks and green spaces, while lower-income communities may have limited access to public parks and green spaces. In conclusion, the relationship between urban forests and socioeconomic factors is complex, and it is important to consider the ways in which access to urban forests and green spaces may be influenced by factors such as income, race, and urban development patterns.

The spatial variation in socioeconomic factors leads to significant differences in the supply and demand of ecosystem services provided by the green infrastructure in cities (Wilkerson et al., 2018). As shown in Table 1.1, multiple studies have examined associations between urban forest structure and socioeconomic situations in recent years. Socioeconomic drivers are critical determinants of urban forest structure and fragmentation.

1.2.3 Urban tree canopy extraction and land cover mapping

Urban tree canopy, defined as the ground area covered by the layer of tree leaves, branches, and stems (Grove et al., 2006), is among the most widely used indicators of urban forest studies. Manual field measurement was the earliest method used to study urban forests. In this method, the whole city or some parts of an area are randomly selected for sampling (Nowak et al., 2008). Conventional approaches to measuring tree canopy focus on using field-based surveys to manually calculate the area of urban vegetation (Fuller & Gaston, 2009; King & Locke, 2013; Loughner et al., 2012). These methods are costly and labor-intensive, prompting scientists to integrate traditional tools into new remote sensing systems,

such as different satellite images, passive optical systems, and active sensors (Shojanoori & Shafri, 2016).

Table 1.1 Studies examining associations between indicators of urban forests and socioeconomic factors.

Variables	Positive association	Negative association
Median Income	(Chan and Hopkins 2017; Dobbs, Kendal, and Nitschke 2014; Fan et al. 2019; Grove, Locke, and O'Neil-Dunne 2014; Iverson and Cook 2000; Lowry, Baker, and Ramsey 2012b; Mennis 2006; Schwarz et al. 2015)	
Poverty	(Talarchek, 1990)	
Education level	(Dobbs et al., 2014; Fan et al., 2019a; Grove et al., 2014; Mennis, 2006; Talarchek, 1990)	
Median home-built year	(Sorrensen et al., 2015)	(Conway & Hackworth, 2007; Fan et al., 2019a; Mennis, 2006)
Housing value	(Grove, Locke, and O'Neil-Dunne 2014; Mennis 2006)	
Household density	(Grove, Locke, and O'Neil-Dunne 2014; Mennis 2006)	(Fan et al., 2019a; Iverson & Cook, 2000)
Crime		(Grove, Locke, and O'Neil-Dunne 2014)
Cultural diversity	(Dobbs et al., 2014)	(Fan et al., 2019a; Mennis, 2006; Talarchek, 1990)
Percent Latino/ Hispanic		(Frey, 2017; Ogneva-Himmelberger et al., 2009; Sorrensen et al., 2015) (Fan et al., 2019a)
Percent black		(Frey, 2017; Schwarz et al., 2015)
Percent white	(Sorrensen et al., 2015)	
Percent renter		(Fan et al., 2019a; Frey, 2017; Riley & Gardiner, 2020; Talarchek, 1990)
Population	(Frey, 2017)	(Conway and Hackworth 2007; Dobbs, Kendal, and Nitschke 2014; Frey 2017; Grove, Locke, and O'Neil-Dunne 2014; Mennis 2006)

Table 1.2 lists several pieces of information on popular satellite sensors. Compared to the laborious manual field measurement, the remote sensing images cover a more considerable spatial extent with higher temporal resolution. For example, the Moderate Resolution Imaging Spectroradiometer – Vegetation Continuous Field (MODIS-VCF) is medium-resolution imagery used to monitor urban forests yearly with a spatial resolution of 250–500 m (*LP DAAC - MODIS Overview*, n.d.). The other medium-resolution satellite systems, such as Landsat Thematic Mapper (TM) / Enhanced Thematic Mapper Plus (ETM+)/ Operational Land Imager (OLI) used in urban studies, are Landsat systems. These data can provide a means to rapidly detect and monitor the land cover changes (*Landsat Satellite Missions*, n.d.). Landsat images can be acquired every 16 days. Moreover, QuickBird and IKONOS are two commonly used high-resolution satellite imageries in urbanization and sustainability studies. Both have panchromatic and four multispectral bands (i.e., red, green, blue, and near-infrared) (Satellite Imaging Corp n.d.). In recent years, very high-resolution satellite images have increasingly been employed in forest management and mapping at the meter and sub-meter levels. The National Agriculture Imagery Program (NAIP) product coupled with airborne Light Detection and Ranging (LiDAR) is a popular data combination, especially for fine-scale urban forest mapping and analysis (Alonzo et al., 2014, 2016; MacFaden et al., 2012).

Table 1.2 Several satellite sensor characteristics

Data (Product)	Spatial resolution (MSS)	Temporal resolution	Temporal extent
MOIDS-VCF	250 m	Yearly	2000-2020
Landsat	30 m	16 days	1972-present
SPOT-5	10 m	Global	1986-present
IKONOS	4 m	3-5 days	2000-2015
QuickBird	2.9 m	1-3.5 days	2001- present
Worldview-2	1.84 m	1.1-3.7 days	2009-present
LiDAR	0.15~1 m	/	/
NAIP	0.6-3 m	/	2007-2020

*MSS (multispectral)

Urban tree canopy extraction in land cover mapping using remote-sensing images identifies and classifies trees leading to the urban tree canopy and greenspace mapping. Mapping is conducted via urban forestry monitoring methods, classified into three groups: manual interpretation, pixel-based, and object-based approaches (Shojanoori & Shafri, 2016). Conventional methods involve supervised classification such as maximum likelihood algorithm and minimum distance and unsupervised classification such as Iterative Self-Organizing Data Analysis Technique (ISODATA) and K-means (Ahmed & Akter, 2017; Manakos et al., 2000). Urban land cover mapping has benefited mainly from the development of these image classification algorithms. For example, Yang et al. (2003) applied the decision tree approach to agriculture mapping over Quebec, Canada, using hyperspectral satellite images (Yang et al., 2003). Zhu and Blumberg (2002) applied the support vector machine (SVM)-based classification method to the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data to map urban land cover

features in Beer-Sheva, Israel (G. Zhu & Blumberg, 2002). Ahmed and Akter (2017) utilized the K-means approach to analyze the land cover changes in southwest Bengal delta, Bangladesh (Ahmed & Akter, 2017).

Many conventional pixel-based classification methods failed to achieve satisfied classification results because they solely relied on the spectral signatures of individual pixels (Myint et al., 2011). The object-based Image Analysis (OBIA) approach offered a more effective solution for extracting the required information rapidly and efficiently (Yadav, Rizvi, and Kadam 2015). The OBIA approach first groups spectrally similar pixels into discrete objects and then employs a classification algorithm to assign the segmented objects into different classes. It fully utilizes spatial characteristics of segmented objects such as shape, size, and compactness to classify the land surface features at a local level (Blaschke, 2010; Hay & Castilla, 2006). Such a process can improve classification accuracy and map small urban elements, such as mature individual trees or small shrubs (Mathieu et al., 2007). The OBIA approach has become a mainstream approach to urban land cover mapping in the recent decade, and its effectiveness has been demonstrated in many studies across the globe (De Luca et al., 2019; Hossain & Chen, 2019; X. Li et al., 2014; MacFaden et al., 2012). Despite its effectiveness, however, the OBIA approach relies heavily on expert and subjective knowledge to determine the appropriate scales, parameters, functions, and classifiers, making its classification results vary widely from case to case and person to person (Blaschke, 2010). In recent years, to process a large amount of data, deep learning has been widely used in remote sensing image classification (Diakogiannis et al., 2020; Weinstein et al., 2019; Yuan et al., 2020). In contrast to the OBIA approach, deep learning

models automatically learn the features in the study area without too much human intervention on parameter decisions (Liang & Li, 2016).

1.2.4 Deep learning in remote sensing and image analysis

As a subfield of machine learning, deep learning has gained recent popularity in remote sensing for its ability to characterize complex patterns (Alonzo et al., 2014). Compared with the traditional rule-based image classification methods and machine learning algorithms, the deep learning-based classification methods have significant advantages in classification accuracy, especially in complex urban areas. The combined use of multisource data also enhances the accuracy of land cover mapping. However, the manually labeled land cover dataset for the training of deep learning models is still inadequate for most tasks. This inadequacy limits the wide practical application of deep learning for land cover classification (Yuan et al., 2020). CNN, in particular, has been successfully applied in various tasks, such as object detection and semantic segmentation (Ondruska et al., 2016; Qian et al., 2015). CNN supports the fast acquisition of rich spatial information from image features (Krizhevsky et al., 2012a) and automatically learns features requiring a minimal amount of expert knowledge. Convolutional networks are powerful visual models that yield hierarchies of features and are generally used for image-level classification. Long et al. illustrated a fully convolutional network (FCN) to classify images at the pixel level, thus solving the problem of semantic segmentation (Long et al., 2015a) (Figure 1.1).

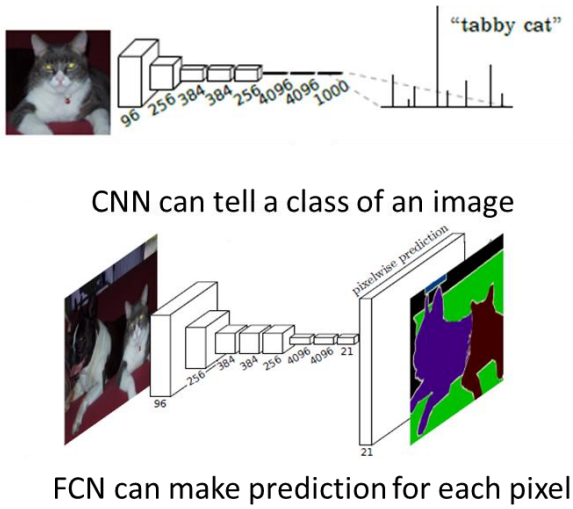


Figure 1.1 General difference between CNN and FCN.

Moreover, compared to other deep learning methods, the U-net proposed by Ronneberger (2015) appears to be more popular and more quickly adopted and modified for remote sensing image segmentation (Feng et al., 2018; Ronneberger et al., 2015). It was first applied to biomedical segmentation and developed further to refine boundary delineation (Falk et al., 2019). It also works with smaller datasets and yields more precise semantic segmentation (Pan et al., 2020b). Despite its adoption in biomedical fields, some researchers gradually have applied the U-net to land cover mapping in recent years. For instance, Wagner et al. (2020) proposed an instance-segmentation-based U-net architecture to delineate individual buildings using the WordView-3 imagery over the Joanópolis City in Brazil (Wagner et al., 2020). Feng et al. (2018) applied the U-net architecture to the WordView-2 images and GaoFen-2 images to extract water bodies (Feng et al., 2018). Another study employed the U-

net to detect palm trees in Jambi, Indonesia and Bengaluru, India, using the WorldView-2 images (Freudenberg et al., 2019).

However, one problem is the limitation of labeled information for newly collected remote sensing images. This phenomenon will make it even more challenging for the deep learning models to process the RS images. With the development of modern satellite sensors and easy access to remote sensing data, the problem of processing such a large amount of data becomes even more severe and urgent. A straightforward consideration is to resort to existing labeled remote sensing data to help with the unknown new data. To achieve this purpose, transfer learning-based frameworks that can overcome the semantic gap between different datasets have become a research frontier in remote sensing data processing (Hu et al., 2015). The information of existing labeled data is exploited to predict the label of newly collected remote sensing data. Razavian et al. (2014) used a model trained for image classification. They conducted a series of transfer learning experiments to investigate a wide range of recognition tasks such as object image classification, scene recognition, and image retrieval (Sharif Razavian et al., 2014). Hu et al. verified the utility of transfer learning from pre-trained CNN models to perform remote-sensing scene classification (Hu et al., 2015). Weinstein et al. refined a CNN model using many higher-quality hand-annotated aerial images to detect the individual tree crowns in a much larger area (Weinstein et al., 2019). Yosinski et al. concluded that using transfer learning from distant tasks performs better than training CNN models from scratch (with randomly initialized weights) (Yosinski et al., 2014).

1.3 Data and methodology

1.3.1 Data

The data used in this dissertation include Vaihingen and Potsdam datasets by International Society for Photogrammetry and Remote Sensing (ISPRS) (*ISPRS Benchmark Test on Urban Object Detection and Reconstruction - ISPRS*, n.d.), Phoenix land cover and land use data by Central Arizona-Phoenix Long-Term Ecological Research (CAP LTER) program (Y. Zhang & Billie, 2020), the LiDAR data by United States Geological Survey (USGS) (*USGS.Gov | Science for a Changing World*, n.d.), and socioeconomic data from U.S. Census government (Bureau, n.d.).

The Vaihingen and Potsdam datasets are two datasets created by the ISPRS. They are widely used for evaluating the performance of remote sensing image analysis algorithms, specifically in the area of semantic segmentation. The ISPRS Vaihingen dataset consists of high-resolution aerial imagery of an urban area in Germany, along with corresponding ground truth annotations for building and road detection. The ISPRS Potsdam dataset is similar, but includes a larger area and more classes, including cars, low vegetation, trees and impervious surfaces. Both datasets are widely used in the remote sensing and computer vision community for training and evaluating machine learning algorithms for semantic segmentation tasks.

The land cover mapping for Phoenix obtained from CAPLTER program is derived based on 2015 National Agriculture Imagery Program (NAIP) imagery. The project uses aerial imagery collected by the National Agriculture Imagery Program (NAIP) in 2015 as input data. The objective is to classify the images into different land cover classes, such as urban, agriculture, forest, water, and others, using computer vision and machine learning techniques.

The resulting map provides valuable information for various applications, including natural resource management, urban planning, and environmental monitoring.

The Phoenix USGS LiDAR Data is a Light Detection and Ranging (LiDAR) dataset collected by the USGS in 2014 for the Metro Phoenix area. LiDAR is a remote sensing technology that uses laser pulses to measure the distance to the ground and generate highly accurate 3D representations of the earth's surface. The 2014 Phoenix USGS LiDAR Data provides high-resolution topographic information for the region, including elevation data and information about the structure of the terrain and vegetation. This information can be used for a variety of purposes, such as flood risk assessment, urban planning, and environmental monitoring. Additionally, the dataset can be combined with other geospatial data, such as aerial imagery and satellite data, to support a wide range of applications in the fields of GIS, geography, and environmental science.

The US Census Bureau collects and publishes a wide range of socioeconomic data about the US population, including information on demographics, education, employment, income, poverty, housing, and more. This data is collected through the decennial census, as well as through various ongoing surveys and programs. The socioeconomic data from the US Census is widely used by businesses, governments, researchers, and the public to better understand the makeup and needs of the US population. It provides valuable insights into demographic trends, economic conditions, and other factors that shape the nation and inform decision-making across many sectors.

1.3.2 Methodology

Chapter 2 and Chapter 3 focus on urban tree canopy extraction and land cover mapping problem using the deep learning techniques. U-net is the main neural network used in the

dissertation. U-Net is a deep learning architecture for image segmentation tasks, developed for biomedical image analysis. It consists of a contracting path (downsampling) and an expansive path (upsampling) with skip connections between them. The architecture is symmetrical and named U-Net due to its shape. The output of the network is a segmentation mask with the same spatial resolution as the input image. Furthermore, transfer learning concept plays an important part in Chapter 3. Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task. The aim is to leverage the knowledge gained from the first task to improve performance on the second task. Fine-tuning method was applied to predict the land cover in the study area. It is a process in transfer learning where a pre-trained model is further trained on a new task, usually with a smaller learning rate, to adapt to the specific characteristics of the new task. The pre-trained weights serve as initialization, providing a good starting point for the fine-tuning process, speeding it up and reducing the amount of training data needed.

Chapter 4 conducted spatial analysis to explore the socio-economic indicators of urban forest characteristics. In order to understand the relationships and spatial patterns between the urban forest structure and tis socioeconomic indicators, this dissertation takes full advantage of recent advancements in GIS and exploratory data analysis method, including ordinary least square (OLS) regression, geographically weighted regression (GWR) and multi-scale GWR (MGWR).

OLS, GWR, and MGWR are all methods for regression analysis, but they have different approaches to modeling the relationship between the dependent and independent variables.

OLS is a method for estimating the parameters of a linear regression model by minimizing the sum of squared differences between the observed and predicted values of the dependent

variable. OLS assumes that the relationship between the dependent and independent variables is linear and that the errors are normally distributed and have constant variance.

GWR is a variant of OLS that allows for the parameters of the regression model to vary across space. In a GWR model, the regression equation is estimated separately for each location in the study area, taking into account the spatial heterogeneity of the data. This allows the model to better capture the spatial variation in the relationship between the dependent and independent variables. MGWR is an extension of GWR that allows for the spatial scale of the regression model to vary. In an MGWR model, multiple regression equations are estimated for each location, each with a different spatial scale, and the final model is a combination of these regression equations. This allows the model to capture different patterns of spatial variation at different scales, making it well-suited for data with complex spatial patterns.

OLS, GWR, and MGWR are all methods for regression analysis that have different approaches to modeling the relationship between the dependent and independent variables. OLS is the simplest method and assumes a linear relationship and constant variance, while GWR and MGWR allow for more flexible modeling of spatial heterogeneity. The choice between these methods will depend on the nature of the data and the research question being addressed.

1.4 Research objectives

The distribution of urban tree canopy and its benefits vary in the US, with urban land expansion causing its importance to increase. Most studies on urban forests have focused on their structure represented by urban tree canopy cover, with satellite imagery and GIS used for cost-effective and timely information. Deep learning methods, such as the Convolutional

Neural Network (CNN), have been applied to urban forest studies with success. However, challenges such as the availability of high-resolution imagery and lack of publicly available training datasets remain. Mapping urban tree canopy and land cover changes is crucial in researching urban forest loss, which is occurring in the US due to urbanization and natural disturbances. Urban forests are becoming increasingly important as the US becomes more urbanized, with 36 million trees being lost annually and urban populations outpacing national growth.

Therefore, there are three objectives in this dissertation project: 1) extend the application of deep learning in urban tree canopy mapping in a proof-of-concept pilot study, and 2) assess the effectiveness and robustness of a novel transfer learning based deep learning framework in urban land cover mapping in a desert city and (3) to better understand urban forest structure and its underlying socioeconomic determinants in fast-urbanization regions, supplying a scientific basis for green infrastructure planning and toward a more sustainable and equitable urban development.

1.5 Organization of the dissertation

Chapter 1 gave the introduction of urban forest, basic deep learning concepts and related literature reviews. Moreover, Chapter 1 provides a quick glance at the data and methods used in the dissertation.

Chapter 2 discussed an extend application of a deep learning architecture to urban tree canopy mapping. I applied a state-of-the-art airborne image dataset provided by the International Society for Photogrammetry and Remote Sensing (ISPRS) to the U-net model to extract the urban tree canopy. Then, I tested the effectiveness of the U-net in comparison with the object-based image analysis (OBIA) approach. Finally, to further assess the

performance and robustness of the U-net, a total of four experiments performed at different scales (16 cm, 32 cm, 50 cm, and 100 cm). Meantime, four trained U-net models will be exported for further application in the second chapter.

Chapter 3 introduced a novel proposed approach for land cover mapping integrated with deep learning and transfer learning. The study uses a U-Net, a semantic segmentation neural network architecture, and fine-tuning method to apply transfer learning to create a framework for automatically classifying land cover features in Phoenix.

Chapter 4 conducted a study to explore the socio-economic determinants of urban forest structure in a desert city. This chapter explore urban forest patterns and socio-economic inequity in a desert city, Phoenix, Arizona, understand the relationships between urban forest characteristics and socio-economic drivers at varying spatial scales and compare the impacts of different urban forest characteristics on socio-economic status.

Chapter 5 discusses and concludes with the findings of the whole dissertation and highlights the research significance and possible pathways for studies in the future.

Chapter 2 Application and evaluation of a deep learning architecture to urban tree canopy mapping

Abstract

An urban forest is a dynamic urban ecosystem that provides critical benefits to urban residents and the environment. Accurate mapping of urban forests plays an important role in greenspace management. In this study, I apply a deep learning model, the U-net, to urban tree canopy mapping using high-resolution aerial photographs. I evaluate the feasibility and effectiveness of the U-net in tree canopy mapping through experiments at four spatial scales—16 cm, 32 cm, 50 cm, and 100 cm. The overall performance of all approaches is validated on the ISPRS Vaihingen 2D Semantic Labeling dataset using four quantitative metrics, Dice, Intersection over Union, Overall Accuracy, and Kappa Coefficient. Two evaluations are performed to assess the model performance. Experimental results show that the U-net with the 32-cm input images performs the best with an overall accuracy of 0.9914 and an Intersection over Union of 0.9638. The U-net achieves state-of-the-art overall performance in comparison with the object-based image analysis approach and other deep learning frameworks. The outstanding performance of the U-net indicates a possibility of applying it to urban tree segmentation at a wide range of spatial scales. The U-net accurately recognizes and delineates tree canopy for different land cover features and has great potential to be adopted as an effective tool for high-resolution land cover mapping.

2.1 Introduction

Urban forests are an integral part of urban ecosystems. They provide a broad spectrum of perceived benefits, such as improved air quality, lower surface and air temperatures, and reduced greenhouse gas emissions (Buyantuyev & Wu, 2010; Dwyer & Miller, 1999;

Loughner et al., 2012; Nowak & Crane, 2002). Additionally, urban trees can improve human mental health by adding aesthetic and recreational values to the urban environment (Pandit et al., 2014; Payton et al., 2008; Ulmer et al., 2016). The last decade has witnessed a dramatic decline of urban tree cover. From 2009 to 2014, there was an estimated 70820 hectares of urban tree cover loss throughout the United States (*Cities and Communities in the US Losing 36 Million Trees a Year*, n.d.). A better understanding of tree canopy cover is more important than ever for the sustainable monitoring and management of urban forests.

Urban tree canopy, defined as the ground area covered by the layer of tree leaves, branches and stems (Grove et al., 2006), is among the most widely used indicators to understand urban forest patterns. Conventionally, field-based surveys are conducted to manually measure the area of urban vegetation (Fuller & Gaston, 2009; King & Locke, 2013; Loughner et al., 2012). The field-based surveys are mostly conducted by regional forestry departments and various research programs, and many of the field surveys are highly costly and labor-intensive (*Tree Cover % — How Does Your City Measure Up?*, n.d.). The availability of digital images and advances in remote sensing technologies provide unique opportunities for effective urban tree canopy mapping (Fan et al., 2019b; M. Li et al., 2014). For instance, the Moderate Resolution Imaging Spectroradiometer–Vegetation Continuous Field (MODIS–VCF) product provides global coverage of percent tree canopy cover at a spatial resolution of 500 meters (*LP DAAC - MODIS Overview*, n.d.). This product is one of the most adopted vegetation map products to study global tree canopy trends and vegetation dynamics. Other satellite images such as the Landsat imagery (*The Thematic Mapper « Landsat Science*, n.d.) and the *Satellite Pour l' Observation de la Terre* (SPOT) imagery are frequently used for studying tree canopy cover at local and regional scales (Baeza & Paruelo, 2020; Ferri et al.,

2014; *SPOT - CNES*, n.d.; Tran et al., 2017). Very high-resolution satellite images have increasingly been employed in forest management and mapping at the meter and sub-meter levels. The National Agriculture Imagery Program (NAIP) product coupled with airborne Light Detection and Ranging (LiDAR) is a popular data source for fine-scale urban forest mapping and analysis (Alonzo et al., 2014, 2016; MacFaden et al., 2012).

Urban land cover mapping has largely benefited from the development of novel image classification algorithms (MacFaden et al., 2012; Ronda et al., 2017). Yang et al. (2003) applied a decision tree approach to agriculture mapping over Quebec, Canada using hyperspectral satellite images (Yang et al., 2003). Zhu and Blumberg (2002) applied the support vector machine (SVM)-based classification method to the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data for mapping urban land cover features in Beer-Sheva, Israel (G. Zhu & Blumberg, 2002). Ahmed and Akter (2017) employed the K-means approach to analyze the land cover changes in the southwest Bengal delta, Bangladesh (Ahmed & Akter, 2017).

Conventional pixel-based classification methods have their drawbacks because they solely rely on the spectral signatures of individual pixels (Myint et al., 2011). Wang et al. (2007) compared the performance of three pixel-based methods, i.e., the iterative self-organizing data analysis technique (ISODATA), maximum likelihood, and the minimum distance method, with an object-based approach in mapping water and buildings in Nanjing, China. They concluded that compared to the pixel-based methods, the object-based approach yielded much better results when working with IKONOS images with a spatial resolution of 4 meters (P. Wang et al., 2007). Object-based image analysis (OBIA) works by first grouping spectrally similar pixels into discrete objects and then employing a classification algorithm to

assign the class labels to different objects. The OBIA has become a commonly used approach to urban land cover mapping, and its success has been evidenced in many studies (De Luca et al., 2019; Hossain & Chen, 2019; X. Li et al., 2014; MacFaden et al., 2012). Despite its effectiveness and popularity, the OBIA approach suffers from several limitations. For example, it has a heavy reliance on expert experience and local knowledge in the process of determining the most appropriate scales, parameters, functions, and classifiers. As a result, the robustness of the OBIA is compromised, making the classification results vary widely from case to case, from person to person.

Deep learning has recently received widespread attention among scholars from numerous fields (Pashaei et al., 2020; X. X. Zhu et al., 2017). The convolutional neural network (CNN), in particular, has been successfully applied in object detection and semantic segmentation (Ondruska et al., 2016; Rongqiang Qian et al., 2015). CNN is a powerful visual model that shows excellent performance in target recognition (Yuan et al., 2020). It supports the fast acquisition of rich spatial information from image features and automatically learns features with minimal need of expert knowledge (Krizhevsky et al., 2012b). Based on the CNN architecture, Long et al. (2015) proposed a fully convolutional network (FCN) architecture to perform dense prediction at the pixel level. Unlike CNN which mainly focuses on target recognition, the FCN supports multi-class classification with the capacity of assigning a class label to each pixel (Long et al., 2015a). Further, the FCN can capture details of image features and transfer these details to be recognized in the neural network (Pan et al., 2020a).

Built from the FCN, the U-net architecture was developed to further refine boundary delineation (Falk et al., 2019). The U-net was first applied to biomedical segmentation and

later applied to various fields with great success, including medical image reconstruction and speech enhancement (Çiçek et al., 2016; Esser et al., 2018; Macartney & Weyde, 2018). Compared to other neural networks, the U-net uses smaller datasets yet achieves better results (Feng et al., 2018; Ronneberger et al., 2015). Despite its adoption in biomedical fields, the application of the U-net in land cover mapping is still limited, with a few exceptions. For instance, Wagner (2020) proposed an instance-segmentation-based U-net architecture to delineate individual buildings using the WordView-3 imagery (Wagner et al., 2020). Another study employed the U-net to detect palm trees in Jambi, Indonesia, and Bengaluru, India, using the WorldView-2 images (Freudenberg et al., 2019). While this study achieved an overall accuracy of over 89%, the accuracy was assessed based on the location of tree stems rather than the extent of tree canopies.

Compared to other urban land cover types, urban tree canopy mapping is a challenging task because tree canopies can take a variety of shapes and forms depending on the age, size, and species of the trees. As trees and other types of green vegetation (e.g., shrubs/grass) are usually planted together, accurate segmentation of trees from other vegetation types is rather difficult due to their spectral similarity. In this study, we utilized the U-net architecture to map the urban tree canopy over Vaihingen, Germany. Coupling aerial images and the deep U-net model, this study aims to: (1) apply the U-net to urban tree canopy mapping using aerial photographs, (2) assess the performance of the U-net architecture at multiple spatial scales, and (3) test the effectiveness of the U-net in comparison with the OBIA approach.

2.2 Methods and data

2.2.1 Study Area and Data

This study utilized an image dataset published by the International Society for Photogrammetry and Remote Sensing (ISPRS). Images were taken over Vaihingen, Germany in 2013. The dataset contains 33 patches, each of which consists of an orthophoto and a labeled ground truth image. The orthophotos have three bands, near-infrared (NIR), red, and green with a spatial resolution of 8 centimeters (cm).

