A WALK IN THE CITY
Using Large Data Sets to Analyze Urban Sidewalks

by

MARYAMSADAT HOSSEINI

A Dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Urban Systems

Written under the direction of

Professor Karen Franck

And approved by

_____

_____

_____

_____

Newark, New Jersey

October, 2022

ABSTRACT OF THE DISSERTATION

A Walk in the City

Using Large Data Sets to Analyze Urban Sidewalks

by MARYAM SADAT HOSSEINI

Dissertation Director:

Karen Franck

While cities worldwide are increasingly promoting streets and public spaces that prioritize pedestrians over vehicles, significant data gaps have made mapping, analysis, and assessment of pedestrian infrastructure, challenging to carry out. Even in industrialized economies, most cities still lack information about the location, connectivity, and quality of their sidewalks, making it difficult to implement research on pedestrian infrastructure and holding the technology industry back from developing accurate, location-based Apps for different users. Moreover, despite the growing attention to urban data analysis, there is a gap between the real needs of researchers and practitioners directly studying urban problems and the urban analysis tools being developed. Standing at the intersection of economics, urban planning, and computer science, my dissertation aims at addressing both issues by providing theory-rich tools for large-scale assessment of urban sidewalks at two scales: at the human scale, using street-level images by proposing CitySurfaces for classifying eight classes of surface materials, and at the city, scale using aerial imagery, by proposing Tile2Net to create pedestrian networks from aerial imagery. Both studies use computer vision techniques to design frameworks and models for analyzing pedestrian facilities.

This dissertation addresses some of the challenges of semantic segmentation models regarding the high cost of image annotation by employing different techniques, such as active learning to offer solutions tailored to the specific qualities of urban problems.

ACKNOWLEDGMENTS

Today, it has been six years and two months since the day I left my loved ones at the airport in Tehran. Six transforming years that, at some points, just felt like a never-ending journey, and I could have never made it through all the challenges it presented if it were not for the love, support, help, and patience of my family and friends.

I would like first to thank my mom and dad, who have been my solid rock, source of hope, inspiration, and encouragement to push through challenging times. My loving dad, who always wanted to cut the talking time over WhatsApp short, so I can go back to my studies, and my mom who just wanted to continue till her heart gets settled. My dad was the main motivation behind my choosing Ph.D. and his genuine interest and excitement about all PhD related things I do always brought pure joy to my heart.

I should also thank Mahdi, my best and closest friend, partner, love, and the one who has always been there for me through thick and thin. I know I could have not done it if it was not for his constant support, help, brainstorming time and time again, and patience with my crazy work schedule and deadline pressures. I am thankful to my brilliant brother, Mohammad, for the thought provoking conversation, my amazing nephew Ali and my sister in-law Fatemeh, to my sister Sedighe, who has a heart of gold, Mehdi, who is such a blessing to the whole family, Nazanin, my sweet niece, my eldest sister, the kind, energetic and high-spirited Elahe, and to Masoud, who truly helped me on many occasions, and to my lovely nieces Sajedeh and Mahdieh.

It gives me immense pleasure to thank Karen Franck, my advisor, mentor, and friend. A successful woman I have always admired and looked up to, professionally and personally. I was so lucky to have you as my advisor; learned a lot from you and your careful, thoughtful, and constructive comments on my writing, critical thinking, and approach to research. I am thankful, Karen, for all the times we spent together, all the great conversations, and all you have taught me.

I would like to express my most sincere gratitude to Claudio Silva, who has been a great mentor and advisor. He was the one who trusted me in the first place, back in 2018, when I was losing hope in ever being able to find an advisor who could understand the jump I was about to make from a social science-based program in Urban Systems to the technical world of computer science and provided me with the environment wherein I could thrive, learn, and grow. Thank you for supporting me through all that and for believing in me.

I also wish to sincerely thank Andres Sevtsuk for all his support, great guidelines, interesting discussions, and all that I learned from him. Since 2018 when I first met him at the Harvard Graduate School of Design, where we talked about sidewalk accessibility assessments, to 2020, when he accepted to be on my dissertation committee, and we started working on the Mapping the Walk paper, till now that I am looking forward to starting my postdoc with him, I have continuously learned from and enjoyed working with him. I am also very thankful to Woojin Jung and Vonu Thakuriah for accepting to be on my committee and for their thoughtful comments.