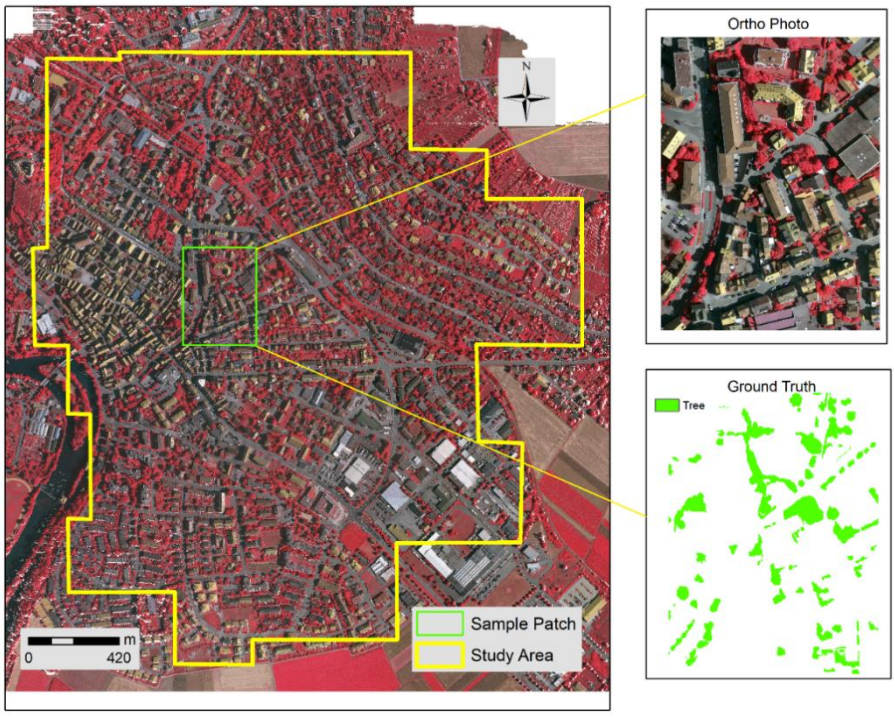


Figure 2.1 Study area located in Vaihingen, Germany.

2.2.2 U-net Architecture

The U-net architecture was designed for boundary detection and localization built from the FCN architecture (Ronneberger et al., 2015). It can be visualized as a symmetrical, U-shaped

process with three main operations, convolution, max-pooling, and concatenation. Figure 2.2 gives a visual demonstration of the U-net architecture for tree canopy segmentation.

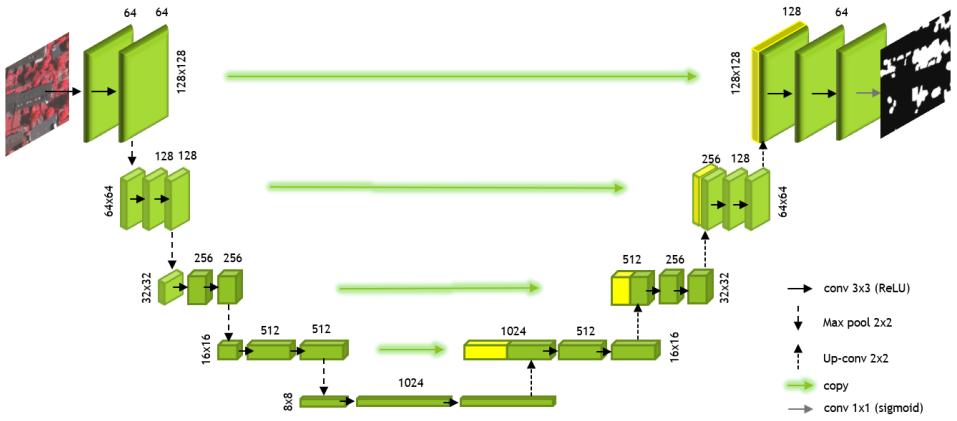


Figure 2.2 The U-net architecture for tree canopy segmentation.

The U-net architecture consists of two paths, a contracting path as shown on the left side of the U-shape, and an expansive path as shown on the right side of the U-shape (Figure 2.2). In the contraction path, the black solid arrows refer to the convolution operation with a 3 x 3 kernel (conv 3x3). With the convolution, the number of channels increased from 3 to 64. The black dash arrows pointing down refer to the max-pooling operation with a 2 x 2 kernel (Max pool 2x2). In the max-pooling operation, the size of each feature map was reduced from 128 x 128 to 64 x 64. The preceding processes were repeated four times. At the bottom of the U-shape, an additional convolution operation was performed twice.

The expansive path restores the output image size from the contraction path to the original 128 x 128. The black dash arrows pointing up refer to the transposed convolution operation, which increases feature map size while decreasing channels. The green arrows pointing horizontally refer to a concatenation process that concatenates the output images from the previous step with the corresponding images from the contracting path. The concatenation

process combines the information from the previous layers to achieve a more precise prediction. The preceding processes were repeated four times. The gray solid arrow at the upper right corner refers to a convolution operation with a 1 x 1 kernel (conv 1x1) to reshape the images according to prediction requirements.

2.2.3 Model Training

A total of four experiments were performed, with the spatial scales of input images being downsampled from 8 cm to 16 cm, 32 cm, 50 cm, and 100 cm, respectively. Figure 2.3 shows the model training workflow for the 16-cm experiment. The same workflow was repeated in the other three experiments. First, the original image with an 8 cm resolution was cropped into tiles of 256 × 256 pixels. Then, both the training and test datasets were downsampled to 16 cm, resulting in an image size of 128 x 128. 90% of the tiles was used for training and 10% was used for testing. In the training process, 85% of the dataset was used for training and 15% was used for validation. We performed two evaluations to assess the model performance. In the first evaluation, the predicted output was first upsampled to 8 cm and then compared with the original 8-cm ground truth data. In the second evaluation, the ground truth data were resampled to the spatial resolution of the predicted output dataset before comparison. For example, to evaluate the performance of the 16-cm model, the ground truth data were downsampled to 16 cm before the accuracy assessment.

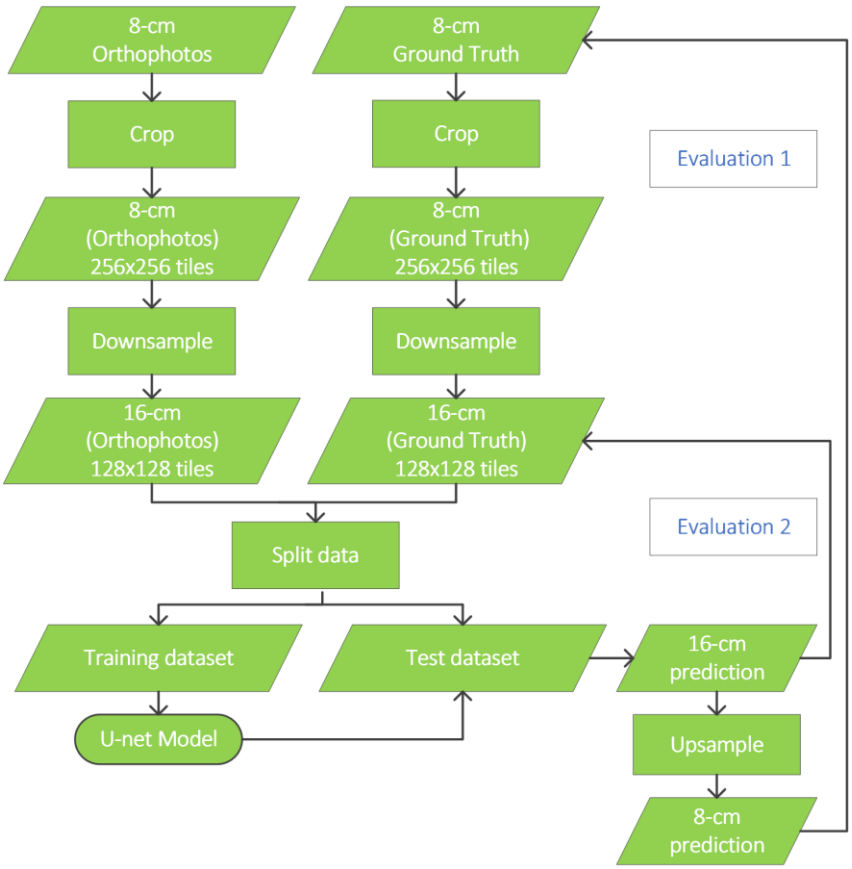


Figure 2.3 Flowchart of the U-net model for the 16-cm experiment.

2.2.4 Dice Loss Function

In deep learning, neural networks are trained iteratively. At the end of each training, a loss function is used as a criterion to evaluate the prediction outcome. In this study, we employed the Dice loss function in the training process. The Dice loss function can be calculated from the Dice similarity coefficient (DSC), a statistic developed in the 1940s to gauge the similarity between two samples (Dice, 1945). The Dice loss is given by

$$\text{Dice}_{\text{loss}} = 1 - \frac{2\sum_i p_i \cdot g_i}{\sum_i {p_i}^2 + \sum_i {g_i}^2} \tag{2.1}$$

where p_i and g_i represent the pixel values of the i^{th} pixel in the training output and the corresponding ground truth images, respectively. In this study, the pixel values in the ground truth images are 0 and 1, referring to the non-tree class and tree class respectively. The Dice loss was calculated during the training process to provide an assessment of the training performance at each iterated epoch. The value of the Dice loss ranges from 0 to 1, where a lower value denotes a better training performance.

Figure 2.4 shows how the Dice loss changes at each epoch for the four experiments during the model training process. As the number of epochs increases, the Dice loss drops rapidly and stabilizes at Epoch 99, 243, 97, and 300 for the 16-cm, 32-cm, 50-cm, and 100-cm experiment, respectively.

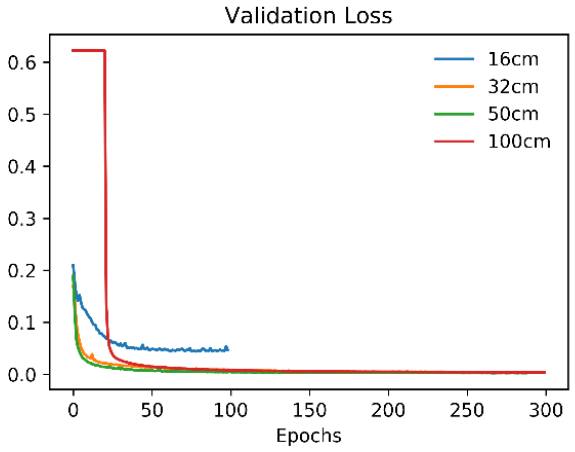


Figure 2.4 Dice loss for the four models in the training process.

2.2.5 Model Parameters and Environment

We used randomly selected tiles as our training datasets. For the experiment at each scale, 300 epochs with 8 batches per epoch were applied, and the learning rate was set at 0.0001 for all training models (Table 1). The Adam optimizer was utilized during the training process. We applied the horizontal shift augmentation to all images to increase the number of tiles in

the training dataset. The number of shift pixels was different at each scale to ensure enough training samples were included in each model. The U-net architecture, as well as the whole semantic segmentation procedure, were implemented on the Tensorflow and Keras deep learning framework in Python programming language. All other processing and analyses were carried out using open-source modules, including GDAL, NumPy, Pandas, OpenCV, Scikit-learn, among others. The deep learning network experimentation and modeling were executed on the Google co-lab platform.

Table 2.1 Model parameters for the four experiments.

	Number of epochs	Number of batches	Finish epoch	Learning rate	Number of tiles
16 cm	300	8	91	0.0001	12627
32 cm	300	8	42	0.0001	14887
50 cm	300	8	238	0.0001	6428
100 cm	300	8	133	0.0001	4683

2.2.6 Performance Evaluation

We used four accuracy metrics to evaluate the performance of the U-net model, including the overall accuracy (OA), the DSC, the Intersection over Union (IoU) and the kappa coefficient (KC). All metrics were calculated based on a confusion matrix (Table 2.2), which records the percentage of pixels that are true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN).

Table 2.2 Confusion matrix for accuracy assessment.

		Ground Truth	
		Tree	Non-tree
Prediction	Tree	TP[a]	FP[b]
	Non-tree	FN[c]	TN[d]

[a] TP: true positive.
[b] FP: false positive.
[c] FN: false negative.
[d] TN: true negative.

The OA was computed as the percentage of correctly classified pixels (Equation 2.2). The DSC was a commonly used metric in semantic segmentation (Bertels et al., 2019). It was employed in Section 2.4 for calculating the loss function and was used here again for accuracy assessment (Equation 2.3). The IoU, also known as the Jaccard Index, represents the ratio of the intersection to union between the predicted output and ground truth labeling images (Equation 2.4) (Hamers, 1989). The KC, used widely in remote sensing applications, is a metric of how the classification results compare to values assigned by chance (Equation 2.5) (Stehman, 1996). All metrics range from 0 to 1 with higher scores indicating higher accuracies.

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2.2}$$

$$\mathrm{DSC} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2.3}$$

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2.4}$$

$$\mathrm{KC} = \frac{\mathrm{N} \times [(\mathrm{TN} + \mathrm{FN}) \cdot (\mathrm{TN} + \mathrm{FP}) + (\mathrm{TP} + \mathrm{FP}) \cdot (\mathrm{FN} + \mathrm{TN})]}{\mathrm{N}^2 - [(\mathrm{TN} + \mathrm{FN}) \cdot (\mathrm{TN} + \mathrm{FP}) + (\mathrm{TP} + \mathrm{FP}) \cdot (\mathrm{FN} + \mathrm{TN})]} \quad \text{where N is the number of pixels} \tag{2.5}$$

2.2.7 Object-based Image Classification

The output of the U-net model was further compared to that of the object-based classification, a popular and widely applied approach to fine-scale land cover mapping. The object-based classification involves two main steps, segmentation and classification. We first used the multiresolution segmentation to group spectrally similar pixels into discrete objects. The shape, compactness, and scale factors were set to 0.1, 0.7, and 150, respectively. Based on a series of trial-and-error tests, an object with a mean value in the near-infrared band greater than 120 and a maximum difference value greater than 0.9 is being classified as tree canopy. The object-based classification was implemented in the eCognition software package (*ECognition | Trimble Geospatial*, n.d.).

2.3 Results

2.3.1 Performance of the U-net Model at Multiple Scales

We performed two evaluations to assess the model performance. Evaluation 1 compared the predicted results with the 8-cm ground truth images. The accuracy metrics are shown in Table 2.3. All accuracy metric scores were higher than 0.91 except for those at the 100-cm scale. Consistently across all metrics, there was a score increase when the scale changed from 16 cm to 32 cm, followed by a slight decrease from 32 cm to 50 cm, and a drastic drop from 50 cm to 100 cm. The U-net model achieved the best performance on the 32-cm dataset and the worst on the 100-cm dataset. The highest metric score was 0.9914 (OA) and the lowest score was 0.7133 (IoU).

Table 2.3 Accuracy metric scores at the four scales in Evaluation 1

Scale	OA	DSC	IoU	KC
16 cm	0.9791	0.9550	0.9138	0.9411
32 cm	**0.9914**	**0.9816**	**0.9638**	**0.9770**
50 cm	0.9881	0.9741	0.9496	0.9664
100 cm	0.9324	0.8327	0.7133	0.7917

Evaluation 2 compared the predicted results with the ground truth images after adjusting to the spatial resolution of the predicted output. Table 2.4 shows the accuracy metric scores for Evaluation 2 at the four scales. First, all metric values were higher than those in Evaluation 1 regardless of scale. Second, all metrics were above 0.99 except for the 16-cm experiment. Similar to Evaluation 1, there was a substantial increase in all four metrics when the scale went from 16 cm to 32 cm. Different from Evaluation 1 though, the trend flattened out when the spatial resolution changed from 32 cm to 100 cm. Lastly, Evaluation 2 yielded much higher metric scores than Evaluation 1 at the 100-cm scale with the highest and lowest metric score of 0.9984 (OA) and 0.9934 (IoU), respectively.

Table 2.4 Accuracy metric scores at the four scales in Evaluation 2

Scale	OA	DSC	IoU	KC
16 cm	0.9798	0.9568	0.9171	0.9436
32 cm	0.9982	0.9962	0.9925	0.9952
50 cm	**0.9987**	**0.9972**	**0.9944**	0.9963
100 cm	0.9984	0.9967	0.9934	**0.9983**

2.3.2 Visual Evaluation of the U-net Performance

To visually assess the performance of the U-net model, we selected an example area with moderate tree canopy cover and compared the predicted output with the ground truth images. Figure 5 shows the selected area on an aerial photo (Figure 2.5(a)), ground truth images at 8 cm, 16 cm, and 100 cm (Figures 2.5(b), 2.5(d), and 2.5(f)), and predicted output at 16 cm and 100 cm (Figures 2.5(c) and 2.5(e)). Overall, both U-net predictions (Figures 2.5(c) and 2.5(e)) showed great consistency with the ground truth images (Figures 2.5(b), 2.5(d), and 2.5(f)). Specifically, the similar patterns in Figures 2.5(c) and 2.5(b) echoed the high accuracy scores at 16 cm in Evaluation 1 (Table 2.3). Because Figure 2.5(d) was downscaled to match the spatial resolution of Figure 5(c), there was a higher level of similarity between Figures 2.5(c) and 2.5(d) than that between Figures 2.5(c) and 2.5(b). This was corroborated by the higher accuracy scores in Table 2.4 than Table 2.3 at the 16-cm scale. A similar pattern was identified when comparing Figure 2.5(e) with Figures 2.5(b) and 2.5(f). The lower metric values at the 100-cm scale in Table 3 were in part due to the blurry canopy boundaries as seen in Figure 2.5(e) as a result of downsampling. A better result was identified comparing Figure 2.5(e) with Figure 2.5(f), consistent with the much higher accuracy scores at 100 cm in Table 2.3.

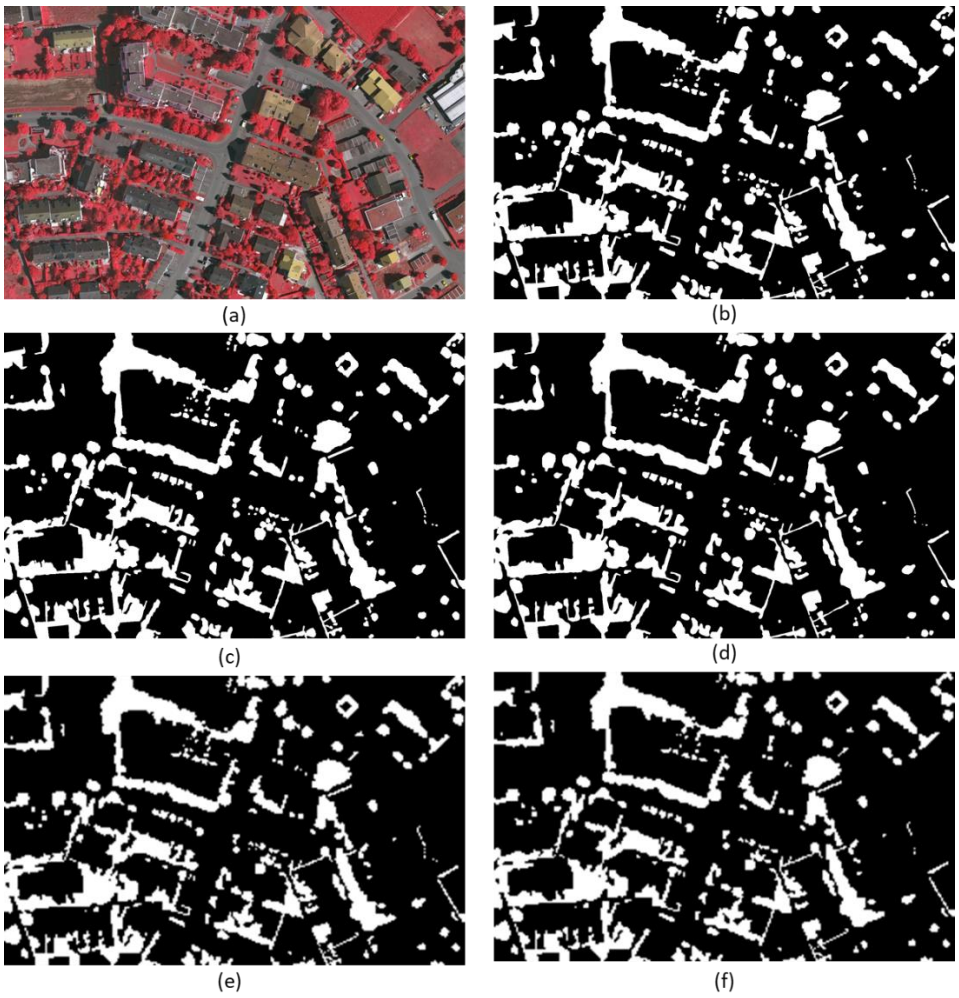


Figure 2.5 An example area of the tree canopy output map from the U-net: (a) original orthophoto, (b) 8-cm ground truth image, (c) 16-cm predicted tree canopy, (d) 16-cm ground truth image, (e) 100-cm predicted tree canopy, and (f) 100-cm ground truth image (white areas refer to tree canopy pixels; black areas refer to non-tree pixels).

2.3.3 Performance Comparison between the U-net and OBIA

Table 2.5 shows the accuracy scores of tree canopy mapping generated by the OBIA approach and the U-net model at the 16-cm scale. Both models were compared with the 8-cm ground truth data. For the OBIA, the highest metric score was 0.857 (OA) and the lowest was 0.489 (IoU). All the OBIA scores were significantly lower than the U-net scores. It is worthy

of note that even the highest score for the OBIA (OA: 0.857) was lower than the lowest score for the U-net (IoU: 0.9138), indicating that the U-net was superior to the OBIA in accurately mapping tree canopy at the 16-cm scale.

Table 2.5 Comparison of tree canopy mapping accuracy between OBIA and U-net (16 cm)

	IoU	DSC	OA	KC
OBIA	0.489	0.657	0.857	0.5681
U-net (16 cm)	**0.9138**	**0.9550**	**0.9791**	**0.9411**

Figure 2.6 shows a comparison between the predicted output of the U-net and the OBIA in reference to the 8-cm ground truth image. By visual inspection, the U-net output (Figure 2.6(c)) showed a much better consistency with the ground truth image (Figure 2.6(b)) compared to the OBIA output (Figure 2.6(d)). There were much more misclassified pixels in the OBIA output than that in the U-net output, especially when identifying trees from grass and shrubs (yellow rectangle in Figure 2.6(d)). Further, the U-net model successfully distinguished trees from buildings, while both the OBIA and the ground truth failed to do so (red rectangle in Figure 2.6(d)). Overall, the U-net outperformed both the OBIA and the ground truth image in accurately extracting tree canopy cover from other complex urban land cover features.

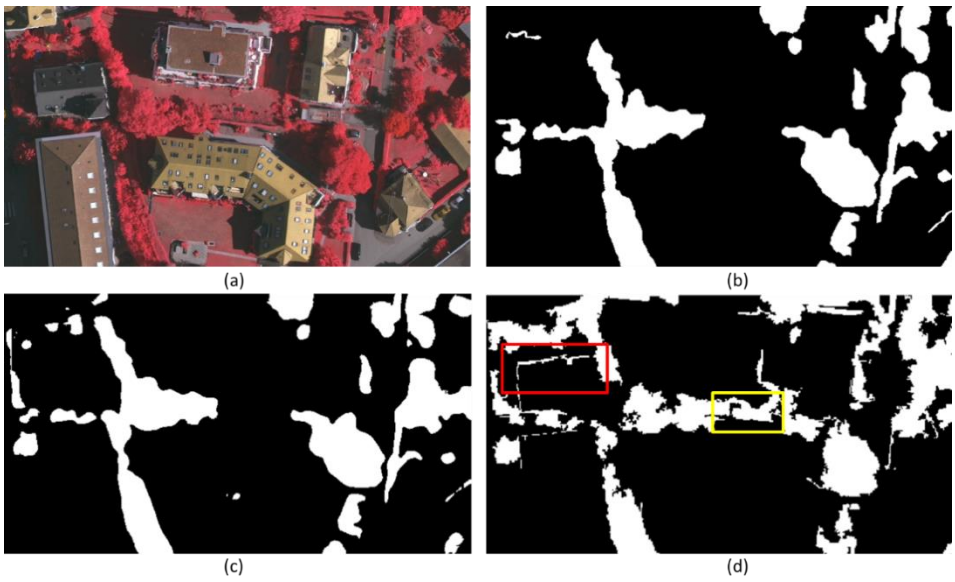


Figure 2.6 Comparison of the urban tree canopy segmentation between the U-net and OBIA: (a) original orthophoto, (b) ground truth image, (c) predicted output of the U-net, and (d) OBIA classification result (white areas refer to tree canopy pixels; black areas refer to non-tree pixels).

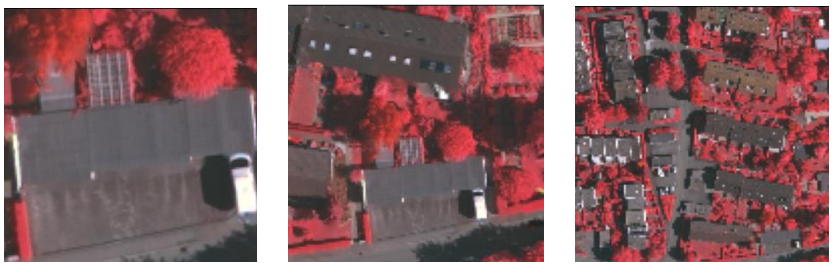
2.4 Discussion

2.4.1 Performance of the U-net in Urban Tree Canopy Mapping

In this study, we tested the effectiveness of the U-net in urban tree canopy mapping. We conducted the experiments at four different scales and performed two evaluations to assess the model performance from two different angles. Evaluation 1 (Table 2.3) compared the predicted output with the 8-cm ground truth. It aims to test the model performance operated on the datasets at different scales. Our results show that the U-net performed the best on the 32-cm dataset with an overall accuracy of 0.9914 (Table 2.3). While the 32-cm dataset is a coarser-resolution dataset compared to the 8-cm and 16-cm datasets, each image patch of the 32-cm dataset contains more geographic features than the other two. In deep learning, the spatial extent of the input is called “the receptive field”. It is defined as the size of the region

in the input that produces the feature (Araujo et al., 2019). In this context, the receptive field indicates how many land cover features can be perceived from an input patch. Figure 2.7 shows examples of input image patches of three scales: (a) 16 cm, (b) 32 cm, and (c) 100 cm. According to Figure 2.7(a), despite the higher spatial resolution, a 16-cm patch is too small to cover enough land cover objects such as buildings and trees with large crowns (Araujo et al., 2019). This in part explains the lower accuracies with the 8-cm and 16-cm datasets as they come with too small of a receptive field.

An overly large receptive field may also be problematic. Figure 2.7(c) shows an example of a 100-cm input patch. While it is large enough to include a significant number of land cover features, it comes with a cost of degradation of spatial details, resulting in a loss of locational accuracy especially at the tree edges. This is part of the reason why the DSC and KC values in the 100-cm experiment are much lower than the experiments performed at other scales. It is recognized that the size of the receptive field plays an important role in the training process of a deep learning neural network. Too small of a receptive field can limit the amount of contextual information while too large of a receptive field may cause a loss of spatial details and decline of locational accuracy (Qin et al., 2020). An optimal receptive field ensures a good number of land cover features go to the training model while retaining the spatial accuracy of the dataset. That is the case for the 32-cm dataset in our study, which achieves the highest accuracy scores and the best model performance (Table 3). It is worthy of note that even with the 100-cm dataset, the U-net is still able to achieve an OA of 0.9324 (Table 3), indicating the overall effectiveness of the U-net architecture in tree canopy mapping.



(a) (b) (c)

Figure 2.7 Examples of the input image patch for the U-net: (a) 16-cm input image patch, (b) 32-cm input image patch, (c) 100-cm input image patch.

We performed Evaluation 2 to assess the accuracy of the U-net models based on input and output of the same spatial resolution. This evaluation is essential because high-resolution ground truth data are not always attainable. Results show that the performance of the U-net architecture is exceptional based on the incredibly high accuracy metric scores (Table 2.4). All metric scores are above 0.99 for scales from 32 cm to 100 cm. These results indicate promising applications of the U-net architecture. An example is the National Agriculture Imagery Program (NAIP), which offers freely accessible satellite imagery across the United States at 1-m (100-cm) spatial resolution. From the results in Table 4, all metric scores are above 0.99 at the 100-cm scale, suggesting a highly effective and promising application of the U-net model to fine-scale land cover mapping based on NAIP data.

2.4.2 Comparison between U-net and OBIA

The OBIA has been a mainstream approach for high-resolution land cover mapping during the last decade. Myint et al. (2011) used the OBIA method to extract major land cover types in Phoenix from QuickBird images. In their study, the DSC score for the tree class was 0.8551 (Myint et al., 2011). Over the same study area, Li et al. (2014) developed another set of decision rules using the NAIP imagery and successfully raised the DSC score to 0.88 (X.

Li et al., 2014). Apart from the multispectral satellite imagery, Zhou (2013) supplemented the height and intensity from the LiDAR data and yielded a DSC score of 0.939 (Zhou, 2013). With a large number of existing studies using the OBIA approach, our study using the U-net model achieves a higher mapping accuracy than almost all the OBIA-based studies in the literature (DSC: 0.9816).

To further compare the performance of the U-net with the OBIA, we selected a sample study area with a variety of land cover types and applied the 16-cm U-net model and the OBIA approach to the same area. According to Table 5, all metric scores generated from the U-net model are higher than those from the OBIA approach. Even the highest OBIA score is lower than the lowest U-net score. Figure 6 provides a visual comparison between the OBIA and U-net output. Urban tree canopy mapping is challenging because trees are typically planted on grassland or closely adjacent to buildings (Figure 2.6(a)). The U-net model has the unique capacity of accurately distinguishing trees from grass and buildings (Figure 2.6(c)) while the OBIA approach is not as effective (Figure 2.6(d)).

One of the major downsides of the OBIA is a requisite for enough expert knowledge of the study area and the land cover types under investigation (Whiteside et al., 2011). In contrast, the U-net automatically learns the features in the study area without too much human intervention on parameter decisions (Liang & Li, 2016). Moreover, the U-net, as a deep learning network, comes with a very high level of automation with a minimal need of manual editing after the classification. This is a prominent advantage of the U-net over the OBIA because the accuracy of the OBIA depends to a great extent on post-classification manual editing which is time-consuming and labor-intensive. As the U-net is free of manual editing,

it has a great potential to become a mainstream mapping tool especially when dealing with large amounts of high-resolution data.

2.4.3 Comparison between the U-net and Other Deep Learning Methods

The test dataset of this study was made available from the ISPRS 2D Semantic Labeling Contest. The dataset contains orthophotos, digital surface model (DSM) images, and normalized DSM (nDSM) images over the Vaihingen city in Germany. The same set of data was utilized by a number of deep learning studies on tree canopy mapping. Table 2.6 lists a couple of these studies along with their methods, datasets, and DSC values on the tree class. Audebert et al. (2016) utilized the multimodal and multi-scale deep networks on the orthophotos. The DSC score on the tree class was 0.899 (Audebert et al., 2016). Sang and Minh (2018) used both the orthophotos and the nDSM images to train a fully convolutional neural network (FCNN). The classification accuracy was not improved in spite of adding the nDSM images on top of the orthophotos (Sang & Minh, 2018). Paisitriangkrai et al. (2015) took advantage of the entire dataset (all three data sets) to train a multi-resolution convolutional neural network, yielding a DSC value of 0.8497. Compared to the above studies using the Vaihingen dataset, our U-net model conducted at 16 cm, 32 cm, and 50 cm outperforms all of them with a DSC of above 0.95. Also note that adding DSM images fails to improve the overall model performance. Our best-performing model, the 32-cm U-net model, achieves a DSC of 0.9816. This surpasses the DSC of the contest winner (0.908), indicating the exceptional effectiveness of the U-net architecture on tree canopy mapping.

Table 2.6 Performance comparison of the U-net with other deep learning methods on the same dataset

Method	Dataset	DSC
Multimodal and Multi-scale Deep Networks (Audebert et al., 2016)	Orthophoto	0.899
Fully Convolutional Neural Network (FCNN) (Sang & Minh, 2018)	Orthophoto +nDSM	0.899
Multi-resolution Convolutional Neural Network (Paisitkriangkrai et al., 2015)	Orthophoto + nDSM +DSM	0.8497
32-cm U-net (this study)	Orthophoto	0.9816

Numerous studies have made modifications to the U-net architecture in an attempt to improve the model performance (Çiçek et al., 2016; J. Zhang et al., 2019; Z. Zhang et al., 2018). For instance, Diakogiannis et al. (2020) integrated the U-net with the residual neural network using high-resolution orthophotos and DSM images (Diakogiannis et al., 2020). The dataset they used was also published by the ISPRS 2D Semantic Labeling Contest (*2D Semantic Labeling - Potsdam*, n.d.) and was similar to the Vaihingen dataset used in this study. The modified framework achieved a DSC score of 0.8917 on the tree class compared to 0.9816 in this study (Table 3). While the U-net is simpler, it is more effective than the other, more complex, deep learning architecture. This is consistent with Ba and Caruana (2014) that while depth can make the learning process easier, it may not always be essential (Ba & Caruana, 2014). Choosing the most efficient and suitable neural network is a top priority to ensure the best overall performance of a deep learning framework.

2.5 Conclusions

Mapping urban trees using high-resolution remote sensing imagery is important for understanding urban forest structure for better forest management. In this study, we applied

the U-net to urban tree canopy mapping using high-resolution aerial photos. We tested the effectiveness of the U-net at four different scales and performed two evaluations to assess the model performance from two different angles. Evaluation 1 shows that the U-net performed the best on the 32-cm dataset, with an overall accuracy of 0.9914. The experiments conducted at four scales indicate the significance of an optimal receptive field for training a deep learning model. Evaluation 2 shows that the U-net can be used as a highly effective and promising tool for fine-scale land cover mapping with exceptional accuracy scores. Moreover, the comparison experiment shows the outstanding performance of the U-net compared to the widely used OBIA approach and other deep learning methods.

This study shows the utility of the U-net in urban tree canopy mapping and discusses the possibility of extending its use to other applications. A broad application of the U-net to high-resolution land cover mapping faces several challenges. First, as with any fine-scale land cover mapping tasks, the availability of freely accessible high-resolution imagery is an issue. U-net model training often requires satellite images with a spatial resolution of 1 meter or finer. It remains a challenge to acquire very high-resolution data for regions of interest at desired times. Second, the lack of publicly available training datasets poses another problem. Ground truth data are usually produced by local government or research institutions through field surveys or manual digitization. The process of generating accurate ground truth data is a complex and laborious task. A possible solution is to introduce the techniques and strategies in transfer learning. Approaches such as pre-training, fine-tuning, and domain adaptation can alleviate the dependence on large, labeled dataset. Therefore, integrating the U-net and transfer learning is a potential direction of future research.

Chapter 3 A Cross-scale Transfer Learning Framework for Deep Learning–based Urban Land Cover Mapping

Abstract

One popular topic in remote sensing is land cover classification in urban areas. The emergence of deep learning has provided a new approach to solving these problems. Deep learning models are able to express and process data with greater power and have shown high accuracy and precision rates in applications. Chapter 3 introduces the concept of deep learning and transfer learning (TL) to remote sensing image classification. The study uses a U-Net, a semantic segmentation neural network architecture, and fine-tuning to apply transfer learning to create a cross-scale transfer learning framework for automatically classifying land cover features in Phoenix based on U-net. The target dataset was National Agriculture Imagery (NAIP) data from 2015, and the source dataset was the Potsdam dataset from the International Society for Photogrammetry and Remote Sensing. The study also tested the effectiveness and feasibility of the proposed framework at different spatial scales from 15 cm to 100 cm. The highest overall accuracy achieved was 0.8201. To further explore the generalizability of the proposed framework, the study classified NAIP images of Las Vegas in 2016 and Phoenix in 2019. OA values in all test images are higher than 70%. The prediction results showed great potential for the framework to be used for land cover mapping, both spatially and temporally. In addition, the study compared the proposed method with two conventional classification methods, Maximum Likelihood and Object-Based Image Analysis. Cross-scale framwork performed better at object segmentation than the traditional methods, with the highest OA value of 85%. Overall, the proposed framework

outperformed the conventional approaches and provided a more effective way to classify urban areas without laborious efforts.

Keywords: Remote sensing; Deep Learning; Transfer Learning; U-net; Land cover classification

3.1 Introduction

Data quality, spatial resolution, revisit periods, and the area covered by remote sensing have all improved dramatically in recent decades. According to Emery and Camps, the ability to study the Earth from low Earth orbit and geostationary satellites has been steadily improving (Navalgund et al., 2007). Such an increase necessitates a major shift in how we use and manage remote sensing images. The enhanced spatial resolution allows for the development of unique methodologies, opening up new possibilities for remote-sensing picture processing and interpretation, allowing us to examine the ground surface in greater depth. However, the increased amount of data available has posed significant hurdles in managing image collection. One of the fundamental remote sensing tasks is image classification.

Image classification is essential for many practical remote sensing applications, such as urban planning, land management, land cover and land use mapping. Traditional classification methods categorize images based on spatial units, including pixels, moving windows, objects, and scenes (Myint et al., 2011; M. Wang et al., 2019). However, distinguishing complex land structures or patterns using limited rules is often tricky because traditional methods only involve low-level features in spectral and spatial domains in classification (Yuan et al., 2020). Thus, classification approaches using a considerable number of high-level features are desirable. Deep learning has been recently introduced to land cover mapping and has obtained optimal results due to its superiorities in multiscale and

multilevel feature extraction (Diakogiannis et al., 2020; Pan et al., 2020b; Yuan et al., 2020). Compared with the traditional rule-based and machine learning methods, the deep learning-based classification method has significant advantages in classification accuracy, especially in complex urban areas.

The emergence of deep learning provides a new approach to solving problems. Deep learning models have more powerful abilities to express and process data and have shown excellent accuracy and precision rates in applications. Building from the convolutional neural network (CNN) architecture, Long et al. (2015) proposed the fully convolutional network (FCN) to perform dense prediction at the pixel level. Unlike traditional CNN, which primarily focuses on target identification, the FCN supports multi-class classification with the capability of assigning a class label to each pixel (Long et al., 2015b). FCN can detect the details of image features and make these details recognized in the neural network (Pan et al., 2020b). Built from the FCN, the U-net was designed to refine boundary delineation further(Long et al., 2015b). The U-net was first used in biomedical segmentation and then successfully applied to various fields, including medical image reconstruction and speech enhancement (Çiçek et al., 2016, p. 21; Esser et al., 2018; Macartney & Weyde, 2018). The U-net employs smaller datasets than conventional neural networks but generates superior results(Feng et al., 2018; Ronneberger et al., 2015). Despite its application in biomedical fields, the adoption of the U-net in mapping land covers is currently limited, with a few exceptions. Besides image classification, CNNs have also achieved satisfactory accuracy in object detection and image segmentation (Chang et al., 2020).

In my previous study, I have already evaluated the feasibility and effectiveness of the U-net in tree canopy extraction (Z. Wang et al., 2021). My findings indicate that U-net achieves

state-of-the-art overall performance. The U-net's remarkable performance suggests applying it to urban tree mapping at various spatial scales. The U-net accurately identifies and delineates tree canopy for different land cover elements. It has a strong potential for use as an effective tool for high-resolution land cover mapping. There also remain some limitations. First, as with any fine-scale land cover mapping effort, the availability of freely accessible high-resolution images is a challenge. Aerial imagery with a spatial resolution of one meter or higher is frequently required for U-net model training. Another issue is the scarcity of publicly available training datasets. One possible solution to the abovementioned issues is adding transfer learning techniques and strategies.

Transfer learning (TL) is one of the most popular approaches in deep learning. We can say transfer learning is a machine learning method. The transfer of learned skills and knowledge from one learning situation to another. In remote sensing, transfer learning can be particularly useful for addressing the problem of limited training data, which is a common challenge due to the high cost and difficulty of collecting and annotating large amounts of remote sensing data. Transfer learning enables the use of pre-trained models that have been trained on large datasets, which can then be fine-tuned on smaller target datasets to improve their performance. Moreover, many pre-trained models, such as VGG, ResNet, and U-Net, have been pre-trained on large-scale image datasets and are available for transfer learning in remote sensing applications.

While transfer learning has been widely used in remote sensing applications, it is true that there may be relatively few studies that have used transfer learning with free-access remote sensing products. Free-access remote sensing products, such as those provided by NASA and ESA, typically have lower spatial resolution and limited spectral bands compared to

commercial high-resolution data. This may limit the potential for transfer learning since the pre-trained models may not be well-suited for the unique features and characteristics of the free-access remote sensing products. Moreover, transfer learning requires large and diverse datasets to train the pre-trained models, which may be limited in the case of free-access remote sensing products. This can make it challenging to fine-tune the pre-trained models on the target dataset, which can lead to overfitting or poor generalization performance.

Therefore, based on my previous findings in this study, I proposed a cross-scale transfer learning framework to extract the land cover features automatically. The findings and results will help with remote sensing image classification while processing large amounts of data.

3.2 Methods and Materials

3.2.1 Study Area

We selected a small area in Phoenix. Phoenix is a city located in the southwestern United States, in the state of Arizona. It is known for its hot, dry desert climate, with hot summers and mild winters. The city is situated in a valley surrounded by mountains, which provide some relief from the intense heat. It is a typical desert city with average high temperatures in the summer can reach over 100°F (38°C), while average lows in the winter are around 50°F (10°C). The city receives minimal rainfall, with the majority of its precipitation occurring during the monsoon season from July to September.

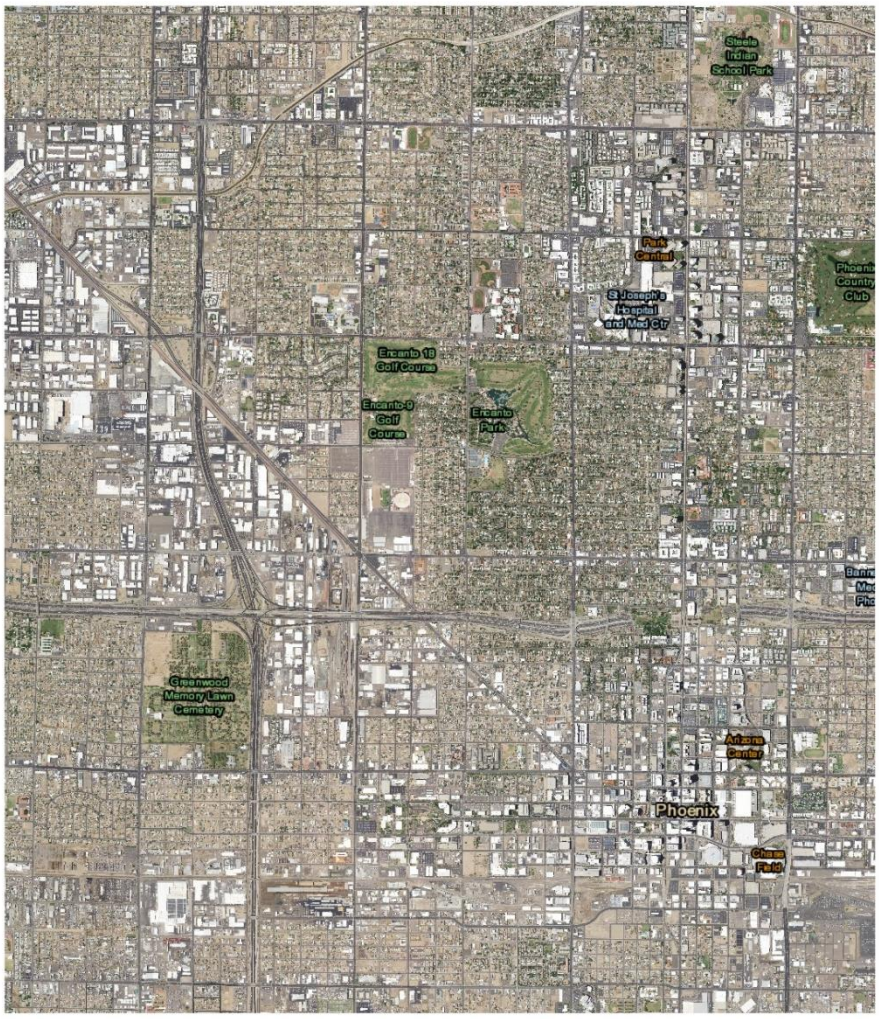


Figure 3.1 Study Area, Phoenix

Phoenix has experienced significant growth in recent decades, and understanding the changes in land-cover is crucial for managing the city's development. Land-cover changes can have significant impacts on the local environment, including increased heat island effect, loss of wildlife habitat, and altered water cycles. Studying land-cover can help identify and mitigate these impacts. Understanding land-cover can also inform decision-making around resource management, such as water allocation, urban planning, and conservation efforts. The desert climate of Phoenix is already challenging, and understanding how land-cover changes may

affect temperature, precipitation, and other climate variables is important for adaptation and mitigation efforts. Therefore, I chose Phoenix, a typical desert city as the study area.

3.2.2 ISPRS Dataset

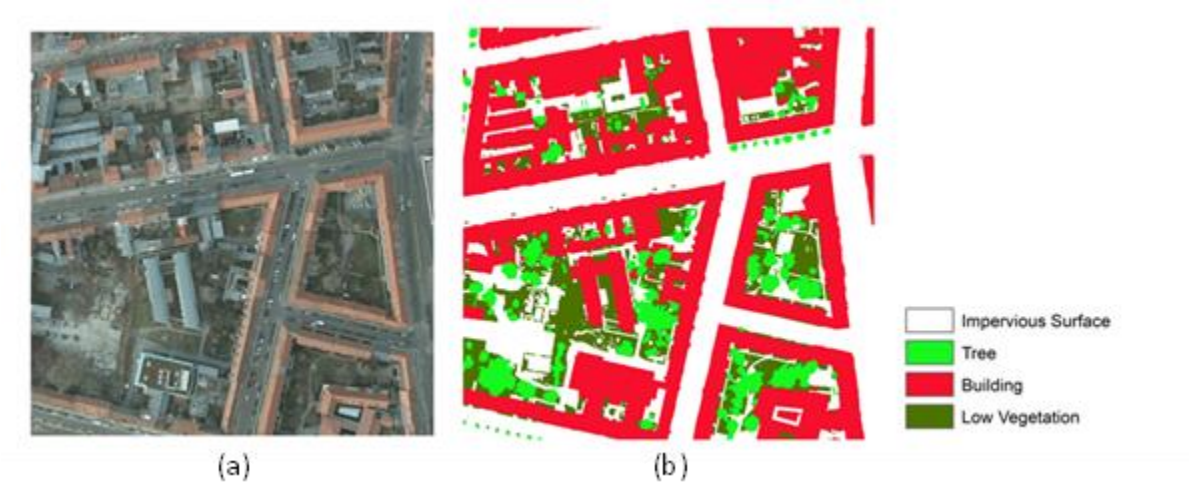


Figure 3.2 Example patches of the ISPRS Potsdam Dataset (a) orthophoto and (b) ground truth

The International Society published the source Potsdam image dataset for Photogrammetry and Remote Sensing (ISPRS). The ISPRS Potsdam dataset is a widely used benchmark dataset for evaluating remote sensing image analysis methods. The labeled Potsdam dataset was downloaded from ISPRS 2D Semantic Labeling Challenge 1 (*ISPRS Benchmark Test on Urban Object Detection and Reconstruction - ISPRS*, n.d.). All images have a spatial resolution of 5 cm. The average dimension is about 6000 x 6000. The ISPRS Potsdam dataset comprises 24 sets of orthophotos and their corresponding labeled images. Each orthophoto consists of four spectral bands, namely, red (R), green (G), blue (B), and near-infrared (IR). Figure 3.2 presents a comparison between the reference and corresponding orthophoto from the dataset. The dataset includes six land cover classes, which are impervious surface, building, low vegetation, tree, car, and background. Upon detailed inspection of the labels,

the car and background classes were merged into the impervious surface class, resulting in a final dataset of four classes, namely, impervious surface, tree, building, and low vegetation, as shown in Figure 3.2 (b). These four classes were used for training the models in the subsequent analysis.

3.2.3 Phoenix NAIP Land cover product

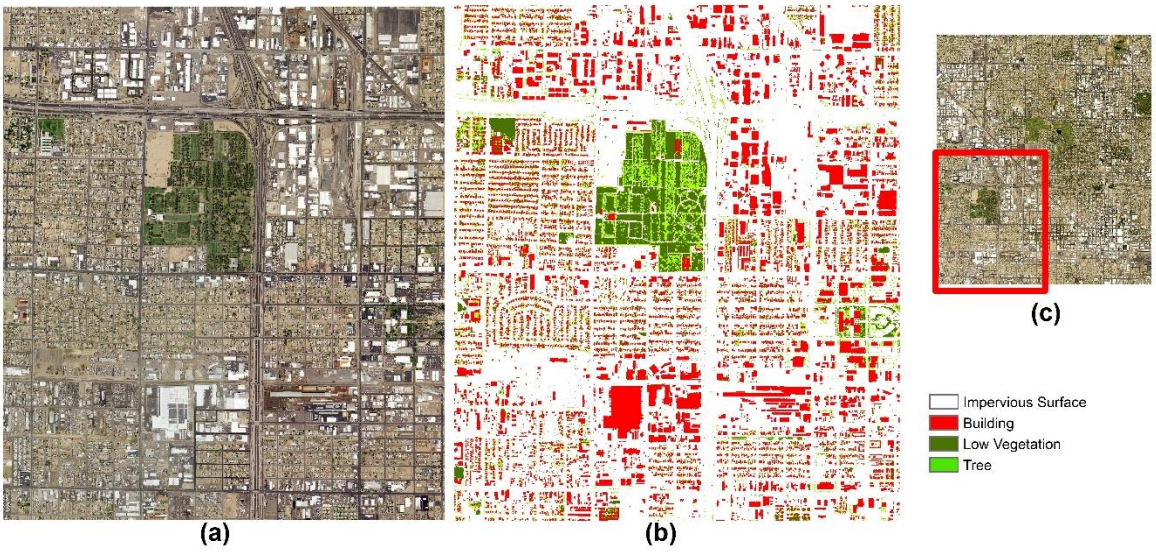


Figure 3.3 NAIP data in Phoenix in 2015 (a) NAIP data used for training and finetuning (b) Corresponding ground truth and (c) the location of the data used for training

The The target dataset used in this study was obtained from the National Agriculture Imagery Program (NAIP). The program is specifically designed to collect aerial imagery during the active agriculture season with high spatial resolution. NAIP imagery is available across the United States and serves as an essential data source for many land cover mapping applications.

Figure 3.2 presents the data used in this study, which is a selected area in Phoenix, Arizona. Figure 3.2 (a) shows the NAIP image from 2015, while Figure 3.2 (b) depicts the land cover

and land use (LULC) map generated by the Central Arizona-Phoenix Long-Term Ecological Research (CAPLTER) project. The CAPLTER project in central Arizona generated the LULC map using NAIP data, providing detailed information on the distribution of different land cover and land use types in the Phoenix metropolitan area. Figure 3.2 (a) represents one-quarter of Figure 3.2 (c), and the LULC data were clipped as labels for Figure 3.2 (c).

However, only one-quarter of the NAIP data and its corresponding label were used for training the finetuning process. The proposed framework predicts the entire Figure 3.2 (c) based on Figures 3.2 (a) and (b). The remaining LULC labels were also used to evaluate the performance of the model.

3.2.3 U-net

U-net was developed and first used for biomedical image segmentation. Its architecture is an encoder network followed by a decoder neural network. The U-net architecture consists of a contracting path, which gradually reduces the spatial resolution of the input image, and an expanding path, which gradually increases the spatial resolution of the output segmentation map. The contracting path is composed of convolutional and pooling layers, while the expanding path is composed of deconvolutional layers, which upsample the feature maps to the original spatial resolution. Unlike classification tasks, where the end result of the deep network is all that matters, semantic segmentation involves not just discrimination at the pixel level, but also a technique to project the discriminative features learned at different stages of the encoder onto the pixel space. I explained it in detail in Chapter 2.

3.2.4 Fine-tuning

Transfer learning is one of the most popular approaches in deep learning. We can say transfer learning is a machine learning method. The application of skills, knowledge that were

learned in one situation to another learning situation (Rostami et al., 2019). In this study, we can say that a model developed for a task that was reused as the starting point for a model on a second task. These two tasks normally have some similar properties. The two datasets used in this study have some differences. They were taken in different places; two cities have two different urban structures. However, the two datasets have more similarities. They are high-resolution orthophotos with four same spectral bands that are taken in urban areas with similar urban features, like trees, buildings, roads and grass. Those are the knowledge that can be transferred through the trained models.

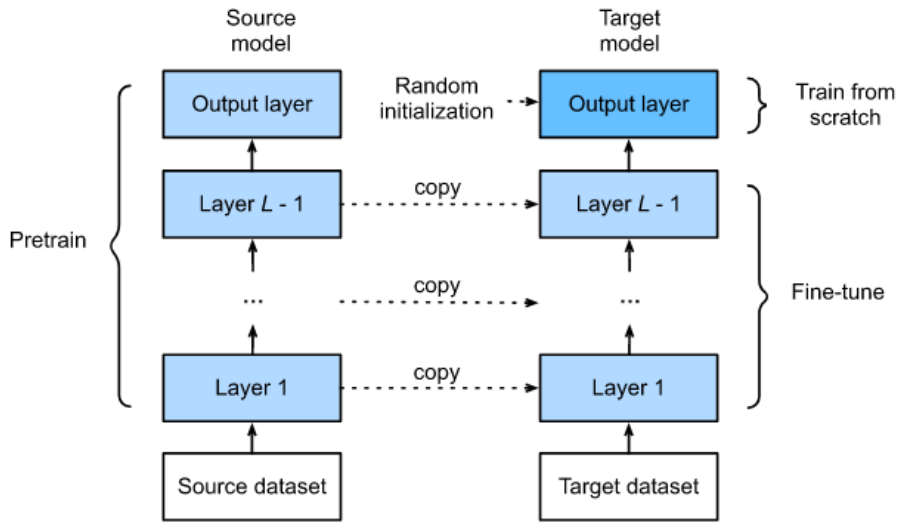


Figure 3.4 Fine-tuning workflow (*14.2. Fine-Tuning — Dive into Deep Learning 1.0.0-Alpha1.Post0 Documentation,* n.d.)

There are some methods of transfer learning, like fine-tuning and domain adoption. In this study, I adopted the fine-tuning approach. As shown in figure 4., fine-tuning consists of the following steps:

a. Pretrain a neural network model, i.e., the source model (the U-net in this study), on a source dataset (the ISPRS dataset in this study). This process involves training the

model on the source dataset to learn useful features that can be transferred to other tasks. The resulting model, called the source model, can then be fine-tuned for a specific target task using a smaller dataset. This approach can improve the model's performance on the target task by leveraging the knowledge learned from the source dataset.

b. Create a new neural network model, which will be the target model (still the U-net in this study). Except for the output layer, this duplicates all model designs and their parameters from the source model. I assume that these model parameters contain the information gained from the source dataset and that this knowledge is also applicable to the target dataset. I also assume that the source model's output layer is strongly related to the labels in the source dataset. Thus, it is not employed in the target model.

c. Add an output layer to the target model with the same number of outputs as the target dataset's categories. Then, at random, set the model parameters for this layer.

d. Train the target model on the target dataset (in this case, the NAIP data), such as a chair dataset. The output layer will be trained from scratch, while the other layers' parameters will be fine-tuned based on the parameters of the source model.

3.2.5 Model Training and Experiments

This study aimed to test the effectiveness of the U-Net for urban land cover mapping and evaluate the performance of the transfer learning-based U-Net through a series of four experiments and two evaluations. The first experiment involved training a U-Net model on the Potsdam ISPRS dataset, both with and without changes to the dataset's spectral bands. This allowed us to determine the importance of spectral information when training remote sensing data in deep learning models and establish baseline hyperparameters.

The second experiment assessed the effectiveness of the U-Net in land cover mapping at different scales. Experiment 3 involved transfer learning by using pre-trained models from the second experiment to classify NAIP images at different scales. In the fourth experiment, the models were fine-tuned using a limited NAIP dataset with labels and then applied to classify the rest of the NAIP data. Figure 3.4 represents the workflow of experiment 1-4.

Evaluation 1 compared the performance of the proposed cross-scale transfer learning framework with that of two commonly used remote sensing classification methods, OBIA and ML. Finally, evaluation experiments were conducted to assess the robustness and feasibility of the framework. In Evaluation 2, we further explored the effectiveness of our proposed framework by evaluating its performance over time and in different geographic areas. To achieve this, we downloaded remote sensing data from the same study area but in two different years and different study areas but in the same year. The purpose of this evaluation was to assess the robustness of the framework for detecting land cover changes over time and to investigate its feasibility for use in areas with varying land cover characteristics.

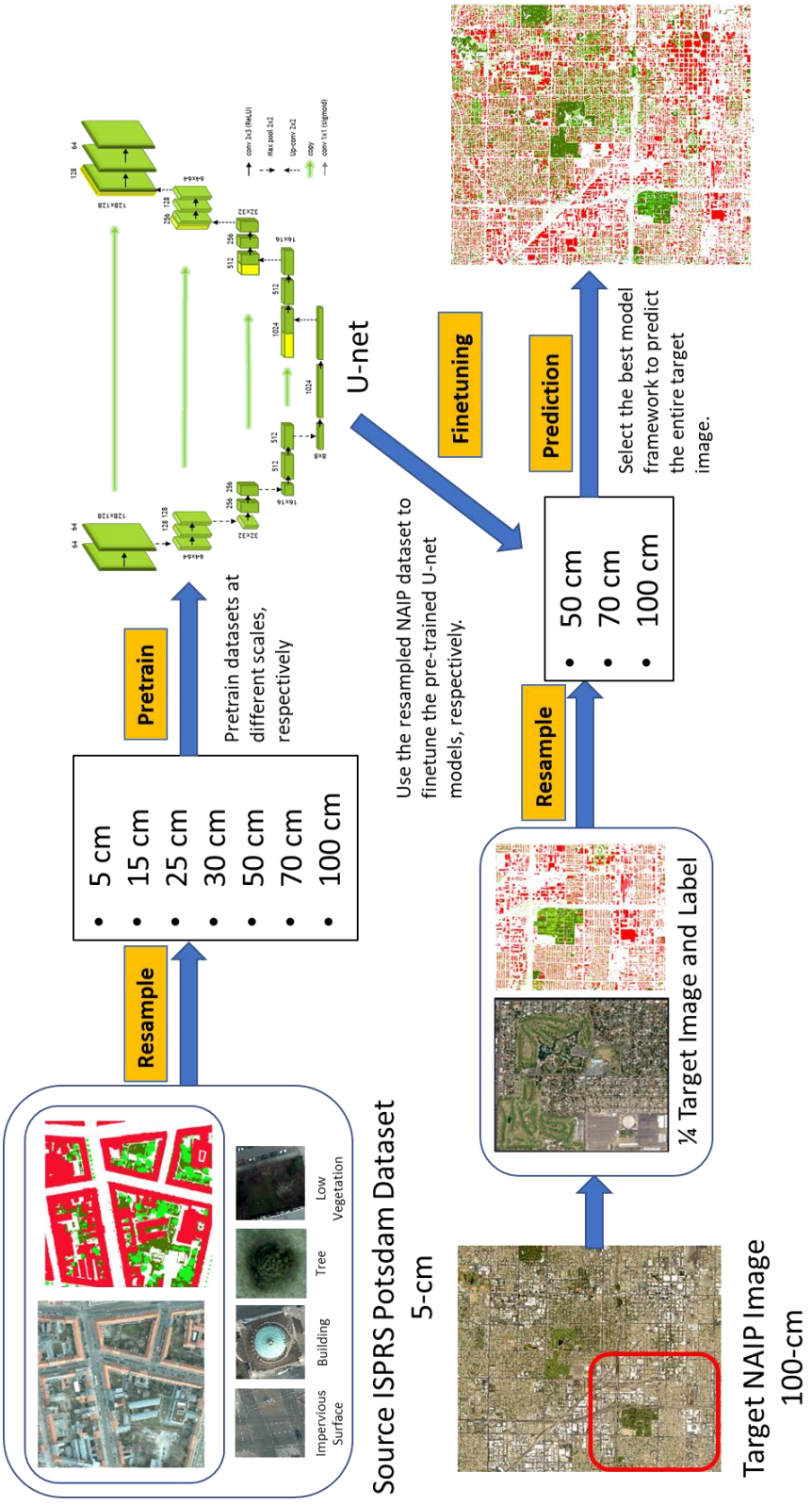


Figure 3.5 Flowchart of the cross-scale framework

3.2.6 Experiment Setup

I used randomly selected tiles as our training datasets. The original image has a 5-cm spatial resolution. The patch size is 128x128. 90% of the tiles were used for training, and 10% were used for testing. In the training process, 85% of the dataset was used for training, and the remaining 15% was used for validation. Then, 300 epochs with 8 batches per epoch were applied, and the learning rate was set at 0.0001 for all training models. All other processing and analyses were carried out using open-source modules and, including GDAL, NumPy, Pandas, OpenCV, Scikit-learn, among others. The deep learning network experimentation and modeling were executed on the Google co-lab platform. The evaluation metrics are dice coefficient (DSC) and overall accuracy (OA).

3.3 Results

3.3.1 Experiment 1

Table 3.1. shows the basic hyperparameters of the first-step model. This experiment aimed to train a baseline model to find the best hyperparameters. And the hyperparameters won't be changed in the following model training experiments. Batch size defines the number of samples we use in one epoch to train a neural network. Batch size controls the accuracy of the estimate of the error gradient when training neural networks (Schmeiser, 1982). It is 64 in the trained model. Learning rate is a hyper-parameter that controls the weights of our neural network with respect to the loss gradient. It defines how quickly the neural network updates the concepts it has learned (Smith et al., 2017). The learning rate is 0.00004. An optimizer is a function or an algorithm that modifies the attributes of the neural network, such as weights and learning rate. Thus, it helps in reducing the overall loss and improve the accuracy (Cochocki & Unbehauen, 1993). Adam optimizer was applied while training the U-net. Data

augmentation is a technique of artificially increasing the training set by creating modified copies of a dataset using existing data. I applied vertical and horizontal flip (V/H flip), brightness and contrast adjustment, and grid distortion to increase the number of tiles in the training dataset U-net was trained with randomly cropped patches of size 128 × 128. The spatial resolution of all images is 5 cm. For the trained baseline U-net model, the OA value is 0.7910.

Table 3.1 Model parameters and metric scores of 5-cm U-net model

Batch size	64
Learning rate	1.00E-04
Optimizer	Adam
Data Augmentation	V/H flip BrightnessContrast GridDistortion
Patch size	128×128
Resolution (cm)	5
Training DSC	0.8972
Validation DSC	0.8253
OA	0.7910

Table 3.2-3.4 presents the comparative performance evaluation of U-net models using different band combinations as input for land cover classification. Models that included the infrared (IR) band achieved better results in classifying Low vegetation and Tree. In fact, all the metric scores of Low vegetation and Tree in the RGIR model and RGBIR model are higher than those of the RGB model. These findings confirm the assumption that the accuracy of vegetation class should be higher in the model with the input of the IR band since vegetation is more sensitive to the IR band. Conversely, the RGB model performed

better in the building class with the highest recall and dice scores of 0.9176 and 0.8502, respectively. The highest OA value was achieved by the RGBIR model with a value of 0.7910. The results imply that remote sensing images with more bands can provide more information, which is reasonable. Therefore, the RGBIR model was used as the baseline model for the subsequent experiments. It is worth mentioning that this study's findings underscore the importance of band selection in U-net models for land cover classification.

Table 3.2 Accuracy metric scores of each class (RGIR model)

Class	Precision	Recall	Dice
Impervious surface	0.8377	0.7494	0.7911
Building	0.7987	0.9016	0.8470
Low vegetation	0.6486	0.8289	0.7278
Tree	0.7754	0.7730	0.7742
OA		0.7791	

Table 3.3 Accuracy metric scores of each class (RGB model)

Class	Precision	Recall	Dice
Impervious surface	0.8439	0.7419	0.7896
Building	0.7919	0.9176	0.8502
Low vegetation	0.6431	0.8045	0.7148
Tree	0.7498	0.7635	0.7566
OA		0.7750	

Table 3.4 Accuracy metric scores of each class (RGBIR model)

Class	Precision	Recall	Dice
Impervious surface	0.7956	0.7414	0.7675
Building	0.7907	0.9105	0.8464
Low vegetation	0.6651	0.8044	0.7282
Tree	0.7627	0.7938	0.7779
OA		0.7910	

3.3.2 Experiment 2

Table 3.5 exhibits the relative parameters and results of U-net models implemented at various scales. ISPRS resolution refers to the spatial resolution of the image input. In training a neural network, stride refers to the number of pixels the filter (kernel) is shifted across the input volume or image when performing a convolution operation. The stride determines how much overlap exists between neighboring receptive fields. The stride can affect the output size of a layer and the computational efficiency of the network. A larger stride reduces the spatial size of the output volume and thus may lead to a faster computation, but it may also result in a loss of information due to reduced overlap between receptive fields. Training size, test size, and validation size refer to the number of patches used for training, test, and validation, respectively. DSC stands for Dice Similarity Coefficient. It is a widely used evaluation metric that measures the overlap between the predicted segmentation mask and the ground truth segmentation mask. The DSC ranges between 0 and 1, where 0 indicates no overlap between the predicted and ground truth masks, and 1 indicates a perfect match. During training, the DSC loss is computed for each image in the training set, and the gradients of the loss with respect to the model parameters are computed and used to update the parameters using an optimization algorithm such as stochastic gradient descent (SGD) or

Adam. By iteratively adjusting the parameters to minimize the DSC loss, the model learns to produce more accurate segmentation masks. OA refers to the overall accuracy to evaluate the overall performance of the model.

As expounded in Chapter 2, the effectiveness of U-net varied with scale. The study conducted in Chapter 2 showed that the U-net performed optimally on a 32-cm model with the original spatial resolution of 8 cm. In this section, I resampled the ISPRS Potsdam dataset across a broader range at 7 distinct scales, ranging from 5cm to 15cm, 25cm, 30cm, 50cm, 70cm, and 100cm, respectively. As the spatial resolutions of images decreased, the pixel size increased. To maintain a patch size of 128x128, I adjusted the stride size to ensure sufficient image tiles for training. Table 3.5 presents the results not only at various scales but also with distinct stride and training image patch numbers. The 15-cm model achieved the highest OA value, with an overall accuracy of 0.7947, which is 0.037 higher than the baseline model (5-cm model). Additionally, the 70-cm model also demonstrated a relatively good score, with an OA value of 0.7917. These models will be utilized for classifying NAIP images in experiment 4 and fine-tuning NAIP images in experiment 5.

Table 3.5 Model parameters and metric scores of models at different scales in experiment 3

ISPRS Resolution (cm)	Stride	Training size	Test size	Validation size	Train DSC	Validation DSC	OA
5	72	165336	17672	13254	0.8972	0.8253	0.7910
15	36	67416	2048	1536	0.9072	0.8230	0.7908
15	16	334176	2048	1536	0.9338	0.8218	0.7947
25	36	23064	800	600	0.9012	0.8068	0.7770
30	16	75264	512	384	0.9591	0.8038	0.7878
50	16	23064	200	150	0.9154	0.7685	0.7544
50	8	86400	200	150	0.9393	0.7727	0.7645
70	16	9600	128	96	0.9054	0.7525	0.7896
70	8	36504	128	96	0.9272	0.7561	0.7917
100	16	3456	72	54	0.7765	0.6634	0.6512
100	8	12696	72	54	0.8874	0.6988	0.7139
100	4	46464	72	54	0.9247	0.7104	0.7216

3.3.3 Experiment 3

Table 3.6 shows the results of experiment 3. Except for training the original 100-cm NAIP data, NAIP images were resampled from 100 cm to 50cm, 70cm as the data input, respectively. In table 3.6, the first column refers to the spatial resolution of the image that used in the pre-trained model from Experiment 2. The OA values presented in the last three columns correspond to the results obtained from applying the pre-trained models to the NAIP data at different resolutions without finetuning.

The results are not good. No model achieved an OA value better than 0.7. Even the best metric score is only 0.6366, achieved from the 5-cm U-net model with 50-cm NAIP input data. The lowest OA value is only 0.3824, achieved from the 50-cm U-net model with 70-cm NAIP input data. These results indicate that the pre-trained model, which was trained on a different dataset, does not fully capture the relevant features of the target dataset, leading to

reduced performance. This phenomenon, known as "dataset shift," can occur when the data distribution of the target dataset is significantly different from that of the original dataset. To overcome this issue, fine-tuning the pre-trained model on the target dataset can enable the model to adapt to the new data distribution, capture the relevant features, and improve the model's performance and generalization ability. Therefore, fine-tuning the pre-trained model on the target dataset is a crucial step in achieving better segmentation results.

Table 3.6 Model parameters and metric scores of models at different scales in experiment 4

ISPRS Spatial Resolution (cm)	Stride	OA on NAIP 100cm	OA on NAIP 70cm	OA on NAIP 50cm
5	72	0.5872	0.5879	0.6366
15	36	0.6272	0.5877	0.6484
15	16	0.5798	0.5631	0.6286
25	36	0.4578	0.4739	0.5214
30	16	0.4216	0.3886	0.4096
30	36	0.5547	0.5365	0.5992
50	8	0.3821	0.3942	0.4140
50	16	0.5352	0.5174	0.5767
70	8	0.3869	0.4021	0.4465
70	16	0.5378	0.4961	0.5246
100	8	0.5587	0.4972	0.5424
100	4	0.5032	0.4798	0.5382

3.3.4 Experiment 4

Table 3.7 shows the results of experiment 4. The experiment utilized the trained U-net models at different scales as pre-trained models and applied the NAIP data with different spatial resolutions to fine-tuning the models, respectively. The first column in Table 3.7 refers to the spatial resolution of the image that used in the pre-trained model from Experiment 2. The last three columns refer to the OA values of the finetuned models using

the NAIP data. The table includes six different ISPRS resolutions, ranging from 5 to 100 cm, three different NAIP resolutions ranging from 50 cm to 100 cm and different strides, ranging from 8 to 72.

Results indicate that the best OA value is 0.8245 while using the 15-cm trained U-net as a pre-trained model and 50-cm NAIP data fine-tuned. Compared with the above experiments, the best OA value is 0.79 while training the ISPRS dataset. However, in experiment 5, seven models achieved OA values higher than 0.80 and all models achieved OA values higher than 0.7. Interestingly, the table also shows that the performance of the models varies depending on the spatial resolution of the NAIP data. For example, some models achieved high OA values on NAIP data at one spatial resolution but not on another. This suggests that the spatial resolution of the input data is an important factor in the performance of the model.

Overall, these results suggest that the best model framework was achieved by pre-training with the 15-cm ISPRS dataset and fine-tuning with 50-cm NAIP data. This is the framework that I used in Experiment 5 and following evaluations.

Table 3.7 Model parameters and metric scores of models at different scales in experiment 5

ISPRS resolution (cm)	Stride	OA On NAIP 100cm	OA on NAIP 70cm	OA on NAIP 50cm
5	72	0.7431	0.7007	0.7327
15	36	0.8133	0.7961	0.8245
15	16	0.6926	0.7864	0.8213
25	36	0.7898	0.7502	0.7639
30	16	0.809	0.7899	0.8155
30	36	0.7017	0.7802	0.8147
50	8	0.7902	0.775	0.8019
50	16	0.712	0.7654	0.8001
70	8	0.7373	0.6969	0.7302
70	16	0.6724	0.7077	0.7324
100	8	0.7627	0.7357	0.7457

3.3.5 Evaluation 1

To further explore the effectiveness of the land cover classification framework proposed in this study, additional data in the same area but from a different year were downloaded and analyzed. Specifically, NAIP data for the year 2019 in Phoenix were obtained. Unlike the previous analysis, no labels were available for this dataset, so the best model achieved in the previous steps was used to classify the data. The spatial resolution of the NAIP data was resampled from 100 cm to 50 cm, and the resampled dataset was used as input into the best model. A selected area of the prediction results was examined to evaluate the accuracy and reliability of the classification framework for the 2019 dataset.

To provide a comparison of the classification framework's performance, the results were compared with two commonly used remote sensing classification methods, object-based

image analysis (OBIA) and maximum likelihood (ML). Both OBIA and ML methods require extensive manual labor to achieve accurate and reliable results. The comparison provides insights into the efficiency and effectiveness of the classification framework in relation to other widely-used classification methods.

Figure 3.6 presents the predicted land cover maps using the cross-scale framework proposed in this study, as well as two commonly used remote sensing classification methods, OBIA and ML. Figure 3.6 (b) shows the predicted map from the cross-scale framework, while Figure 3.6 (c) and (d) show the predicted maps from OBIA and ML, respectively. From the maps, it is clear that the segmentation of land cover features is evident. The cross-scale framework was able to classify more impervious surface and better recognized open land areas as impervious surfaces, which is important for urban land management and planning. ML, on the other hand, struggled to delineate building boundaries, resulting in irregularly-shaped buildings in Figure 3.6 (d). OBIA also had some classification errors, wrongly classifying some roads as buildings and some low vegetation as trees, which can be problematic for urban planning and management.

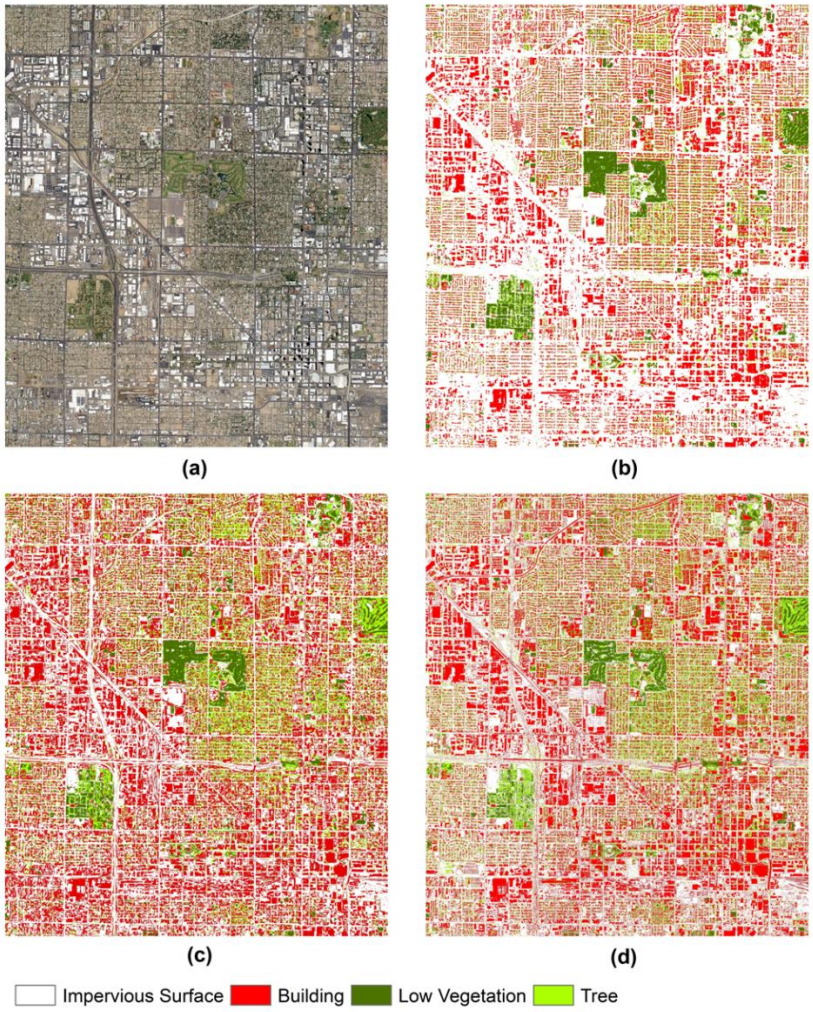


Figure 3.6 Comparisons of land cover classification performance between TL-Multiscale-U-net, OBIA and ML (a) NAIP data of the selected area in Phoenix in 2019, (b) Prediction result using the proposed framework (TL-Mutiscale-U-net) in this study, (c) Prediction result using OBIA method, (d) Prediction result using ML method.

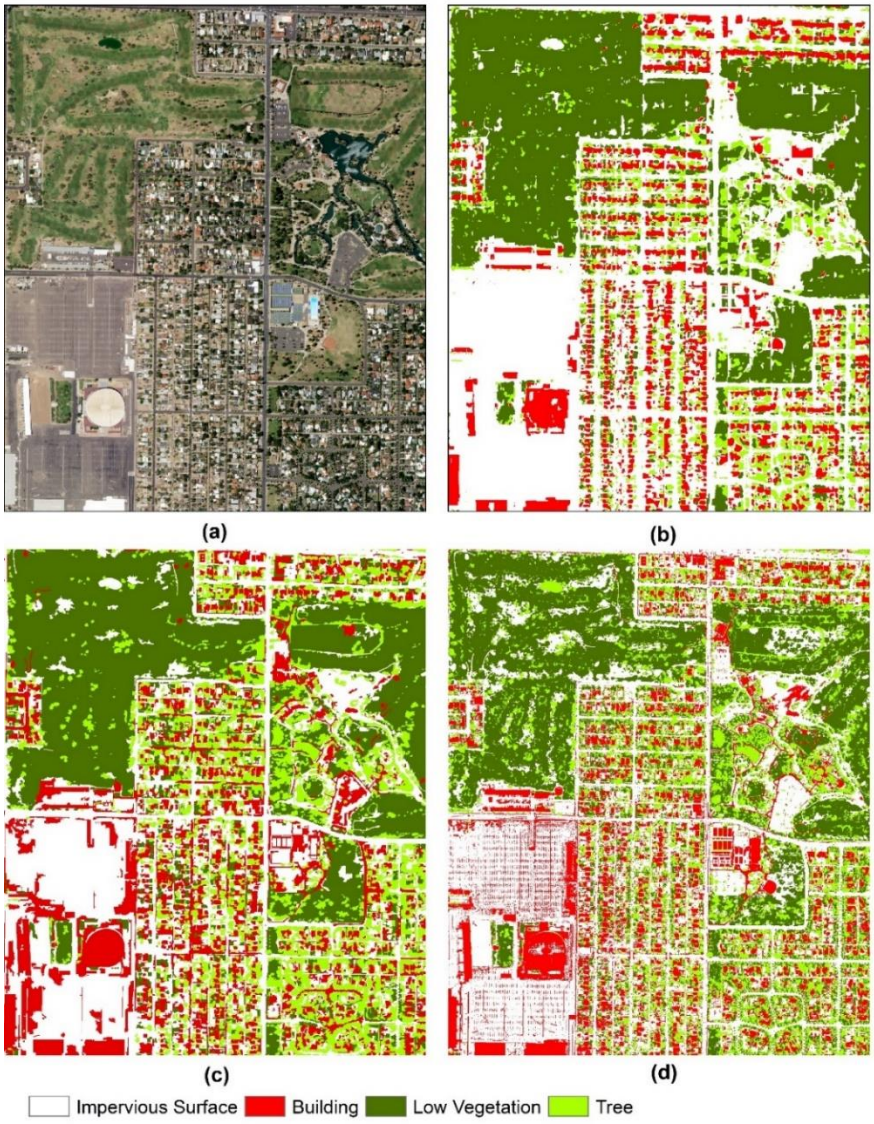


Figure 3.7 Comparisons of land cover classification performance between TL-Multiscale-U-net, OBIA and ML detailed in a selected neighborhood (a) NAIP data of the selected area in Phoenix in 2019, (b) Prediction result using the proposed framework (TL-Mutiscale-U-net) in this study, (c) Prediction result using OBIA method, (d) Prediction result using ML method.

Table 3.8 Confusion matrix and metric scores of different models

	Predicted Class	True Class							
		Impervious Surface	Building	Low Vegetation	Tree	Accuracy	Precision	Recall	F1
OBIA OA: 71.7%	Impervious Surface	362	45	8	8	0.748	0.86	0.65	0.74
	Building	161	211	5	7	0.774	0.55	0.8	0.65
	Low Vegetation	16	5	51	2	0.955	0.69	0.7	0.69
	Tree	14	3	9	93	0.957	0.78	0.85	0.81
ML OA: 72.8%	Impervious Surface	379	34	7	4	0.781	0.89	0.69	0.78
	Building	156	225	1	1	0.803	0.59	0.85	0.7
	Low Vegetation	2	0	45	26	0.944	0.62	0.62	0.62
	Tree	16	5	20	79	0.928	0.66	0.72	0.69
Cross-scale framework OA: 85%	Impervious Surface	506	37	9	17	0.89	0.89	0.92	0.9
	Building	39	217	2	18	0.894	0.79	0.82	0.8
	Low Vegetation	3	1	58	6	0.975	0.85	0.79	0.82
	Tree	5	9	4	69	0.941	0.79	0.63	0.7

Table 3.8 is the confusion matrix that shows the classification results for four land cover classes (Impervious Surface, Building, Low Vegetation, and Tree) using three different classification methods (OBIA, ML, and the cross-scale framework). The table includes various performance measures, including overall accuracy, precision, recall, and F1 score. The overall accuracy measures the proportion of correctly classified samples over the total number of samples. In this case, the cross-scale framework achieved the highest overall accuracy (85%), indicating that it had the best overall performance among the three methods.

The cross-scale framework outperforms the other two methods, OBIA and ML, in terms of overall accuracy (OA) for all classes except for the "Building" class. For the "Impervious Surface" class, The cross-scale framework achieved an OA of 89%, which is higher than the

OBIAs 75% and ML's 78%. Similarly, for the "Low Vegetation" class, the cross-scale framwork achieved an OA of 97.5%, which is significantly higher than OBIA's 95.5% and ML's 94.4%. For the "Tree" class, the proposed method achieved an OA of 94.1%, which is higher than OBIA's 95.7% but lower than ML's 92.8%. For the "Building" class, ML achieved the highest OA (80.3%), followed by The cross-scale framework (89.4%) and OBIA (77.4%). Additionally, looking at the precision, recall, and F1 score measures for each class, we can see that the cross-scale framework generally outperforms the other two methods. For example, for the "Impervious Surface" class, the cross-scale framework achieved a precision of 0.89, recall of 0.92, and F1 score of 0.9, which are higher than OBIA's precision of 0.86, recall of 0.65, and F1 score of 0.74, and ML's precision of 0.89, recall of 0.69, and F1 score of 0.78.

Overall, the results suggest that the cross-scale framework method that you proposed outperforms the other two methods, especially for classes such as "Impervious Surface" and "Low Vegetation." These findings highlight the effectiveness of your proposed method and demonstrate its potential for accurately classifying land cover using remote sensing data.

3.3.6 Evaluation 2

To further explore the effectiveness of the classification framework proposed in this study, additional data in the same area but from different years were downloaded and analyzed. Data in the study area were obtained for the years 2013 and 2019 (Figure 3.8 (a) and 3.88 (a)). The spatial resolution of the NAIP data was resampled to 50 cm, and the best model achieved in the previous steps was applied to the resampled data. This approach enabled the evaluation of the classification framework's performance on different datasets and over time. By using the best model trained on the original dataset, we can compare the performance of

the classification framework on new data with different characteristics. The results of the classification were analyzed to determine whether the framework can achieve high accuracy and provide reliable land cover classification for different years in the same area. The use of additional data provides a more comprehensive evaluation of the framework's effectiveness and improves its applicability for different time periods.

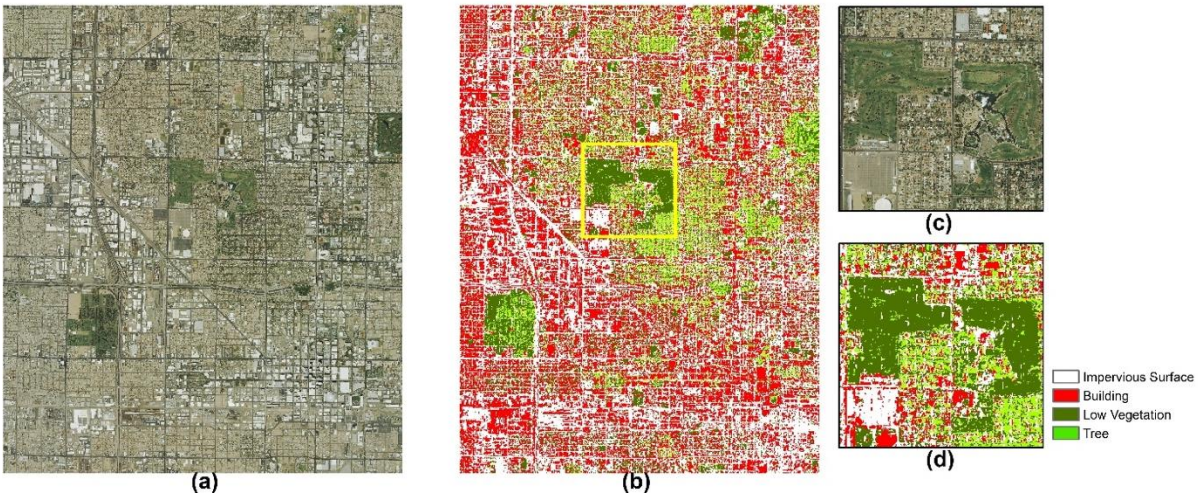


Figure 3.8 Land cover classification results of Phoenix in 2013: (a) NAIP data in Phoenix. (b) Corresponding predicted land cover map to (a). (c) NAIP data of a subset area. (d) Corresponding predicted land cover map to (c).

Since ground truth data for the 2013 and 2019 datasets were not available, a sample of 1000 points within the study area was randomly selected, and each point was manually labeled to create a ground truth dataset. This approach ensured that the accuracy of the classification framework could be evaluated using a reliable and representative ground truth dataset. Once the ground truth data were established, the confusion matrix was calculated to assess the classification framework's performance on the 2013 and 2019 datasets. The confusion matrix provided detailed information about the classification performance for each land cover class

and helped identify areas for improvement in the classification framework. The use of manual labeling to create the ground truth dataset ensured the accuracy of the evaluation and allowed for a more reliable comparison between the different datasets.

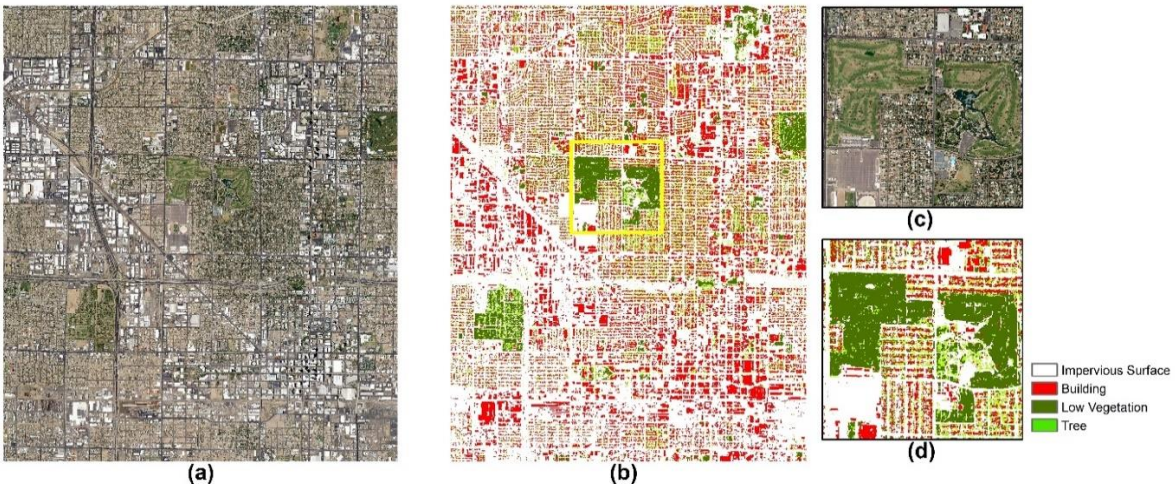


Figure 3.9 Land cover classification results of Phoenix in 2019: (a) NAIP data in Phoenix. (b) Corresponding predicted land cover map to (a). (c) NAIP data of a subset area. (d) Corresponding predicted land cover map to (c).

Table 3.9 is a confusion matrix that shows the classification results for four land cover classes (Impervious Surface, Building, Low Vegetation, and Tree) for two different years (2013 and 2019). The table includes various performance measures, including overall accuracy, precision, recall, and F1 score.

The overall accuracy of the classification model improved from 71.3% in 2013 to 85% in 2019, indicating an improvement in the classification performance over time. The "Impervious Surface" and "Low Vegetation" classes were well classified in both years, with high accuracy and precision values. However, the "Building" class was more difficult to classify in both years, with relatively low precision and F1 score values. The "Tree" class

was classified relatively well in both years, with high precision and recall values, although the precision was lower in 2019. Overall, the confusion matrix provides insights into the strengths and weaknesses of the classification model and can guide the selection of the most appropriate method for land cover classification.

Table 3.9 Confusion matrix and metric scores in different years

		True Class							
	Predicted Class	Impervious Surface	Building	Low Vegetation	Tree	Accuracy	Precision	Recall	F1
2013 OA:71.3%	Impervious Surface	410	81	50	3	0.767	0.75	0.81	0.78
	Building	82	149	16	5	0.806	0.59	0.62	0.61
	Low Vegetation	11	5	109	11	0.895	0.8	0.58	0.67
	Tree	6	5	12	45	0.958	0.66	0.7	0.68
2019 OA: 85%	Impervious Surface	506	37	9	17	0.89	0.89	0.92	0.9
	Building	39	217	2	18	0.894	0.79	0.82	0.8
	Low Vegetation	3	1	58	6	0.975	0.85	0.79	0.82
	Tree	5	9	4	69	0.941	0.79	0.63	0.7

To test the robustness in a different city, I chose the city of Las Vegas in the same year, 2015. Las Vegas is also a city on the floor of the Mojave Desert. It has a similar landscape and climate environment to Phoenix. Figure 3.10 shows the orthophoto and corresponding prediction results.

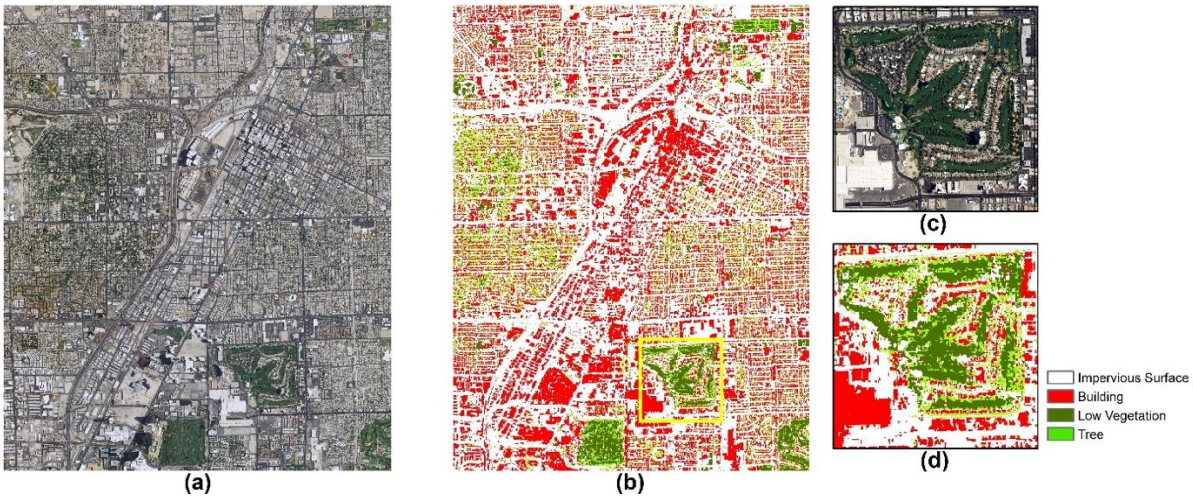


Figure 3.10 Land cover classification result of Las Vegas in 2015. (a) NAIP data in Las Vegas and corresponding classification result (b)

Table 3.10 is a confusion matrix shows the classification results for four land cover classes (Impervious Surface, Building, Low Vegetation, and Tree) in Las Vegas. The overall accuracy (OA) of the classification model is moderate at 75.92%. The "Low Vegetation" class was well classified with high accuracy, precision, and recall values, indicating that the classification model performed well for this class. The "Building" class has the lowest recall value, indicating that some of the building pixels were misclassified as other classes. The "Impervious Surface" and "Tree" classes were classified relatively well, with precision, recall, and F1 score values ranging from moderate to high. The confusion matrix provides insights into the strengths and weaknesses of the classification model and can guide the selection of the most appropriate method for land cover classification in Las Vegas.

Table 3.10 Confusion matrix and metric scores in Las Vegas

		True Class							
	Predicted Class	Impervious Surface	Building	Low Vegetation	Tree	Accuracy	Precision	Recall	F1
Las Vegas OA:75.92%	Impervious Surface	0	1	2	3	0.8182	0.81	0.85	0.83
	Building	442	60	21	25	0.8362	0.71	0.74	0.73
	Low Vegetation	56	219	8	24	0.9471	0.72	0.5	0.59
	Tree	6	6	38	2	0.9171	0.66	0.54	0.6

3.4 Discussion

3.4.1 Model Performance

The emergence of deep learning has provided a new approach to remote sensing image classification, particularly in the context of urban land cover mapping. This study introduced the concept of deep learning and transfer learning to remote sensing image classification and evaluated the effectiveness of the U-Net model in land cover mapping at different scales.

The first experiment explored the importance of spectral information when training remote sensing data in deep learning models. It showed that the inclusion of the infrared band in the model's input led to better classification results for low vegetation and trees, which are more sensitive to the infrared band. Conversely, the RGB model performed better in classifying buildings. This experiment highlighted the importance of band selection in U-Net models for accurate land cover classification.

The second experiment evaluated the effectiveness of the U-Net model in land cover mapping at different scales. The study found that the U-Net model performed optimally on a 32-cm model with the original spatial resolution of 8 cm. However, when the ISPRS Potsdam dataset was resampled across a broader range of seven scales, ranging from 5cm to 100cm,

the 15-cm model achieved the highest overall accuracy, with an overall accuracy of 0.7947. This experiment suggested that spatial resolution plays a critical role in the performance of the model.

The third experiment involved transfer learning by using pre-trained models from the second experiment to classify NAIP images at different scales. However, the pre-trained models did not capture the relevant features of the target dataset, leading to reduced performance. This phenomenon, known as "dataset shift," can occur when the data distribution of the target dataset is significantly different from that of the original dataset.

To overcome this issue, the fourth experiment fine-tuned the pre-trained models on a limited NAIP dataset with labels and then applied them to classify the rest of the NAIP data. The study found that the best model framework was achieved by pre-training with the 15-cm ISPRS dataset and fine-tuning with 50-cm NAIP data. This framework achieved an overall accuracy of 0.8245, which is higher than the accuracy achieved by training the ISPRS dataset and the accuracy achieved by using commonly used remote sensing classification methods.

In recent years, there have been several studies on deep learning-based land cover classification using remote sensing data. Wang et al. (2020) proposed a deep learning-based approach for land use classification from remote sensing imagery using a stacked autoencoder neural network. The authors used the deep neural network to extract high-level features from the remote sensing images and classify them into different land use categories. The results showed that the proposed method achieved an overall accuracy of 77.3%. Wang et al. (2019) used a transfer learning-based framework to classify urban land cover features in high-resolution remote sensing images. The authors used a pre-trained CNN on a large-scale dataset and fine-tuned it on a smaller dataset of high-resolution images. The results showed

that the proposed method achieved an overall accuracy of 79.67%. Another study by Li et al. (2019) used a deep learning-based approach to classify land cover features in remote sensing images of urban areas. The authors used a modified U-Net architecture and transfer learning to fine-tune the model on a small-scale dataset of high-resolution images. The results showed that the proposed method achieved an overall accuracy of 76.89%.

Overall, these experiments demonstrated that the U-Net model can effectively classify land cover features in urban areas, particularly when fine-tuned using transfer learning. Additionally, spatial resolution and band selection play critical roles in the performance of the model. The findings of experiment four suggest that pre-training with a high-resolution dataset and fine-tuning with lower resolution data can improve the model's performance and generalization ability. These findings have practical implications for the development of accurate and efficient land cover classification models for urban areas.

3.4.2 Potential applications on different data

Evaluation 1 focused on testing the proposed land cover classification framework on additional data from a different year and comparing its performance with two commonly used remote sensing classification methods, object-based image analysis (OBIA) and maximum likelihood (ML). The results showed that the proposed framework outperformed the traditional methods in terms of overall accuracy, precision, recall, and F1 score for most land cover classes. Moreover, the framework was able to identify impervious surfaces more accurately, which is crucial for urban land management and planning. This evaluation highlights the effectiveness and efficiency of the proposed framework in dealing with complex urban landscapes, making it a promising approach for future land cover classification studies.

Evaluation 2 aimed to evaluate the framework's performance on different datasets and over time by analyzing data from two different years in the same area. The results showed an improvement in the framework's overall accuracy from 71.3% in 2013 to 85% in 2019, indicating that the proposed framework can be applied to different datasets and provide reliable land cover classification for different years in the same area. This evaluation also tested the framework's robustness by applying it to a different city, Las Vegas, and achieved moderate overall accuracy. This evaluation highlights the potential of the proposed framework in dealing with different datasets and landscapes and its adaptability to different cities and regions.

Overall, the two evaluations highlight the feasibility and potential of the proposed land cover classification framework in future applications. The framework demonstrated its effectiveness in dealing with complex urban landscapes, outperforming traditional remote sensing classification methods and achieving high accuracy, precision, recall, and F1 score for most land cover classes. Moreover, the framework showed its adaptability to different datasets, landscapes, and cities, making it a promising approach for future land cover classification studies. The proposed framework can be useful for urban land management and planning, environmental monitoring, and disaster management, among other applications, providing a reliable and efficient method for land cover classification.

3.4.3 Limitations

Although the proposed classification framework achieved high accuracy in this study, there are several limitations that need to be addressed in future research. Firstly, the limited sample size of the dataset used for training and testing may limit the generalizability of the results. Therefore, using more diverse datasets from different regions can increase the robustness and

generalizability of the classification framework. Additionally, manual labeling was required to create the ground truth datasets for the evaluation, which can be time-consuming and costly. Future studies could explore the use of semi-supervised or unsupervised learning methods to reduce the reliance on manually labeled data. Moreover, the study focused on four land cover classes commonly used in urban land cover classification. However, future studies could explore more complex land cover classes, such as different types of buildings, roads, and water bodies, to provide more detailed information for urban planning and management. Furthermore, the integration of other data sources, such as socioeconomic and demographic data, could provide a more comprehensive understanding of the urban environment. Finally, although the proposed classification framework achieved high accuracy, there is still room for improvement. Future studies could explore the use of more advanced deep learning architectures, such as attention-based models or graph convolutional networks, to further improve the performance of the classification framework. Addressing these limitations and exploring these future directions can enhance the feasibility and applicability of the proposed framework in urban land cover classification and management.

Overall, the proposed classification framework has the potential to be applied in various urban land cover classification tasks, including urban planning, environmental management, and disaster response. By addressing the limitations and exploring the future directions, the classification framework can become more robust and effective in these applications.

3.5 Conclusions

In conclusion, this study proposed a deep learning-based framework for land cover classification in urban areas using transfer learning and multi-scale segmentation. The results demonstrated that the proposed framework achieved high accuracy in classifying impervious

surface, building, low vegetation, and tree classes, with an overall accuracy of 82.45%. Evaluation of the framework's effectiveness on different datasets and over time showed that it can provide reliable and accurate land cover classification for different years in the same area. Comparison with other widely used classification methods, such as object-based image analysis (OBIA) and maximum likelihood (ML), showed that the proposed framework outperforms these methods in terms of overall accuracy, precision, recall, and F1 score.

However, there are still some limitations to this study. The limited sample size and manual labeling required for evaluation may limit the generalizability and increase the cost of future applications. Future studies could explore the use of more diverse datasets from different regions and the integration of other data sources, such as socioeconomic and demographic data, to provide a more comprehensive understanding of the urban environment. Additionally, future studies could also explore more complex land cover classes and advanced deep learning architectures to further improve the performance of the classification framework.

Overall, the proposed framework provides an effective and efficient approach for land cover classification in urban areas using remote sensing data. The findings of this study have important implications for urban planning and management, providing valuable information for decision-making processes. The proposed framework has the potential to be applied in other urban areas and can contribute to the development of more accurate and comprehensive urban land cover maps, which can inform policy-making and planning for sustainable urban development.

Chapter 4 Socio-economic determinants of urban forest structure in a desert city: Spatial modeling using multiscale geographically weighted regression (MGWR)

Abstract

Urban forests provide crucial benefits to people and urban landscapes. This research investigates the influence of stem density and socio-economic indicators on urban forest structure in Phoenix, Arizona, using advanced statistical models. The study finds that minority and low-income populations have limited access to natural amenities, indicating a disproportionate distribution of trees in low-income neighborhoods or those with people of color. However, communities with more trees and shrubs show better educational performance, emphasizing urban forests' positive impacts on public health, well-being, and education. Considering both urban tree canopy and stem density, the research offers a comprehensive understanding of factors affecting urban forest distribution and highlights the need to address socio-economic disparities. Targeted and localized approaches to urban forest management are essential for promoting equitable access to green spaces and enhancing public health and well-being in diverse metropolitan areas like Phoenix. This chapter employs remote sensing and advanced spatial statistical modeling, revealing the importance of understanding interactions between urban forest characteristics and socio-economic factors at different spatial scales. The research findings can inform urban planners, policymakers, and community leaders in developing strategies to promote equitable access to urban forests and green spaces, ultimately improving overall quality of life and well-being in metropolitan areas.

Keywords: Urban forest; Socio-economic factors; Multiscale Geographically Weighted Regression (MGWR)

4.1 Introduction

Urban forests, including urban parks, gardens, street trees, and greenways (*Urban Forests | US Forest Service*, n.d.), constitute an important part of the urban landscape. Urban forests provide vital ecosystem services that bring a wide range of benefits to human and the urban environment. They help improve air quality, conserve energy, reduce carbon emissions, and mitigate the urban heat island effect (Baró et al., 2014; Clark et al., 1997; Escobedo et al., 2010; McDonald, 2009; McPherson, 1994). Furthermore, urban forests strengthen social cohesion and add economic values to urban communities (McDonald 2009; Song et al. 2018).

The spatial distribution of urban forests is found to be associated with the socio-economic status of urban residents. Socio-economic factors like income, education, and population density are key determinants of urban forest distribution. They are also important factors to consider in urban forest maintenance and management. (Wilkerson et al., 2018).

Existing literature have highlighted key linkages between urban forest and the socio-economic status of residents. Socio-economic indicators can affect the quality and quantity of nature in urban environment (Shanahan et al., 2015). For example, residents with higher socio-economic status are found to live in neighborhoods with higher plant diversity. The same group of residents tends to live in areas with more local parks with natural remnant ecosystems (Shanahan et al., 2015). Peterson et al. (2008) pointed out that residents who live in neighborhoods with more green space were older, better educated, and more environmentally conscious than those who live in less vegetated areas (Peterson et al., 2008).

A number of studies has identified a positive relationship between urban tree canopy cover and income (Fan et al., 2019a; Iverson & Cook, 2000; Lowry et al., 2012). Residents' income and awareness of green space's benefits may affect the amount of green space they maintain within their yards (Wilkerson et al., 2018). Further, ethnicity was found to be associated with the pattern of urban forest distribution. Martin et al. (2004) noted that the percentage of Hispanic/Latino residents was positively related to the abundance of Oleander – an evergreen plant in Phoenix, AZ (Martin et al., 2004). Schwarz et al. (2015) identified low tree canopy cover in communities with a high concentration of racial and ethnic minorities in Sacramento and Los Angeles, California (Schwarz et al., 2015). Troy et al. (2007) discovered a positive correlation between the percentage of African-American households and percentage of tree canopy cover in Baltimore, MD (Troy et al., 2007).

Tree canopy cover is widely investigated among urban forest studies (Danford et al. 2014; Locke et al. 2013; Schwarz et al. 2015; Z. Wang, Fan, and Xian 2021). Tree canopy cover is defined as the ground area covered by the layer of tree leaves, branches, and stems (Grove et al., 2006). Stem density is another important attribute of forest structure. Defined as the number of trees in a unit area (Johnson et al., 2019), stem density is frequently used to assess spatiotemporal patterns of tree mortality and forest succession dynamics (Fatehi et al., 2017). A combined use of tree canopy cover and stem density can provide informative guidance for forest management and decision-making (Crowther et al., 2015).

A number of tools has been recently developed to quantify urban forest characteristics. The i-Tree, for instance, is a collection of tools for urban forest analysis and benefits assessment (*Tools | I-Tree*, n.d.). Remote sensing techniques allow the estimation of tree cover and density at varying spatial and temporal scales. The high-resolution commercial satellites,

such as the QuickBird and IKONOS, provide capability to capture vegetation information at a spatial resolution of 2 meters to 10 meters (C. Li et al., 2010). The National Agriculture Imagery Program (NAIP) provides freely accessible aerial photos across the US with a one-meter spatial resolution, which can be effectively used to produce fine-scale vegetation products (X. Li et al., 2014; Loughner et al., 2012; Moskal et al., 2011). Furthermore, active sensors such as synthetic aperture radar (SAR) and light detection and ranging (LiDAR) are widely used to estimate tree height and delineate boundaries of individual tree crowns (Alonzo et al., 2016; Lim et al., 2003; MacFaden et al., 2012).

Previous studies focused on using correlation analysis and global regression models (such as the ordinary least squares (OLS)) to understand the relationships between urban forest characteristics and socio-economic factors (Chan & Hopkins, 2017; Nesbitt et al., 2017; Schwarz et al., 2015; Sorrensen et al., 2015; H.-F. Wang et al., 2016). Grove et al. (2014) found a positive correlation between the urban tree canopy and housing household density in New York City (Grove, Locke, and O'Neil-Dunne 2014). However, Fan et al. (2019) and Iverson and Cook (2000) both reported a negative association in Chicago (Fan et al., 2019a; Iverson & Cook, 2000). Moreover, Sorrensen et al. indicated that the home age was positively associated with urban tree cover in Lubbock, Texas (Sorrensen et al., 2015). But Conway and Hackworth found an opposite association in Toronto, Canada (Conway & Hackworth, 2007). The physical benefits of urban forests are more regional (Nelson et al., 2021). It is recognized that the relationship between urban forest attributes and socio-economic determinants is not constant and tends to vary widely over space.

In order to better understand and analyze these spatially varying relationships, researchers have developed specialized techniques. The geographically weighted regression (GWR) was

developed to allow the relationship to vary spatially. For each location, the GWR determines an optimal spatial scale at which the relationship operates (Fotheringham et al., 2017). The spatial scale is fixed for all processes in a classical GWR model. This assumption was further relaxed in a later variant of the GWR, named the multi-scale geographically weighted regression (MGWR). In the MGWR, the spatial relationship between the response and any covariate can vary locally, regionally, or not vary at all (Oshan et al., 2019). Through eliminating the constraint that all relationships operate at the same spatial scale, the MGWR reduces the potential issues of over-fitting, bias in parameter estimations, and concurvity (Oshan et al., 2019).

The goal of this paper is to understand the spatial pattern of urban forest in a desert city and explore how the pattern relates to socio-economic factors at varying spatial scales. Coupling remote sensing and advanced spatial statistical modeling, this study aims to: (1) explore urban forest patterns and socio-economic inequity in a desert city, Phoenix, Arizona; (2) understand the relationships between urban forest characteristics and socio-economic drivers at varying spatial scales. (3) compare the impacts of different urban forest characteristics on socio-economic status.

4.2 Materials and Methods

4.2.1 Study Area

The study area is the city of Phoenix, located in Maricopa County of Arizona (Figure 4.1). The city has an area of ~372 square miles encompassing 911 block groups. Phoenix has a typical desert climate with extremely long and hot summers and short, mild winters. As of 2019, Phoenix is the 5th most populous city in the United States, with an estimated population around 1.63 million (“Phoenix, Arizona,” 2021). The city is located within one of the

sunniest regions in the world, with more than 300 days of sunshine annually. Being one of the fastest growing regions in the US, Phoenix is known for its urban heat island (UHI) effect, the size and intensity of which have mounted rapidly in the recent years.

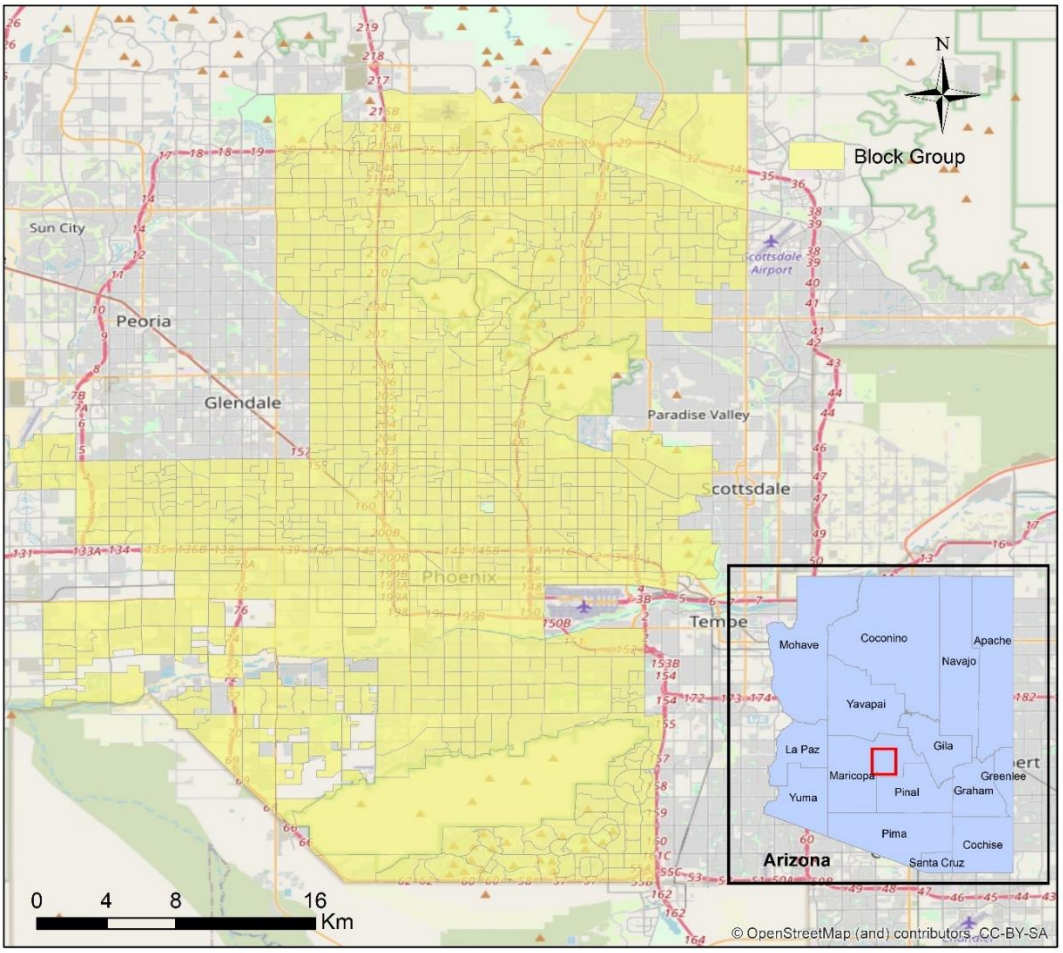


Figure 4.1 Study area located in Phoenix, Arizona.

Urban tree studies in Phoenix, Arizona, have primarily focused on understanding the distribution and benefits of urban trees in the context of the city's hot and arid climate. Research has highlighted the importance of trees in mitigating the urban heat island effect by providing shade and reducing surface temperatures. Studies have also revealed disparities in tree canopy distribution, with wealthier neighborhoods having more extensive tree canopies

than lower-income areas, emphasizing the need for targeted urban forestry programs to address environmental justice concerns. The ecosystem services provided by urban trees, such as improving air quality, reducing stormwater runoff, and sequestering carbon, are also crucial for the city's overall health and livability. Furthermore, research on water usage and drought tolerance has informed the selection of appropriate tree species and irrigation methods to maintain a sustainable urban forest in Phoenix's challenging climate.

4.2.2 Data Processing

Tree canopy cover was extracted from the 2015 land use and land cover (LULC) map created by the Central Arizona–Phoenix Long-Term Ecological Research (CAPLTER) (Y. Zhang & Billie, 2020). The LULC map was originally generated from the 2015 NAIP images with a spatial resolution of 1 meter (*NAIP Imagery*, n.d.). The percent of tree canopy cover (PTCC) was calculated at the block group level. Tree location data were provided by the Arizona State University Map and Spatial hub (Accessing Phoenix LiDAR Data | ASU Library n.d.). The tree location data were generated from the LiDAR dataset with a spatial resolution of 0.5 meter. The LiDAR data were sourced from the United States Geological Survey (USGS) three-dimensional elevation program (3DEP) (*3D Elevation Program - Data & Tools*, n.d.). Stem density (SD) was calculated as the number of trees in a block group over the area of the block group. Table 4.1 shows the summary statistics of the PTCC and SD at the block group level. Tree canopy cover ranges from 0 to 36% across 911 block groups in Phoenix. The mean percent tree canopy cover was only 8.2%. The stem density ranges from 0 to 23.33/acre with an average of 6.95 stems /acre.

Table 4.1 Descriptive statistics of the urban forest characteristics

	Minimum	Maximum	Mean	Std. Deviation
Percent of Tree canopy cover (%)	0	36	8.2	0.05
Stem density (/acre)	0	23.33	6.95	2.71

4.2.3 Socio-economic Indicators

The socio-economic data at the block group level were sourced from the 2010-2014 American Community Survey (ACS) 5-year estimate (Bureau, n.d.). Table 4.2 shows the descriptive statistics of all socio-economic variables collected. We included factors reflecting wealth (median household income, percent of poverty), demographics (population density, median age, percent associate degree or above), minority information (percent of residents that are Black, percent of residents that are Hispanic or Latino), ethnic information (percent of residents that are White) and building characteristics (density of housing units, median building age, median housing value). Density of housing units was calculated as the total number of housing units in a block group over the area of the block group. Median building age in 2014 was obtained by subtracting the median year the house was built from 2014.

Table 4.2 Descriptive statistics of socio-economic variables

	Minimum	Maximum	Mean	Std. Deviation
Population Density (/acre)	0.02	61.41	10.26	6.85
Percent of residents that are Black (%)	0	93	7	9
Percent of residents that are Hispanic or Latino (%)	0	100	39	29
Percent of residents that are White (%)	5	100	75	17
Median Household Income (k)	0	191.5	50.38	29.09
Median Age	16	82	35.42	9.33
Percent of Poverty (%)	0	96	24	20
Density of Housing Units (UNIT)	0	36.09	4.45	3.71
Median Building Age (year)	0	82	42.27	15.39
Median Housing Value (k)	0	1270	154.23	137.14
Percent of Residents with Associate Degree or Above (%)	0	72	18	15

4.2.4 Statistical Analysis

The statistical analysis proceeded in three phases. Firstly, we used the Pearson's correlation to quantify the bivariate relationships between the two urban forest attributes and the socio-economic indicators, respectively. We then conducted the ordinary least squares (OLS) regression to evaluate the relative contribution of each socio-economic variable to the urban forest attributes.

OLS is a method used to estimate the parameters of a linear regression model. A linear regression model is a statistical model that is used to predict a continuous dependent variable (y) based on one or more independent variables (x). The relationship between the variables is assumed to be linear, meaning that a change in x is associated with a constant change in y.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon \tag{4.1}$$

where β_0 is the intercept, β_1, β_2, ..., β_n are the coefficients for each independent variable and ε is the error term. The goal of OLS is to find the values of β_0, β_1, β_2, ..., β_n that minimize the sum of the squared differences between the observed y values and the predicted y values.

The following formula is applied to find the optimal values for the coefficients.

$$\beta = (X^T X)^{-1} X^T y \tag{4.2}$$

where X is the matrix of independent variables, y is the vector of dependent variables, and b is the vector of coefficients. The superscript T indicates the transpose of a matrix. The asterisk (*) symbol denotes matrix multiplication.

GWR stands for "geographically weighted regression." It is a type of regression analysis that allows the strength and form of the relationships between the dependent and independent variables to vary across different spatial locations.

GWR is used to model relationships that are not constant over space, but rather vary depending on the location. It is particularly useful when working with spatial data, as it allows you to model spatial heterogeneity in the relationships between variables.

The basic idea behind GWR is to fit a separate regression model for each location (or "zone") in the study area. The coefficients of the model are estimated using only the data points

within that zone. This allows the model to capture local patterns and trends that may not be evident in the overall dataset.

The mathematical representation of the GWR is (Fotheringham et al., 2017):

$$y_i = \sum_{j=0}^{n} \beta_{ij}\,(u_i, v_i) + \varepsilon_i \tag{4.3}$$

where y_i is the dependent variable at location i. $\beta_{ij}(u_i, v_i)$ is the coefficient associated with the j^{th} variable at location i. ε_i is the error term.

The GWR permits the linkages between the response and predictor variables to vary throughout space rather than giving an "average" global estimate (Fotheringham et al., 2017). However, it assumes that each relationship operates at the same spatial scale. The MGWR provides a more flexible framework to examine multiscale processes. MGWR allows not only for the coefficients to vary in space, but also for the scale to vary across different explanatory variables, incorporating various bandwidths across the study area surface(Fotheringham et al., 2017). he MGWR is given by

$$y_i = \sum_{j=0}^{n} \beta_{bwj}\,(u_i, v_i) + \varepsilon_i \tag{4.4}$$

where $\beta_{bwj}\,(u_i, v_i)$ is …… bwj is the bandwidth used for calibration for the j^{th} conditional relationship (Fotheringham et al., 2017). The remaining parameters are defined in the same way as in Equation (1). Different bandwidths reflect different spatial scales, thus MGWR can capture spatial heterogeneity more precisely by quantifying the effect of scale on spatial processes (Fotheringham et al., 2017).

4.3 Results

4.3.1 Urban Forest Structure

Most urban forest research has used canopy cover to represent forest structure. The term "Urban Tree Canopy" (UTC) refers to the layer of tree leaves, branches, and stems that provides tree covering of the ground (*Urban Forests | US Forest Service*, n.d.). The PTCC was calculated as the proportion of land surface that is covered by treetops within each block group area. Stem density is another key characteristic of forest structure and was calculated as the total number of stems in a stand in each block group area.

Figure 2 shows the spatial distribution of percent tree canopy cover and stem density (tree/acre) by block group. The block groups with a higher PTCC (higher than 8.83%) were mainly located at east side of the study area. Even though the highest PTCC is 35%, nearly one third of block groups had a PTCC under 5%. In 2010, tree canopy was estimated to cover 9.0% of the land area (*City of Phoenix, Arizona*, n.d.). In 2015, based on the data we collected, the tree canopy cover was 6.12%. There was a 2.88% decrease in the tree canopy cover which accounts for ~ 6702 acres. The stem density ranges from 0 to 23.3 trees/acre through the study area. By visual observation, areas with a high stem density also tend to have a high PTCC. However, some of the high-PTCC areas had a very low stem density (< 10 trees/acre).

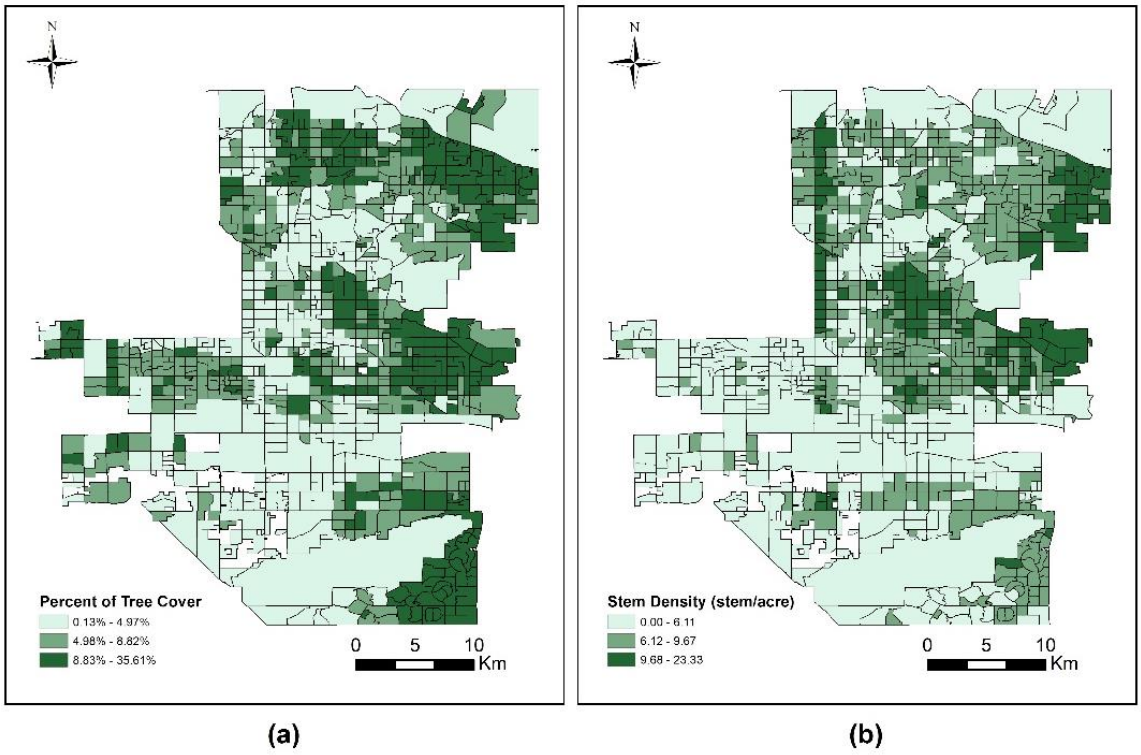


Figure 4.2 Spatial distribution of the percent of tree canopy cover (a) and stem density (b)

4.3.2 Correlation Analysis

Based on current literature, we identified a total of 11 socio-economic variables to be included in the correlation analysis. Table 4.3 shows the Pearson's correlation coefficient between each of the urban forest variable and the socio-economic variables, respectively. All variables are significantly correlated with PTCC at the 0.01 level except median building age. All variables are significantly correlated with SD except population density. Percent of Black and percent of Hispanic or Latino had a significantly negative correlation with both PTCC and SD and Percent of Hispanic or Latino shows much stronger relationships with PTCC and SD. In contrast, percent of White showed a positive association with both variables. Significant correlations were found with wealth variables including median household income, median housing value, and percent of poverty. Percent of residents with

associate degree or above, which reflects the level of education, had strong positive relationships with both urban forest indicators. Median housing value also had very strong positive association with SD. Overall, both forest variables had mostly consistent correlations with the socio-economic variables, but stem density had stronger associations with almost all variables than PTCC (except population density and density of housing units).

Table 4.3 Pearson's correlation between urban forest variables and socio-economic indicators

	Percent of tree canopy cover	Stem density (/acre)
Population Density (/acre)	**0.090****	0.027
Density of Housing Units (/acre)	**0.169****	**0.123****
Percent of Black	**-0.148****	**-0.173****
Percent of Hispanic or Latino	**-0.326****	**-0.413****
Percent of White	**0.132****	**0.243****
Median Household Income (k)	**0.262****	**0.321****
Median Housing Value (k)	**0.297****	**0.432****
Percent of Poverty	**-0.286****	**-0.310****
Median Age	**0.236****	**0.343****
Median Building Age (year)	-0.047	**0.169****
Percent of residents with associate degree or above	**0.346****	**0.426****

** Correlation is significant at the 0.01 level ($p < 0.01$).

4.3.3 Regression Results

The correlation analysis provides a rudimentary grasp of the bivariate correlations. However, it is critical to quantify each predictor's contribution to tree variables while adjusting for other predictors. We included Percent of Black, Percent of Hispanic or Latino, Median Household Income, Housing Density, and Education Level in the multiple regression analysis considering significance and collinearity. Table 4 shows the OLS regression results

for the two urban forest variables. The variance inflation factors (VIF) for both models were below 3, indicating that there was no severe multicollinearity issue. For both models, all indicators were statistically significant at the 0.05 level. Consistent with the correlation analysis, Percent of Black and Percent of Hispanic or Latino had a significant negative relationship with PTCC and SD. Percent of Black is weakly associated with PTCC. Consistent with Table 3, Percent of Hispanic or Latino is a stronger indicator in both PTCC and SD models. Median household income is again positively related to PTCC and, but it is a relatively weaker indicator in SD model. Density of housing units is strongly related to both PTCC and SD and it is the strongest indicator in PTCC model. Education level, similar to the correlation results, is a strongly positive indicator for both PTCC and SD and it is the strongest indicator in SD model. Overall, the SD model provided a better model fit with an R^2 of 0.244.

Table 4.4 OLS regression results

Percent of tree canopy					
	Unstandardized Coefficient	Std. Error	Standardized Coefficient	p-value	VIF
Percent of Black	-0.056	0.019	-0.091	0.004	1.087
Percent of Hispanic or Latino	-0.022	0.008	-0.118	0.008	2.242
Median Household Income (k)	0	0	0.125	0.007	2.371
Density of Housing Units (/acre)	0.003	0	0.24	0	1.184
Percent of residents with associate degree or above	0.064	0.019	0.178	0	2.984
R squared		0.187			
Stem density (tree/acre)					
Percent of Black	-3.187	0.945	-0.102	0	1.087
Percent of Hispanic or Latino	-1.803	0.406	-0.192	0	2.242
Median Household Income (k)	0.09	0.004	0.095	0.03	2.371
Density of Housing Units (/acre)	0.14	0.023	0.191	0	1.184
Percent of residents with associate degree or above	3.977	0.912	0.218	0	2.984
R squared		0.244			

Table 4.5 Mean significant coefficient estimates from MGWR models.

	Percent of Tree Canopy		Stem Density (/acre)	
	Mean Coefficient (significant cases)	Bandwidth	Mean Coefficient (significant cases)	Bandwidth
Percent of Black	-0.071	291	-0.039	291
Percent of Hispanic or Latino	-0.794	43	-0.294	440
Median Household Income (k)	-0.237	45	0.450	83
Density of Housing Units (/acre)	0.494	68	0.919	46
Percent of residents with associate degree or above	N/A	910	0.076	149

Table 4.5 shows the results from the MGWR models including mean coefficient estimates of statistically significant cases and bandwidth of each variable. The bandwidth varied from 43 to 910 for each variable. Mean coefficient of percent of black is negative and shows a relatively weaker correlation with both PTCC and SD. The bandwidth of percent of black keeps the same of 291 in both PTCC and SD model. Mean coefficient of percent of Hispanic or Latino again indicates that it is a strong indicator in PTCC model. The bandwidth of percent of Hispanic of Latino in PTCC model is 43 and it becomes much larger in SD model with the value of 440. There is a negative relationship for mean coefficient of median household income with PTCC which is contrary to the correlation and OLS results. Mean coefficients of density of housing units are consistent with the correlation and OLS results for both PTCC and SD models. Different from previous results, with only significant cases, high

education level is regarded as an insignificant variable in all cases in PTCC model and becomes a weak indicator in SD model. The "Bandwidth" column represents the spatial scale at which these relationships are significant. A smaller bandwidth implies that the relationship is significant at a more local scale, while a larger bandwidth suggests the relationship is significant at a broader spatial scale. The bandwidth of percent of residents with associate degree or above in PTCC model is the largest which is 910 and much smaller in SD model which is 149. This indicates that the relationship between the percentage of residents with associate degrees or higher and stem density is significant at a more local scale compared to its relationship with tree canopy coverage.

Table 4.6 shows a summary of model goodness of fit for the OLS, GWR and MGWR models. We used the R^2 and AICc as diagnostic metrics. A higher R^2 and a lower AICc indicate a better model fit. The R^2 scores were less than 0.3 for the OLS models, and above 0.7 for the MGWR models. For both PTCC and SD models, the MGWR achieved the highest R^2 scores among the three models with the highest R^2 score of 0.781 in the PTCC model. The MGWR also had the lowest AICc score. Overall, the MGWR achieved significantly greater explanatory power than the OLS and GWR with the highest R^2 and lowest AICc values.

Table 4.6 Performance comparison among OLS, GWR, and MGWR models

	PTCC			**SD**		
	OLS	GWR	MGWR	OLS	GWR	MGWR
R^2	0.187	0.681	0.781	0.244	0.659	0.708
AICC	2410.952	1820.894	1590.282	2344.777	1920.268	1786.365

4.3.4 Spatial Patterns of the MGWR estimates

Each explanatory variable uses a different bandwidth to capture varying spatial relationships in the MGWR model. The bandwidth can be explained as the number of block groups used in each local regression in the MGWR model and it varies from predictor to predictor. The bandwidth varied from 1 to 911. When it equals 911, it can be regarded as a global regression model. Therefore, based on the varied bandwidth, we divided the variables into three categories, local, regional and global variables, respectively. Global variable is easy to understand when the bandwidth is close to 911. It is hard to differentiate between a regional and a local variable. In our study, the bandwidth size between variables is relatively large. Therefore, except global variables, we defined as the variables with bandwidth smaller than 100 as local variables and all other variables as regional variables.

Figure 3 shows the spatial distribution of the significant coefficient estimates for percent of residents that are Black in the PTCC and SD models. There were more significant areas in PTCC model than SD model. All cases were significant in the PTCC model, but only partial areas represented significant relationships in SD model. As defined above, percent of Black can be regarded as a regional variable in both models. Significantly negative associations were found in all block groups in percent of tree canopy cover model which was consistent with the mean coefficient in Table 5. However, in the SD model, positive relationships were found in the northeast of the city. But the mean coefficient was negative in Table 4.5.

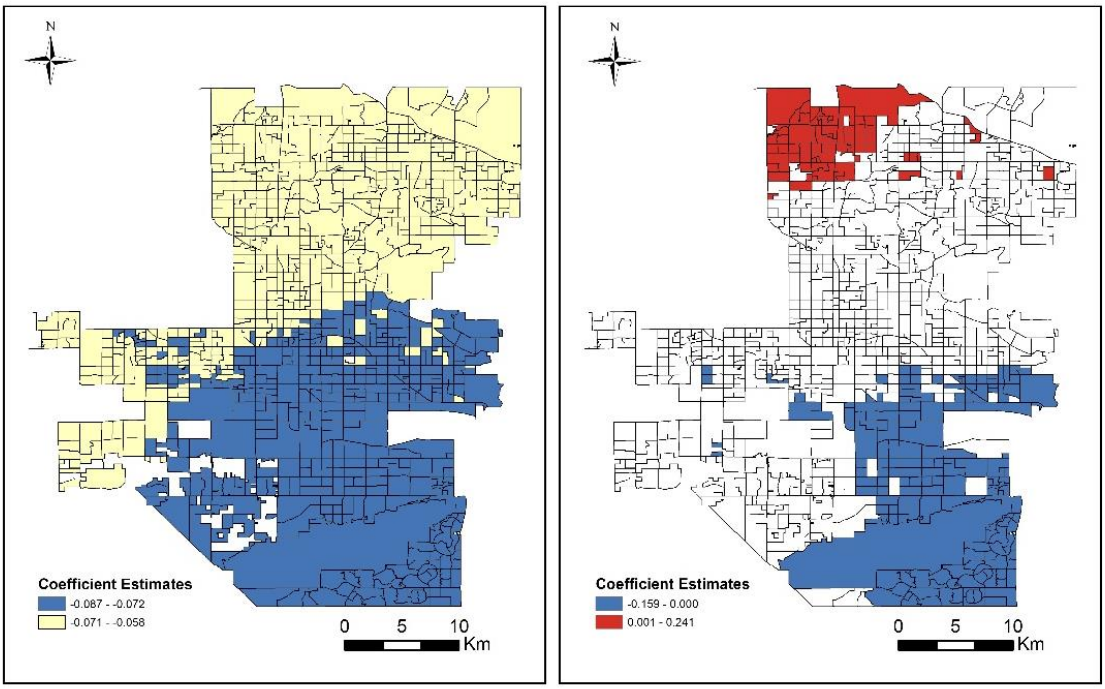


Figure 4.3 PTCC (a) and SD (b) MGWR results for percent of residents that are Black

Figure 4.4 shows the significant estimates for percent of resident that are Hispanic or Latino. All relationships in the significant cases were negative in both models which were consistent with the results in Table 3-5. The bandwidth is different in two models. There were much more significant block groups in SD model than PTCC model. Most cases were not significant in PTCC model. The bandwidth of PTCC model is 43 which percent of Hispanic or Latino can be considered as a local variable. Figure 4.4 (a) clearly shows the effect of local variable that significant cases clustered within a small number of block groups. Moreover, the bandwidth of SD model is 440 which percent of Hispanic or Latino can be considered as a regional variable. In Figure 4.4 (b), the coefficient estimates of a sizeable number of block groups are clustered with the same value range.

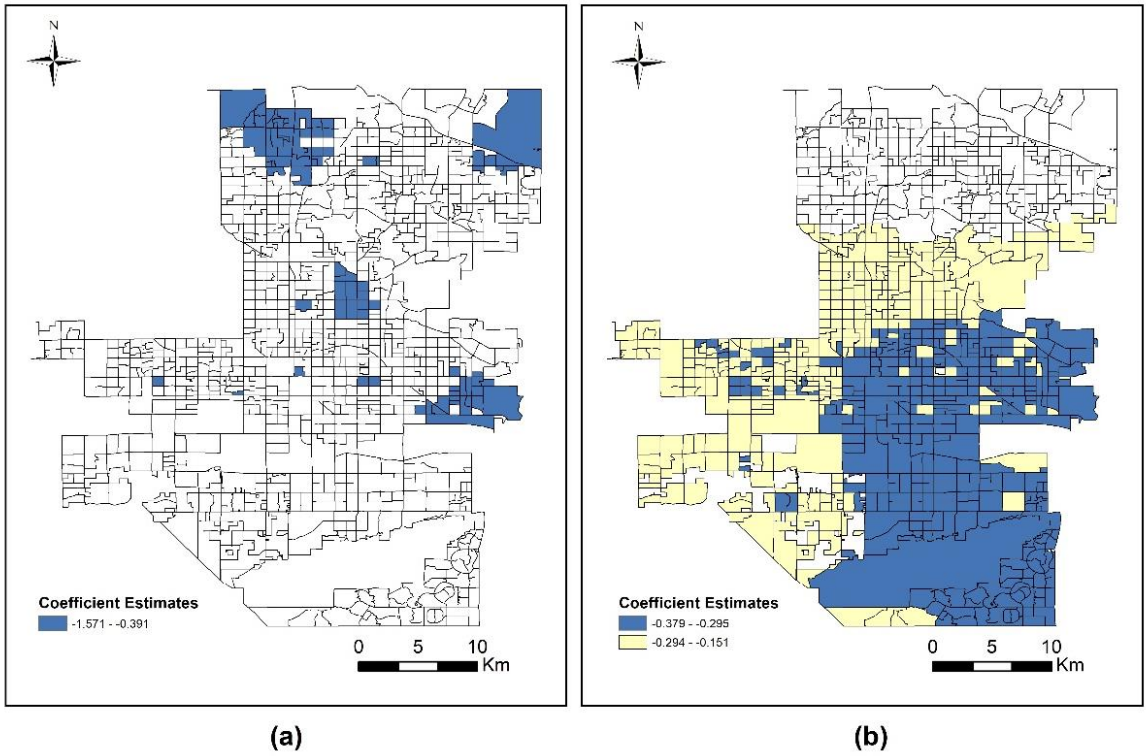


Figure 4.4 PTCC (a) and SD (b) MGWR results for percent of residents that are Hispanic or Latino

Figure 4.5 shows the significant estimates for median household income. Mean household income was considered as a local variable in two models. There were more significant cases in SD model than PTCC model. Correlation analysis and OLS results show a positive relationship between PTCC and median household income. However, in Figure 4.5 (a), only a few positive cases were found in the west of study area and mean estimate is negative since the negative effects in the northwest were stronger. In the SD model, all cases were positive which was consistent with the mean coefficient estimate in Table 4.5.

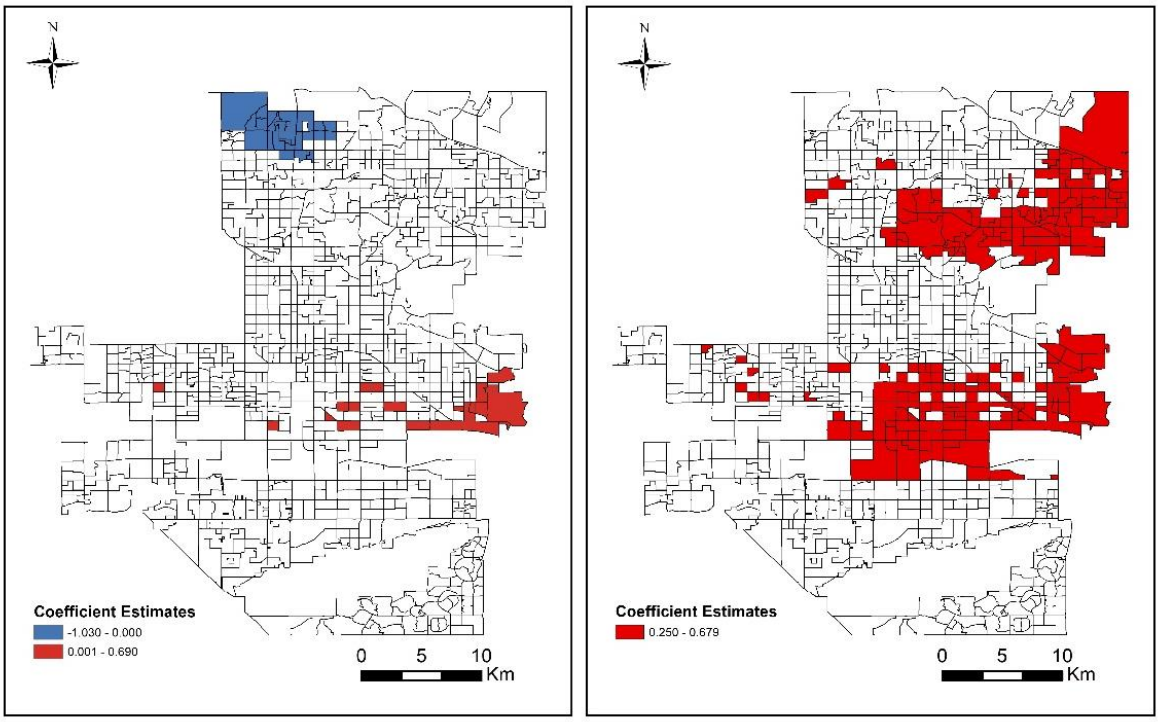


Figure 4.5 PTCC (a) and SD (b) MGWR results for median household income.

Figure 4.6 shows the spatial patterns for significant coefficient estimate for density of housing units. Housing density was considered as a local variable in two models. All significant cases reveal positive relationships and were distributed on the north and south sides of the city. In both models, there was no significant cases in center of the study area.

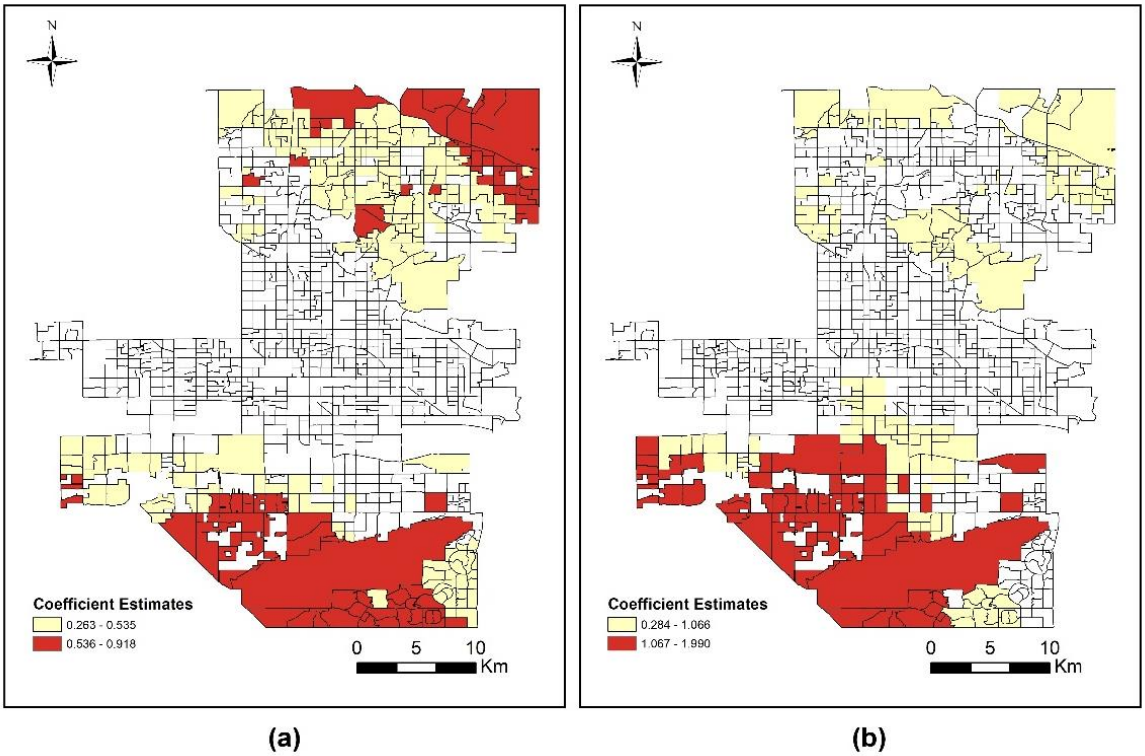


Figure 4.6 PTCC (a) and SD (b) MGWR results for density of housing units.

Figure 4.7 shows the significant estimates for education level (percent of residents with associate degree or above). Education level is identified as a global variable based on its bandwidth in tree canopy percent model. It is interesting that the education level was globally insignificant in PTCC model. It turns into a regional variable in SD model but only a few significant cases are observed. Several block groups in the northwest reveal negative relationships, leaving positive relationships in the southern part of the study area which was consistent with the results in Table 4.3-4.5.

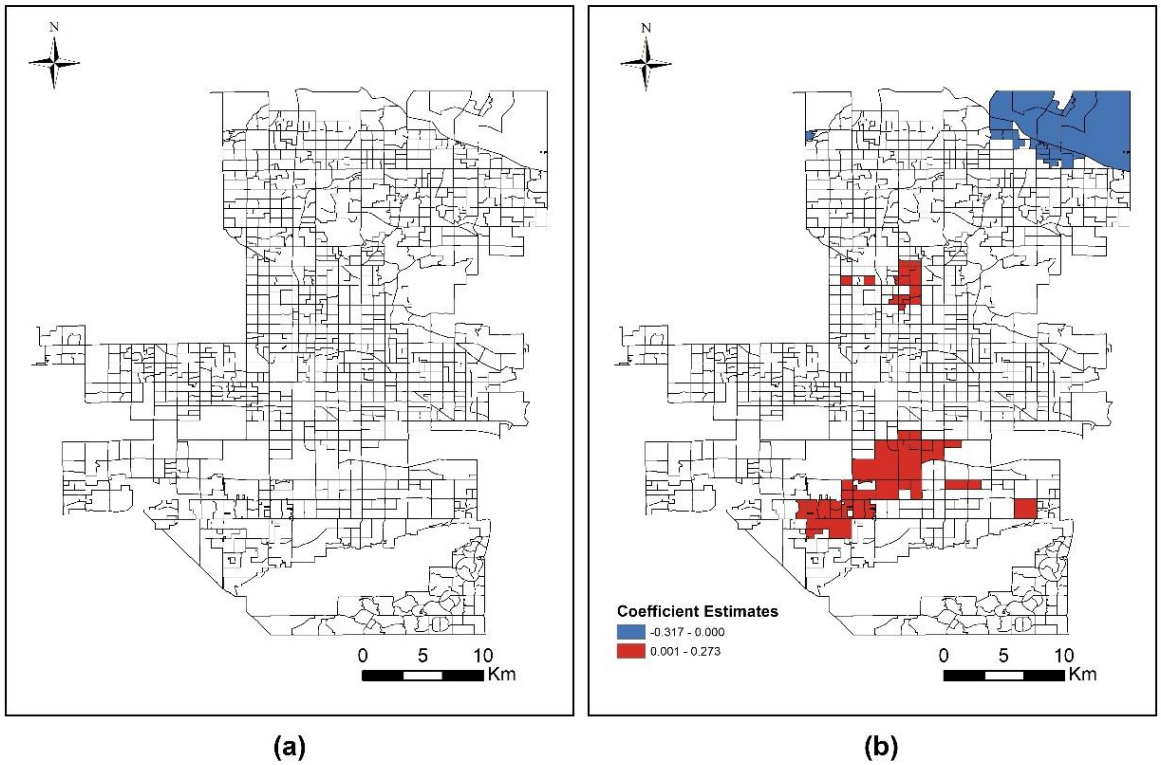


Figure 4.7 PTCC (a) and SD (b) MGWR results for percent of residents with associate degree or above.

4.4 Discussion

4.4.1 The Urban Forest in a Desert City

Urban forests have become an essential feature of modern cities, as they have been shown to provide a range of benefits, such as improved air quality, enhanced public safety, and mitigation of the urban heat island effect. A well-maintained urban forest can contribute to reducing the impact of rising temperatures by providing shade and increasing evapotranspiration from soil and leaves. While cities in forested regions such as the Northeastern United States have been studied extensively for their tree-planting efforts, urban forestry in the arid desert presents different challenges. For instance, choosing the most effective species for an arid environment is critical.

In the city of Phoenix, Arizona, a master plan was developed by the City of Phoenix Parks and Recreation Department in 2010 to achieve a tree canopy cover of 25% by 2030, as the percentage of urban forest coverage was low compared to regional standards. However, the percentage of tree canopy cover was estimated to be only 13% in 2005, with a decrease to 6.12% in 2015. The current state of the urban forest in Phoenix is, therefore, a cause for concern. Monsoon storms and accidents cause the destruction of approximately 1,000 trees annually, and the city's funding for tree replacement was drastically reduced during the recession. Furthermore, a high percentage of block groups have a tree cover percentage of less than 5%, with many areas having only 0-6 trees per acre. This demonstrates the urgent need for more trees and shade in Phoenix.

In addition, understanding the structure of urban forests in a desert city is vital to effective urban forest management. Previous studies that employed tree canopy cover as the sole indicator of urban forest structure have limitations, as stem density and tree density also provide valuable information. In Cook County, Illinois, Chao et al. studied the land use and socio-economic determinants of urban forest structure using three indicators: stem density, urban tree cover, and species diversity. The findings showed a similarity between the results of tree canopy cover and stem density. However, in Phoenix, the structure of the urban forest differs, as the number of trees does not necessarily correlate with a larger canopy area. Therefore, a more targeted greening plan is necessary for desert cities.

In conclusion, the benefits of urban forests are evident, and their role in mitigating the impact of rising temperatures is critical. However, the challenges posed by arid desert environments must be considered when designing urban forest plans. The City of Phoenix has recognized this need and has implemented a master plan to increase tree canopy cover, which requires

significant investment and commitment. Nevertheless, understanding the structure of the urban forest in Phoenix and other desert cities is essential to successful urban forest management.

4.4.2 Importance of the multiscale analysis

The concept of scale is essential in geographic analyses, and it has significant implications for understanding complex processes operating at different levels. While there is a considerable body of literature on the importance of scale in various geographical contexts, measuring the geographic scale over which different processes operate is challenging. Previous studies suggest that the geographically weighted regression (GWR) approach is critical for accurately interpreting local relationships. However, GWR is limited by its assumption that each variable operates at a single scale, which may not hold in many real-world scenarios. In contrast, the multiscale geographically weighted regression (MGWR) approach is more flexible and allows each variable to be processed at varying scales, making it an ideal tool for policy suggestions.

In this study, we utilized the MGWR approach to examine the effects of socio-economic indicators on urban forest distribution at different spatial scales. Our findings demonstrated that the MGWR approach is capable of providing a nuanced understanding of how these indicators influence urban forest distribution at different scales. Specifically, the results of our study showed that the percentage of black population acts as a regional variable, while income and housing density act as local variables in both PTCC and SD models. The percentage of Hispanic or Latino is a local variable in the PTCC model but a regional variable in the SD model. Furthermore, the education level showed global insignificance in the PTCC model and regional significance in the SD model. These results highlight the

importance of accounting for the different scales at which variables operate in landscape scale analyses.

While OLS and GWR provide unique results, they do not consider the varying scales at which variables operate. In contrast, the MGWR approach can provide a more comprehensive and nuanced understanding of the spatial heterogeneity in the statistical relationships between landscape factors and public perceptions. The MGWR approach also allows for the identification of the scales at which different processes operate, providing valuable information for policy and planning. Our study demonstrates the importance of multiscale analysis in landscape scale analyses and highlights the value of the MGWR approach in effectively supporting policy and planning. Future studies could explore the application of MGWR in other geographic contexts and further develop the method to better account for spatial heterogeneity in statistical relationships between landscape factors and public perceptions.

4.4.3 Urban Tree Distribution Inequality in a Desert City

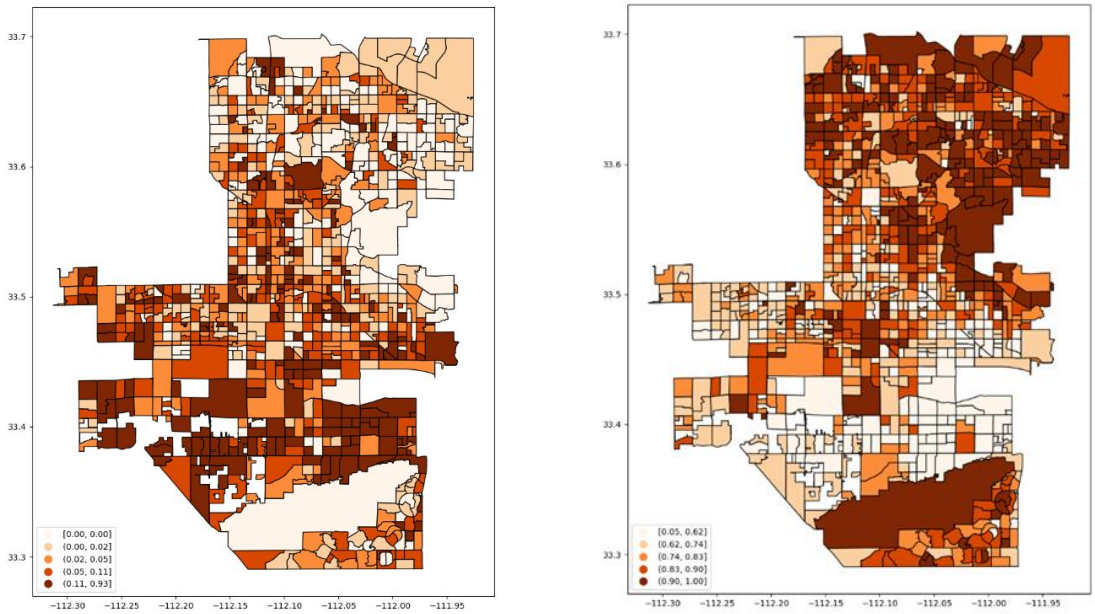
Chapter 4 investigates the complex relationships between urban forest characteristics and socio-economic indicators in the diverse metropolitan area of Phoenix, Arizona, using correlation analysis, OLS, and MGWR models. While the correlation and OLS results offer a broad overview of the associations between PTCC and SD and the socio-economic variables in the study area, the MGWR analysis provides a more nuanced and spatially differentiated perspective of the coefficient estimates and their variability at various scales. The study finds that minority and low-income populations have limited access to natural amenities, which is consistent with earlier research demonstrating that trees are disproportionately distributed in neighborhoods with low-income or people of color. However, MGWR results indicate a

positive association between the percentage of Black and SD, but a negative association with PTCC in the northwest corner of Phoenix, where there is a higher percentage of White residents and a lower percentage of Black residents. These results suggest that the Black population in those areas have more access to green spaces, similar to the White population. Interestingly, the variation of the magnitude of PTCC doesn't align with SD in that specific region, although the shift of the magnitude of PTCC and SD are mostly similar in other areas. In several block groups, the relationships between the percent of Hispanic or Latino and PTCC are only significant, whereas the SD model yields more significant relationships. This inconsistency may be attributed to the unique structure of the urban forest in Phoenix, which is different from other cities due to the extremely hot and dry climate of the desert city.

Moreover, this study finds that communities with more trees and shrubs are associated with higher educational performance. These results provide evidence that green spaces and urban forests can have positive effects on public health and well-being, including educational outcomes, and promote social equity by ensuring equal access to natural amenities for all communities. However, the MGWR results also indicate that not every variable is a significant predictor of PTCC and SD within each block group, highlighting the need for more targeted and localized approaches to urban forest management in diverse metropolitan areas like Phoenix.

In conclusion, this study offers important insights into the complex relationships between urban forest characteristics, socio-economic indicators, and demographic variables in Phoenix, Arizona, using advanced statistical models. The findings have significant implications for urban forest management and planning in desert cities and can inform

policies and programs aimed at promoting equitable access to green spaces and improving public health and well-being.

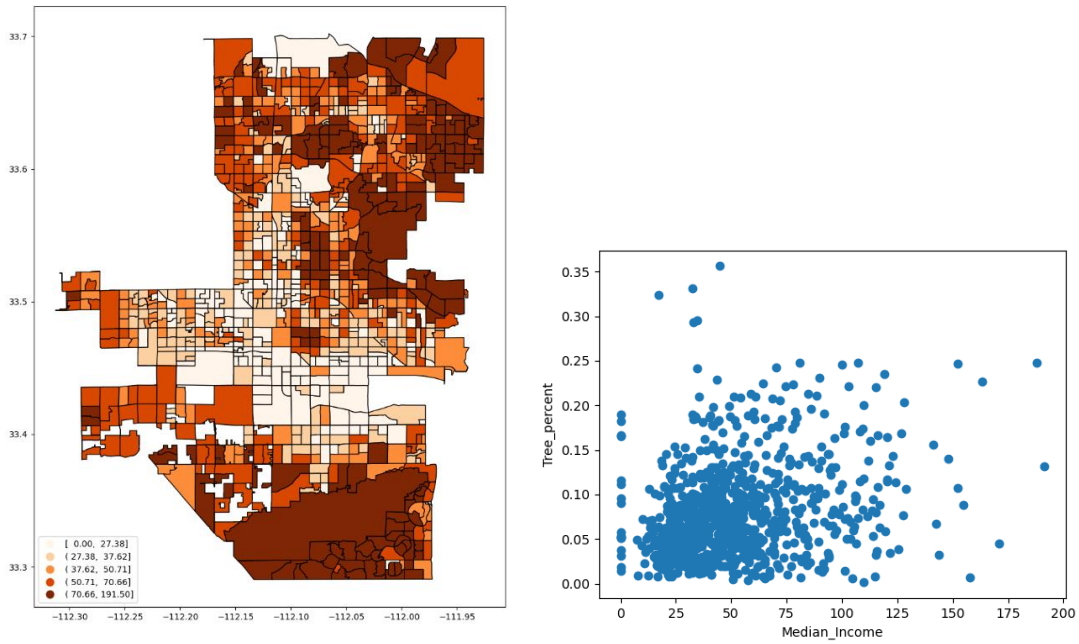


(a) (b)

Figure 4.8 Spatial distribution of the percent of residents that are Black (a) and White (b)

Income is another important indicator discussed a lot in previous studies. Correlation analysis and OLS results in this study support evidence from previous findings that low-income households have less access to green infrastructure (citation). But in MGWR results, contrary to global relationships, a small number of significant cases were observed in the PTCC model result, and some negative associations were found in the northwest corner (Figure 5 (a)). According to Figure 2 (a) and Figure 9 (a). It is reasonable that this area had a relatively higher median income value and a lower PTCC. Correlations analysis and OLS indicate a positive association between income and PTCC which can be observed from Figure 9 (b) as well. However, according to the scatter plot, there are several cases with low income but high tree percent or high income but low tree percent clustered at the origin. This explained why most cases were insignificant in PTCC results. More significant cases are observed in SD model and the values of the coefficient varied slightly because, in these areas, the positive relationships were

consistent that the stem density increased when the median household income was higher in the block group (Figure 5(b) and Figure 9 (a)).



(a) (b)

Figure 4.9 The spatial distribution of median household income (a) and the scatter plot between percent tree canopy cover and median household income (b)

As our results suggest, housing density is positively correlated with PTCC and SD. It is reasonable that trees can increase both the aesthetic values and economic values of the properties. Therefore, most residents and real estate developers would like to plant more trees surrounding their houses. According to Figure 10, the housing density is low south and north of the city. Significant cases were also observed in those areas in both models. In other areas, the spatial distribution of housing density was more irregular, but we can see a regular distribution of PTCC and SD according to Figure 2. This explains why there were no significant cases in these areas.

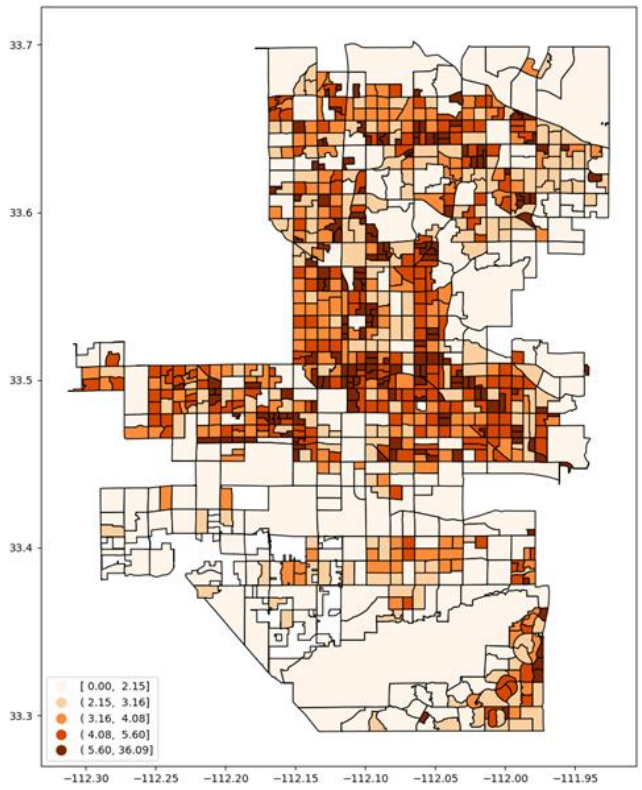


Figure 4.10 Spatial distribution of household density

The high-education level was positively correlated with PCTT in previous studies. Our correlation analysis and OLS results suggest that education level is weakly related to two forest variables. Comparatively, MGWR results are not significantly associated with PCTT and significant correlations with SD can be observed in several block groups Figure 4.11. Figure 4.11 shows the spatial pattern of the residents with a higher education level. Compared to Figure 4.2 with Figure 4.11, the shift trends of the patterns are not consistent. In the study area, most block groups had a percentage of higher education level less than 20 %. However, the PTCC and SD values varied significantly. This indicated that even though the global results reveal a significantly positive relationship, the MGWR results are more reliable without any significant relationships observed locally.

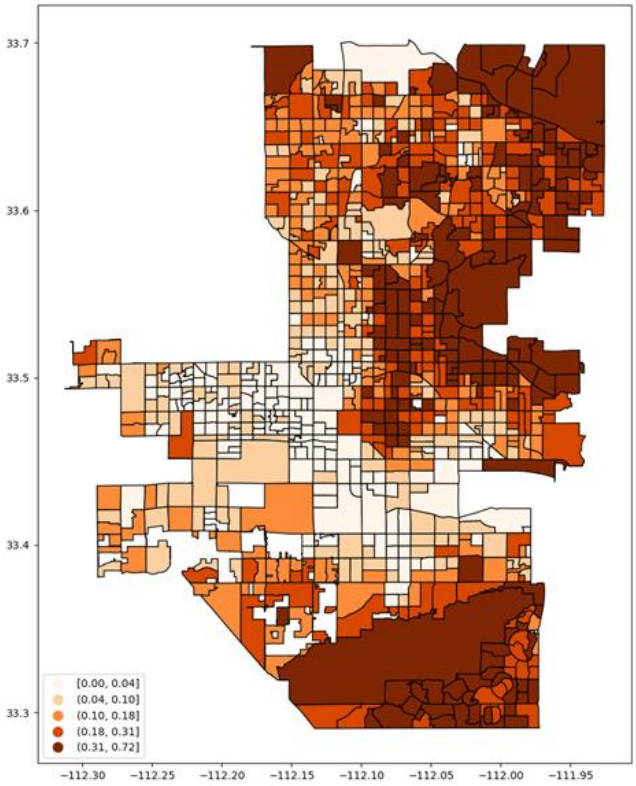


Figure 4.11 Spatial distribution of percentage of associate's or above degree holders

Urban tree inequality is essentially everywhere in the United States. Low-income communities have less tree cover than high-income neighborhoods in 92% of US cities. Wealthy communities have 15% more tree cover and live in districts that are 1.50 degrees Celsius cooler than the poor ones (McDonald et al., 2021). Growing trees in urban areas across the United States is becoming more popular to decrease tree inequality. Government and nonprofit organizations in Boston, Detroit, Phoenix, and other cities are organizing tree-planting events and developing urban forestry plans ("For Tree Equity and Climate Change, How Many Urban Trees Do We Need?," n.d.). But, in order to progress Tree Equity, how many trees should be planted each year? Phoenix is different from the cities like Boston or Chicago. Based on the findings in this study, the tree cover is much lower, and the tree

species is unique due to the desert climate. In April 2021, the City of Phoenix and American Forests signed a Memorandum of Understanding (MOU) to collaborate together to increase tree equality in Phoenix communities. The government intends to increase the canopy of trees in Arizona's low-income, low-canopy communities (*Official Website of the City of Phoenix, Arizona*, n.d.). The results in this study help propose a more targeted plan on planting trees in the low-income communities. Moreover, findings indicate that the tree cover decreased from 2010 to 2015. One important reason might be people would like to remove the trees due to the annual cost of maintaining the trees is high (Carmichael & McDonough, 2018). This aligns to the results of correlation analysis OLS regression that the residents with high-education level have more access to the trees. The Phoenix government said that they intended to educate the public on the benefits of trees, as there is limited understanding of the importance of the urban forest ("Official Website of the City of Phoenix, Arizona" n.d.). Therefore, simply increasing the tree cover in low-cover communities is not the most efficient way to decease the tree inequality. There should be a difference in priority and intensity. Combining our results with policymakers' considerations for local finances and policies can lead to more economical solutions.

4.5 Conclusions

This Chapter investigates the complex relationships between urban forest characteristics and socio-economic indicators in Phoenix, Arizona, using advanced statistical models. The study finds that minority and low-income populations have limited access to natural amenities, which is consistent with earlier research demonstrating that trees are disproportionately distributed in neighborhoods with low-income or people of color. However, the study also finds that communities with more trees and shrubs are associated with higher educational

performance, highlighting the potential positive effects of urban forests on public health and well-being, including educational outcomes. The results of the study suggest that targeted and localized approaches to urban forest management are necessary to promote equitable access to green spaces and improve public health and well-being in diverse metropolitan areas like Phoenix.

Chapter 5 Conclusion

Deep learning has proven to be a potent tool for urban classification due to its capacity to learn and extract intricate features from urban datasets. This dissertation initially implemented the U-Net model for urban tree canopy extraction, demonstrating its cutting-edge performance in urban classification tasks. Concurrently, intriguing insights emerged at various spatial scales. However, a limitation persists in the scarcity of publicly accessible training datasets for urban classification. Consequently, I introduced the concept of transfer learning as a subsequent step. Transfer learning can enhance urban classification by utilizing pre-trained deep learning models on extensive datasets, thereby improving performance on smaller urban datasets with a limited number of labeled samples. The primary contribution of this dissertation is the proposal of an innovative cross-scale framework that amalgamates transfer learning and deep learning for urban land cover mapping. This framework has been proven effective and feasible through evaluations conducted in different cities and years. It offers a promising approach to surmount the challenges posed by the availability of high-resolution imagery and the deficiency of publicly accessible training datasets. The proposed framework has great potential for land cover mapping and can be applied at different spatial and temporal scales.

To extend the application of the urban tree canopy and land cover maps. The dissertation finally has investigated the complex relationships between urban forest characteristics, socio-economic indicators, and their impact on public health and well-being in Phoenix, Arizona. Through a comprehensive literature review and the application of advanced statistical models, this study has highlighted the importance of accurate and timely information about the distribution and structure of urban tree canopy. The findings demonstrate that minority

and low-income populations have limited access to natural amenities, which is consistent with earlier research demonstrating that trees are disproportionately distributed in neighborhoods with low-income or people of color. However, the study also finds that communities with more trees and shrubs are associated with higher educational performance, highlighting the potential positive effects of urban forests on public health and well-being, including educational outcomes. The findings have significant implications for urban forest management and planning, as well as for green infrastructure planning and sustainable and equitable urban development. The study emphasizes the importance of urban forests in promoting public health, well-being, and social equity in urban areas. Moreover, it also suggests that targeted and localized approaches to urban forest management are necessary to promote equitable access to green spaces and improve public health and well-being in diverse metropolitan areas like Phoenix.

In conclusion, this dissertation has contributed to the advancement of deep learning methods for urban forest studies and has provided important insights into the role of urban forests in promoting public health and well-being in diverse communities. The findings of this study provide a scientific basis for green infrastructure planning and sustainable and equitable urban development. The study highlights the need for further research on the socio-economic determinants of urban forest structure in other metropolitan areas, as well as the need for more targeted and localized approaches to urban forest management in diverse communities. Overall, this dissertation emphasizes the importance of urban forests in promoting public health, well-being, and social equity in urban areas, and provides a promising framework for sustainable and equitable urban development.

References

2D Semantic Labeling—Potsdam. (n.d.). Retrieved January 18, 2021, from https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-potsdam/

3D Elevation Program—Data & Tools. (n.d.). Retrieved May 18, 2021, from https://www.usgs.gov/core-science-systems/ngp/3dep/data-tools

14.2. Fine-Tuning—Dive into Deep Learning 1.0.0-alpha1.post0 documentation. (n.d.). Retrieved December 10, 2022, from http://d2l.ai/chapter_computer-vision/fine-tuning.html

Ahmed, K. R., & Akter, S. (2017). Analysis of landcover change in southwest Bengal delta due to floods by NDVI, NDWI and K-means cluster with landsat multi-spectral surface reflectance satellite data. *Remote Sensing Applications: Society and Environment*, *8*, 168–181. https://doi.org/10.1016/j.rsase.2017.08.010

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, *148*, 70–83.

Alonzo, M., McFadden, J. P., Nowak, D. J., & Roberts, D. A. (2016). Mapping urban forest structure and function using hyperspectral imagery and lidar data. *Urban Forestry & Urban Greening*, *17*, 135–147. https://doi.org/10.1016/j.ufug.2016.04.003

Araujo, A., Norris, W., & Sim, J. (2019). Computing receptive fields of convolutional neural networks. *Distill*, *4*(11), e21.

Audebert, N., Le Saux, B., & Lefèvre, S. (2016). Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. *Asian Conference on Computer Vision*, 180–196.

Ba, J., & Caruana, R. (2014). Do deep nets really need to be deep? *Advances in Neural Information Processing Systems*, *27*, 2654–2662.

Baeza, S., & Paruelo, J. M. (2020). Land Use/Land Cover Change (2000–2014) in the Rio de la Plata Grasslands: An Analysis Based on MODIS NDVI Time Series. *Remote Sensing*, *12*(3), 381.

Baró, F., Chaparro, L., Gómez-Baggethun, E., Langemeyer, J., Nowak, D. J., & Terradas, J. (2014). Contribution of Ecosystem Services to Air Quality and Climate Change Mitigation Policies: The Case of Urban Forests in Barcelona, Spain. *AMBIO*, *43*(4), 466–479. https://doi.org/10.1007/s13280-014-0507-x

Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., & Blaschko, M. B. (2019). Optimizing the Dice score and Jaccard index for medical image segmentation: Theory and practice. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 92–100.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16.

Bureau, U. C. (n.d.). *Census.gov*. Census.Gov. Retrieved April 9, 2020, from https://www.census.gov/en.html

Buyantuyev, A., & Wu, J. (2010). Urban heat islands and landscape heterogeneity: Linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology*, *25*(1), Article 1. https://doi.org/10.1007/s10980-009-9402-4

Carmichael, C. E., & McDonough, M. H. (2018). The trouble with trees? Social and political dynamics of street tree-planting efforts in Detroit, Michigan, USA. *Urban Forestry & Urban Greening*, *31*, 221–229. https://doi.org/10.1016/j.ufug.2018.03.009

Chan, A. Y., & Hopkins, K. G. (2017). Associations between sociodemographics and green infrastructure placement in Portland, Oregon. *Journal of Sustainable Water in the Built Environment*, *3*(3), 05017002.

Chang, N.-B., Hossain, U., Valencia, A., Qiu, J., & Kapucu, N. (2020). The role of food-energy-water nexus analyses in urban growth models for urban sustainability: A review of synergistic framework. *Sustainable Cities and Society*, *63*, 102486. https://doi.org/10.1016/j.scs.2020.102486

Cheng, G., Xie, X., Han, J., Guo, L., & Xia, G.-S. (2020). Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3735–3756.

Cheng, K.-S., Su, Y.-F., Kuo, F.-T., Hung, W.-C., & Chiang, J.-L. (2008). Assessing the effect of landcover changes on air temperatu$\times$ re using remote sensing images—A pilot study in northern Taiwan. *Landscape and Urban Planning*, *85*(2), 85–96.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 424–432.

Cities and communities in the US losing 36 million trees a year. (n.d.). ScienceDaily. Retrieved October 19, 2020, from https://www.sciencedaily.com/releases/2018/04/180418141323.htm

Clark, J. R., Matheny, N. P., Cross, G., & Wake, V. (1997). A model of urban forest sustainability. *Journal of Arboriculture*, *23*, 17–30.

Cochocki, A., & Unbehauen, R. (1993). *Neural networks for optimization and signal processing*. John Wiley & Sons, Inc.

Congalton, R. G. (1991). Remote sensing and geographic information system data integration: Error sources and. *Photogrammetric Engineering & Remote Sensing*, *57*(6), 677–687.

Conway, T., & Hackworth, J. (2007). Urban pattern and land cover variation in the greater Toronto area. *The Canadian Geographer/Le Géographe Canadien*, *51*(1), 43–57.

Coutts, C., & Hahn, M. (2015). Green infrastructure, ecosystem services, and human health. *International Journal of Environmental Research and Public Health*, *12*(8), 9768–9798.

Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., Smith, J. R., Hintler, G., Duguid, M. C., Amatulli, G., Tuanmu, M.-N., Jetz, W., Salas, C., Stam, C., Piotto, D., Tavani, R., Green, S., Bruce, G., Williams, S. J., … Bradford, M. A. (2015). Mapping tree density at a global scale. *Nature*, *525*(7568), Article 7568. https://doi.org/10.1038/nature14967

Danford, R. S., Cheng, C., Strohbach, M. W., Ryan, R., Nicolson, C., & Warren, P. S. (2014). What Does It Take to Achieve Equitable Urban Tree Canopy Distribution? A Boston Case Study. *Cities and the Environment (CATE)*, *7*(1), 2.

De Luca, G., N. Silva, J. M., Cerasoli, S., Araújo, J., Campos, J., Di Fazio, S., & Modica, G. (2019). Object-Based Land Cover Classification of Cork Oak Woodlands using UAV Imagery and Orfeo ToolBox. *Remote Sensing*, *11*(10), Article 10. https://doi.org/10.3390/rs11101238

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 94–114.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*(3), 297–302.

Dobbs, C., Kendal, D., & Nitschke, C. R. (2014). Multiple ecosystem services and disservices of the urban forest establishing their connections with landscape structure and sociodemographics. *Ecological Indicators*, *43*, 44–55. https://doi.org/10.1016/j.ecolind.2014.02.007

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 196–212.

Dwyer, M. C., & Miller, R. W. (1999). Using GIS to assess urban tree canopy benefits and surrounding greenspace distributions. *Journal of Arboriculture*, *25*, 102–107.

ECognition | Trimble Geospatial. (n.d.). Retrieved September 10, 2020, from https://geospatial.trimble.com/products-and-solutions/ecognition

Epa, U. (2016). Using trees and vegetation to reduce heat islands. *Heat Islands*.

Escobedo, F., Varela, S., Zhao, M., Wagner, J. E., & Zipperer, W. (2010). Analyzing the efficacy of subtropical urban forests in offsetting carbon emissions from cities. *Environmental Science & Policy*, *13*(5), 362–372.

Esser, P., Sutter, E., & Ommer, B. (2018). A variational u-net for conditional appearance and shape generation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8857–8866.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., & Seiwald, K. (2019). U-Net: Deep learning for cell counting, detection, and morphometry. *Nature Methods*, *16*(1), 67–70.

Fan, C., Johnston, M., Darling, L., Scott, L., & Liao, F. H. (2019a). Land use and socio-economic determinants of urban forest structure and diversity. *Landscape and Urban Planning*, *181*, 10–21.

Fan, C., Johnston, M., Darling, L., Scott, L., & Liao, F. H. (2019b). Land use and socio-economic determinants of urban forest structure and diversity. *Landscape and Urban Planning*, *181*, 10–21. https://doi.org/10.1016/j.landurbplan.2018.09.012

Fan, C., & Wang, Z. (2020). Spatiotemporal characterization of land cover impacts on urban warming: A spatial autocorrelation approach. *Remote Sensing*, *12*(10), 1631.

Fatehi, P., Damm, A., Leiterer, R., Pir Bavaghar, M., Schaepman, M. E., & Kneubühler, M. (2017). Tree density and forest productivity in a heterogeneous alpine environment: Insights from airborne laser scanning and imaging spectroscopy. *Forests*, *8*(6), 212.

Feng, W., Sui, H., Huang, W., Xu, C., & An, K. (2018). Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based

conditional random field model. *IEEE Geoscience and Remote Sensing Letters*, *16*(4), 618–622.

Ferri, S., Syrris, V., Florczyk, A., Scavazzon, M., Halkia, M., & Pesaresi, M. (2014). A new map of the European settlements by automatic classification of 2.5 m resolution SPOT data. *2014 IEEE Geoscience and Remote Sensing Symposium*, 1160–1163.

For Tree Equity and Climate Change, How Many Urban Trees Do We Need? (n.d.). *American Forests*. Retrieved November 10, 2022, from https://www.americanforests.org/tools-research-reports-and-guides/research-reports/climate-change-urban-forests/

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265.

Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., & Kleinn, C. (2019). Large Scale Palm Tree Detection In High Resolution Satellite Images Using U-Net. *Remote Sensing*, *11*(3), 312.

Frey, N. (2017). Equity in the distribution of urban environmental amenities: The case of Washington, DC. *Urban Geography*, *38*(10), 1534–1549.

Fugate, D., Tarnavsky, E., & Stow, D. (2010). A survey of the evolution of remote sensing imaging systems and urban remote sensing applications. *Remote Sensing of Urban and Suburban Areas*, 119–139.

Fuller, R. A., & Gaston, K. J. (2009). The scaling of green space coverage in European cities. *Biology Letters*, *5*(3), 352–355.

Gardiner, D. (n.d.). *Phoenix's urban forest is shrinking—And residents say it's time to change that*. The Arizona Republic. Retrieved July 6, 2021, from https://www.azcentral.com/story/news/local/phoenix/2017/05/27/phoenix-urban-forest-shrinking-residents-say-its-time-change/331751001/

Grove, J. M., Locke, D. H., & O'Neil-Dunne, J. P. (2014). An ecology of prestige in New York City: Examining the relationships among population density, socio-economic status, group identity, and residential canopy cover. *Environmental Management*, *54*(3), 402–419.

Grove, J. M., O'Neil-Dunne, J., Pelletier, K., Nowak, D., & Walton, J. (2006). A report on New York City's present and possible urban tree canopy. *United States Department of Agriculture, Forest Service, Northeastern Area, South Burlington, Vermont*.

Hamers, L. (1989). Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing and Management*, *25*(3), 315–318.

Hay, G. J., & Castilla, G. (2006). Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT). *Proc. 1st Int. Conf. OBIA*, 4–5.

Hegazy, I. R., & Kaloop, M. R. (2015). Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt. *International Journal of Sustainable Built Environment*, *4*(1), 117–124.

Hillel, D., & Rosenzweig, C. (2010). *Handbook of climate change and agroecosystems: Impacts, adaptation, and mitigation* (Vol. 1). World Scientific.

Hossain, M. D., & Chen, D. (2019). Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*, 115–134.

Hu, F., Xia, G.-S., Hu, J., & Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, *7*(11), 14680–14707.

ISPRS Benchmark Test on Urban Object Detection and Reconstruction—ISPRS. (n.d.). Retrieved April 30, 2020, from http://www2.isprs.org/commissions/comm3/wg4/tests.html

Iverson, L. R., & Cook, E. A. (2000). Urban forest cover of the Chicago region and its relation to household density and income. *Urban Ecosystems*, *4*(2), 105–124.

Johnson, P. S., Shifley, S. R., Rogers, R., Dey, D. C., & Kabrick, J. M. (2019). *The ecology and silviculture of oaks*. Cabi.

Joyce, K. E., Belliss, S. E., Samsonov, S. V., McNeill, S. J., & Glassey, P. J. (2009). A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Progress in Physical Geography*, *33*(2), 183–207.

King, K. L., & Locke, D. H. (2013). *A comparison of three methods for measuring local urban tree canopy cover*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012a). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012b). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Kuras, A., Brell, M., Rizzi, J., & Burud, I. (2021). Hyperspectral and lidar data applied to the urban land cover machine learning and neural-network-based classification: A review. *Remote Sensing*, *13*(17), 3393.

Landsat Satellite Missions. (n.d.). Retrieved October 17, 2019, from https://www.usgs.gov/land-resources/nli/landsat/landsat-satellite-missions?qt-science_support_page_related_con=2#qt-science_support_page_related_con

Li, C., Yin, J., & Zhao, J. (2010). Extraction of urban vegetation from high resolution remote sensing image. *2010 International Conference On Computer Design and Applications*, *4*, V4-403.

Li, M., Zang, S., Zhang, B., Li, S., & Wu, C. (2014). A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, *47*(1), 389–411.

Li, X., Myint, S. W., Zhang, Y., Galletti, C., Zhang, X., & Turner II, B. L. (2014). Object-based land-cover classification for metropolitan Phoenix, Arizona, using aerial photography. *International Journal of Applied Earth Observation and Geoinformation*, *33*, 321–330.

Li, Z., Wang, Y., Zhang, N., Zhang, Y., Zhao, Z., Xu, D., Ben, G., & Gao, Y. (2022). Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey. *Remote Sensing*, *14*(10), 2385.

Liang, H., & Li, Q. (2016). Hyperspectral Imagery Classification Using Sparse Representations of Convolutional Neural Network Features. *Remote Sensing*, *8*(2), Article 2. https://doi.org/10.3390/rs8020099

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography: Earth and Environment*, *27*(1), 88–106. https://doi.org/10.1191/0309133303pp360ra

Locke, D. H., Grove, J. M., Galvin, M., O'Neil-Dunne, J. P., & Murphy, C. (2013). Applications of urban tree canopy assessment and prioritization tools: Supporting collaborative decision making to achieve urban sustainability goals. *Cities and the Environment (CATE)*, *6*(1), 7.

Long, J., Shelhamer, E., & Darrell, T. (2015a). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Long, J., Shelhamer, E., & Darrell, T. (2015b). *Fully Convolutional Networks for Semantic Segmentation*. 3431–3440. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html

Loughner, C. P., Allen, D. J., Zhang, D.-L., Pickering, K. E., Dickerson, R. R., & Landry, L. (2012). Roles of urban tree canopy and buildings in urban heat island effects: Parameterization and preliminary results. *Journal of Applied Meteorology and Climatology*, *51*(10), 1775–1793.

Lowry, J. H., Baker, M. E., & Ramsey, R. D. (2012). Determinants of urban tree canopy in residential neighborhoods: Household characteristics, urban form, and the geophysical landscape. *Urban Ecosystems*, *15*(1), 247–266.

LP DAAC - MODIS Overview. (n.d.). Retrieved December 14, 2020, from https://lpdaac.usgs.gov/data/get-started-data/collection-overview/missions/modis-overview/#modis-metadata

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *152*, 166–177.

Macartney, C., & Weyde, T. (2018). Improved speech enhancement with the wave-u-net. *ArXiv Preprint ArXiv:1811.11307*.

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.

Manakos, I., Schneider, T., & Ammer, U. (2000). A comparison between the ISODATA and the eCognition classification methods on basis of field data. *IAPRS*, *33*(Suppl. B7), 133–139.

Martin, C. A., Warren, P. S., & Kinzig, A. P. (2004). Neighborhood socioeconomic status is a useful predictor of perennial landscape vegetation in residential neighborhoods and embedded small parks of Phoenix, AZ. *Landscape and Urban Planning*, *69*(4), 355–368.

Mathieu, R., Freeman, C., & Aryal, J. (2007). Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery. *Landscape and Urban Planning*, *81*(3), 179–192.

McDonald, R. I. (2009). Ecosystem service demand and supply along the urban-to-rural gradient. *Journal of Conservation Planning*, *5*(5), 1–14.

McDonald, R. I., Biswas, T., Sachar, C., Housman, I., Boucher, T. M., Balk, D., Nowak, D., Spotswood, E., Stanley, C. K., & Leyk, S. (2021). The tree cover and temperature disparity in US urbanized areas: Quantifying the association with income across 5,723 communities. *PLOS ONE*, *16*(4), e0249715. https://doi.org/10.1371/journal.pone.0249715

McPherson, E. G. (1994). Using urban forests for energy efficiency and carbon storage. *Journal of Forestry;(United States)*, *92*(10).

Mennis, J. (2006). Socioeconomic-vegetation relationships in urban, residential land. *Photogrammetric Engineering & Remote Sensing*, *72*(8), 911–921.

Moskal, L. M., Styers, D. M., & Halabisky, M. (2011). Monitoring Urban Tree Cover Using Object-Based Image Analysis and Public Domain Remotely Sensed Data. *Remote Sensing*, *3*(10), Article 10. https://doi.org/10.3390/rs3102243

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. Object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, *115*(5), 1145–1161. https://doi.org/10.1016/j.rse.2010.12.017

NAIP Imagery. (n.d.). [Page]. National-Content. Retrieved May 18, 2021, from https://fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index

Navalgund, R. R., Jayaraman, V., & Roy, P. S. (2007). Remote sensing applications: An overview. *Current Science*, 1747–1766.

Nelson, J. R., Grubesic, T. H., Miller, J. A., & Chamberlain, A. W. (2021). The equity of tree distribution in the most ruthlessly hot city in the United States: Phoenix, Arizona. *Urban Forestry & Urban Greening*, *59*, 127016.

Nesbitt, L., Hotte, N., Barron, S., Cowan, J., & Sheppard, S. R. J. (2017). The social and economic value of cultural ecosystem services provided by urban forests in North America: A review and suggestions for future research. *Urban Forestry & Urban Greening*, *25*, 103–111. https://doi.org/10.1016/j.ufug.2017.05.005

New Web Map for Accessing Phoenix LiDAR Data | ASU Library. (n.d.). Retrieved May 18, 2021, from https://lib.asu.edu/geo/news/New-Web-Map-Accessing-Phoenix-LiDAR-Data

Nikolaou, G., Neocleous, D., Christou, A., Kitta, E., & Katsoulas, N. (2020). Implementing sustainable irrigation in water-scarce regions under the impact of climate change. *Agronomy*, *10*(8), 1120.

Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution*, *116*(3), 381–389.

Nowak, D. J., Crane, D. E., Stevens, J. C., Hoehn, R. E., Walton, J. T., & Bond, J. (2008). A ground-based method of assessing urban forest structure and ecosystem services. *Aboriculture & Urban Forestry. 34 (6): 347-358.*, *34*(6).

Official Website of the City of Phoenix, Arizona. (n.d.). Retrieved July 12, 2021, from https://www.phoenix.gov/

Ogneva-Himmelberger, Y., Pearsall, H., & Rakshit, R. (2009). Concrete evidence & geographically weighted regression: A regional analysis of wealth and the land cover in Massachusetts. *Applied Geography*, *29*(4), 478–487.

Ondruska, P., Dequaire, J., Wang, D. Z., & Posner, I. (2016). End-to-end tracking and semantic segmentation using recurrent neural networks. *ArXiv Preprint ArXiv:1604.05091*.

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *ISPRS International Journal of Geo-Information*, *8*(6), Article 6. https://doi.org/10.3390/ijgi8060269

Paisitkriangkrai, S., Sherrah, J., Janney, P., & Van-Den Hengel, A. (2015). Effective semantic pixel labelling with convolutional networks and Conditional Random Fields. *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 36–43. https://doi.org/10.1109/CVPRW.2015.7301381

Pan, Z., Xu, J., Guo, Y., Hu, Y., & Wang, G. (2020a). Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. *Remote Sensing*, *12*(10), 1574.

Pan, Z., Xu, J., Guo, Y., Hu, Y., & Wang, G. (2020b). Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. *Remote Sensing*, *12*(10), Article 10. https://doi.org/10.3390/rs12101574

Pandit, R., Polyakov, M., & Sadler, R. (2014). Valuing public and private urban tree canopy cover. *Australian Journal of Agricultural and Resource Economics*, *58*(3), 453–470.

Pashaei, M., Kamangir, H., Starek, M. J., & Tissot, P. (2020). Review and Evaluation of Deep Learning Architectures for Efficient Land Cover Mapping with UAS Hyper-Spatial Imagery: A Case Study Over a Wetland. *Remote Sensing*, *12*(6), Article 6. https://doi.org/10.3390/rs12060959

Pauleit, S., Hansen, R., Rall, E. L., Zölch, T., Andersson, E., Luz, A. C., Szaraz, L., Tosics, I., & Vierikko, K. (2017). Urban landscapes and green infrastructure. In *Oxford research encyclopedia of environmental science*.

Payton, S., Lindsey, G., Wilson, J., Ottensmann, J. R., & Man, J. (2008). Valuing the benefits of the urban forest: A spatial hedonic approach. *Journal of Environmental Planning and Management*, *51*(6), 717–736.

Peterson, M. N., Chen, X., & Liu, J. (2008). Household location choices: Implications for biodiversity conservation. *Conservation Biology*, *22*(4), 912–921.

Phoenix, Arizona. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Phoenix,_Arizona&oldid=1002989126

Qian, R., Zhang, B., Yue, Y., Wang, Z., & Coenen, F. (2015). Robust chinese traffic sign detection and recognition with deep convolutional neural network. *2015 11th International Conference on Natural Computation (ICNC)*, 791–796.

Qin, X., Wu, C., Chang, H., Lu, H., & Zhang, X. (2020). Match Feature U-Net: Dynamic Receptive Field Networks for Biomedical Image Segmentation. *Symmetry*, *12*(8), Article 8. https://doi.org/10.3390/sym12081230

QuickBird Satellite Sensor | Satellite Imaging Corp. (n.d.). Retrieved July 5, 2020, from https://www.satimagingcorp.com/satellite-sensors/quickbird/

Rasti, B., Chang, Y., Dalsasso, E., Denis, L., & Ghamisi, P. (2021). Image restoration for remote sensing: Overview and toolbox. *IEEE Geoscience and Remote Sensing Magazine*, *10*(2), 201–230.

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.

Riley, C. B., & Gardiner, M. M. (2020). Examining the distributional equity of urban tree canopy cover and ecosystem services across United States cities. *PLOS ONE*, *15*(2), e0228499. https://doi.org/10.1371/journal.pone.0228499

Ronda, R. J., Steeneveld, G. J., Heusinkveld, B. G., Attema, J. J., & Holtslag, A. A. M. (2017). Urban finescale forecasting reveals weather conditions with unprecedented detail. *Bulletin of the American Meteorological Society*, *98*(12), 2675–2688.

Rongqiang Qian, Bailing Zhang, Yong Yue, Zhao Wang, & Coenen, F. (2015). Robust chinese traffic sign detection and recognition with deep convolutional neural network. *2015 11th International Conference on Natural Computation (ICNC)*, 791–796. https://doi.org/10.1109/ICNC.2015.7378092

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rostami, M., Kolouri, S., Eaton, E., & Kim, K. (2019). Deep Transfer Learning for Few-Shot SAR Image Classification. *Remote Sensing*, *11*(11), Article 11. https://doi.org/10.3390/rs11111374

Sang, D. V., & Minh, N. D. (2018). Fully Residual Convolutional Neural Networks for Aerial Image Segmentation. *Proceedings of the Ninth International Symposium on Information and Communication Technology*, 289–296.

Schmeiser, B. (1982). Batch size effects in the analysis of simulation output. *Operations Research*, *30*(3), 556–568.

Schwarz, K., Fragkias, M., Boone, C. G., Zhou, W., McHale, M., Grove, J. M., O'Neil-Dunne, J., McFadden, J. P., Buckley, G. L., & Childers, D. (2015). Trees grow on money: Urban tree canopy cover and environmental justice. *PloS One*, *10*(4).

Sedona, R., Cavallaro, G., Jitsev, J., Strube, A., Riedel, M., & Benediktsson, J. A. (2019). Remote sensing big data classification with high performance distributed deep learning. *Remote Sensing*, *11*(24), 3056.

Shanahan, D. F., Fuller, R. A., Bush, R., Lin, B. B., & Gaston, K. J. (2015). The health benefits of urban nature: How much do we need? *BioScience*, *65*(5), 476–485.

Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 806–813.

Shojanoori, R., & Shafri, H. Z. (2016). Review on the use of remote sensing for urban forest monitoring. *Arboric Urban For*, *42*(6), 400–417.

Smith, S. L., Kindermans, P.-J., Ying, C., & Le, Q. V. (2017). Don't decay the learning rate, increase the batch size. *ArXiv Preprint ArXiv:1711.00489*.

Sorrensen, C. L., Carter, P. L., & Phelps, J. (2015). Urban landscape as mirror of ethnicity: Trees of the South Plains. *Urban Geography*, *36*(7), 1042–1063.

SPOT - CNES. (n.d.). Retrieved June 14, 2020, from https://web.archive.org/web/20131006213713/http://www.cnes.fr/web/CNES-en/1415-spot.php

Stehman, S. (1996). Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing*, *62*(4), 401–407.

Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(5), e1379.

Talarchek, G. M. (1990). The urban forest of New Orleans: An exploratory analysis of relationships. *Urban Geography*, *11*(1), 65–86.

Thapa, R. B., & Murayama, Y. (2009). Urban mapping, accuracy, & image classification: A comparison of multiple approaches in Tsukuba City, Japan. *Applied Geography*, *29*(1), 135–144.

The Thematic Mapper « Landsat Science. (n.d.). Retrieved June 14, 2020, from https://landsat.gsfc.nasa.gov/the-thematic-mapper/

Tools | i-Tree. (n.d.). Retrieved November 2, 2021, from https://www.itreetools.org/tools

Tran, D. X., Pla, F., Latorre-Carmona, P., Myint, S. W., Caetano, M., & Kieu, H. V. (2017). Characterizing the relationship between land use land cover change and land surface temperature. *ISPRS Journal of Photogrammetry and Remote Sensing*, *124*, 119–132. https://doi.org/10.1016/j.isprsjprs.2017.01.001

Tree Cover % — How Does Your City Measure Up? | DeepRoot Blog. (n.d.). Retrieved July 5, 2020, from https://www.deeproot.com/blog/blog-entries/tree-cover-how-does-your-city-measure-up

Troy, A. R., Grove, J. M., O'Neil-Dunne, J. P., Pickett, S. T., & Cadenasso, M. L. (2007). Predicting opportunities for greening and patterns of vegetation on private urban lands. *Environmental Management*, *40*(3), 394–412.

Ulmer, J. M., Wolf, K. L., Backman, D. R., Tretheway, R. L., Blain, C. J., O'Neil-Dunne, J. P., & Frank, L. D. (2016). Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription. *Health & Place*, *42*, 54–62.

Urban Forests | US Forest Service. (n.d.). Retrieved January 26, 2021, from https://www.fs.usda.gov/managing-land/urban-forests

USGS.gov | Science for a changing world. (n.d.). Retrieved January 31, 2023, from https://www.usgs.gov/

Wagner, F. H., Dalagnol, R., Tarabalka, Y., Segantine, T. Y., Thomé, R., & Hirye, M. (2020). U-Net-Id, an Instance Segmentation Model for Building Extraction from Satellite Images—Case Study in the Joanópolis City, Brazil. *Remote Sensing*, *12*(10), 1544.

Wang, H.-F., Qureshi, S., Qureshi, B. A., Qiu, J.-X., Friedman, C. R., Breuste, J., & Wang, X.-K. (2016). A multivariate analysis integrating ecological, socioeconomic and physical characteristics to investigate urban forest cover and plant diversity in Beijing, China. *Ecological Indicators*, *60*, 921–929. https://doi.org/10.1016/j.ecolind.2015.08.015

Wang, M., Zhang, X., Niu, X., Wang, F., & Zhang, X. (2019). Scene classification of high-resolution remotely sensed image based on ResNet. *Journal of Geovisualization and Spatial Analysis*, *3*, 1–9.

Wang, P., Feng, X., Zhao, S., Xiao, P., & Xu, C. (2007). *Comparison of object-oriented with pixel-based classification techniques on urban classification using TM and IKONOS imagery* (W. Ju & S. Zhao, Eds.; p. 67522J). https://doi.org/10.1117/12.760759

Wang, R., Zhao, J., & Meitner, M. J. (2017). Urban woodland understory characteristics in relation to aesthetic and recreational preference. *Urban Forestry & Urban Greening*, *24*, 55–61.

Wang, Z., Fan, C., & Xian, M. (2021). Application and Evaluation of a Deep Learning Architecture to Urban Tree Canopy Mapping. *Remote Sensing*, *13*(9), Article 9. https://doi.org/10.3390/rs13091749

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sensing*, *11*(11), Article 11. https://doi.org/10.3390/rs11111309

Whiteside, T. G., Boggs, G. S., & Maier, S. W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation*, *13*(6), 884–893. https://doi.org/10.1016/j.jag.2011.06.008

Wilkerson, M. L., Mitchell, M. G. E., Shanahan, D., Wilson, K. A., Ives, C. D., Lovelock, C. E., & Rhodes, J. R. (2018). The role of socio-economic factors in planning and managing urban ecosystem services. *Ecosystem Services*, *31*, 102–110. https://doi.org/10.1016/j.ecoser.2018.02.017

Xu, Z., Yang, J., Peng, C., Wu, Y., Jiang, X., Li, R., Zheng, Y., Gao, Y., Liu, S., & Tian, B. (2014). Development of an UAS for post-earthquake disaster surveying and its

application in Ms7. 0 Lushan Earthquake, Sichuan, China. *Computers & Geosciences*, *68*, 22–30.

Yang, C.-C., Prasher, S. O., Enright, P., Madramootoo, C., Burgess, M., Goel, P. K., & Callum, I. (2003). Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*, *76*(3), 1101–1117. https://doi.org/10.1016/S0308-521X(02)00051-3

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *ArXiv Preprint ArXiv:1411.1792*.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., & Wang, J. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716.

Zhang, J., Du, J., Liu, H., Hou, X., Zhao, Y., & Ding, M. (2019). LU-NET: An Improved U-Net for ventricular segmentation. *IEEE Access*, *7*, 92539–92546.

Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 22–40.

Zhang, Y., & Billie, T. I. (2020). *Land-cover mapping of the central Arizona region based on 2015 National Agriculture Imagery Program (NAIP) imagery* [Data set]. Environmental Data Initiative. https://doi.org/10.6073/PASTA/E671ED549A55FDA3338B177A2AD54487

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, *15*(5), 749–753.

Zhou, W. (2013). An object-based approach for urban land cover classification: Integrating LiDAR height and intensity data. *IEEE Geoscience and Remote Sensing Letters*, *10*(4), 928–931.

Zhu, G., & Blumberg, D. G. (2002). Classification using ASTER data and SVM algorithms;: The case study of Beer Sheva, Israel. *Remote Sensing of Environment*, *80*(2), 233–240. https://doi.org/10.1016/S0034-4257(01)00305-4

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.