I should also thank Fabio Miranda, who was a force behind my staying at NYU. I published my first paper in the CS world with him, enjoyed his creative ideas, and learned much from our collaborations and joint projects. Big thanks to Neel Dey for his tremendous help with my research. I learned a lot from him, and I am very thankful for the time he spent introducing me to the world of semantic segmentation and beyond. Also, sincere thanks to Jon Froehlich, whom I always feel lucky to get to know (thanks to Fabio!). His encouraging words, inspiring enthusiasm for and creative take on research, and amazing collaborations we had pushed me forward through challenging times. And to all my great friends at the VIDA lab, Joao, Peter, Harish, Kunal, Ann, and Thales, as well as my col-leagues and collaborators at NYU, Rutgers, and beyond, Marcos, Nivan, Charlie, Graham, Eric, Roberto, Hamed, Alex, Dana, Mikey, Jason, Woojin, Vonue, Esthi, Gulse, M.J, and my program chairs, Sean, Mara, and Jamie. And big thanks to my amazing, supportive, loving, and incredible friends, who have been my family and much more since we arrived in the US. Thank you to Elahe, who was my angel when I was so new to NYU; Sajede, my sister, friend, and shoulder in the tough times; Reyhane, my sweet, kind, strong, and amaz-

ing friend, my talented, and always supportive brother Saeed and the kind, high-spirited, lovely, bright, and full of life Mahta, who helped me a lot with the annotation and image sorting, Ghazaleh, the most generous, big-hearted person I know, Sara, Sabihe and Ali, Fatima and Hammad, Shiva and Ali, Zahra, Ali and their little Boshra, Hanie and her mom, Tannaz, Shaya, Mina, Sama, Samaneh, Fatima, and all my friends who have been sources of warmth, energy, inspiration, and love in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Throughout the history of urbanization *change* has been the steady feature of cities. Aided by technological advancements, the physical form of cities co-evolved with social composition as cities underwent cycles of transformation. The prevalence of personal automobiles redefined the use of urban space, making remote locations accessible (Schaeffer & Sclar, 1980). Taken by the wave of suburbanization, concentrated urban areas were transformed into sprawling metropolitan regions. Large public spaces and shared streets were replaced by wide roads and multi-lane highways, leaving pedestrians with narrow sidewalks. With every array of changes, new challenges emerged, putting the health, safety, social well-being, and economic viability of urban residents at risk.

1.1 Why Sidewalks?

Sidewalks are the focal point of the human scale of the city, where the most basic and widely used form of travel, walking, takes place. As the most important pedestrian-dedicated planned public spaces, sidewalks have been shown to impact various aspects of urban life, from public health to the economy, safety from crime, and social interactions. Well-designed sidewalks can create safe, lively, inclusive, walkable, and accessible cities (Gehl, 2011, 2013; Jacobs, 1961; Speck, 2013), all of that by encouraging people to leave their cars behind and take a walk/roll in the city.

For a large population of people with mobility or vision impairment, sidewalks are the main and often the only means of accessing public spaces (Clarke et al., 2008; Deitz, 2021; Hosseini, Saugstad, et al., 2022). The design of public spaces, specifically sidewalks and crossings, can significantly impact the independence and self-reliance of such groups (Froehlich et al., 2022; Mitchell, 2006b). Environmental barriers, uneven surfaces,

missing curb ramps and crossing demarcation, and narrow sidewalks can turn a short trip to local stores into an impossible mission (Brenner & Clarke, 2019; Eisenberg et al., 2017; Harris et al., 2015).

Accessibility, safety from crime, and walkability often go hand in hand. Safe and accessible sidewalks attract more people, and more frequented sidewalks can bolster the neighborhoods' safety by increasing the opportunity for informal surveillance, as explained by Jane Jacobs (1961) *eyes on the street* theory. Also, the presence and quality of sidewalks have been found to be significant predictors of perceived safety in the pedestrian environment (Landis et al., 2001).

Aside from crime safety, sidewalks can impact two other forms of safety: safety from falling and safety from traffic. Safety from tripping or falling is an important aspect of sidewalk design considerations (Clifton et al., 2007; Emery et al., 2003). Studies show a strong correlation between pedestrian falling and injuries and sidewalk features (Chippendale, 2020; S. Lee, 2018; Twardzik et al., 2019). The existence of potholes, unleveled and heaved surfaces, overcrowded sidewalks, obstructions and obstacles in the way, bad lighting, and slippery surfaces can all lead to pedestrian falling and injuries. This is of crucial importance, specifically for more vulnerable populations such as the elderly, pregnant women, and people with disabilities (Aghaabbasi et al., 2018; Clifton et al., 2007; Haans & De Kort, 2012). Sufficiently wide sidewalks, with furnishing zones that act as buffers between pedestrians and vehicular traffic, or bike lanes, can reduce the risk of (Asadi-Shekari et al., 2013; Crews & Zavotka, 2006).

From an economic perspective, well-designed sidewalks can be one of the most impactful forms of marketing for the adjacent business (Credit, 2018). Walkable and accessible neighborhoods, with retail stores and services located within walking distance, can attract different groups of people and create strong agglomeration forces that support the local economy and boost the productivity of the neighborhood businesses (Sevtsuk, 2020).

Accounting for selection bias, sidewalks that promote walking can increase physical

activity and hence, decrease the risk of diseases associated with a sedentary lifestyle, such as cardiovascular disease (Krizek, 2003; McCormack et al., 2017; Slater et al., 2013), provided that the environmental condition such as noise and air pollution are also within the safe range. Moreover, lively and inviting sidewalks, which provide places for staying, such as benches or building stoops, can encourage social interactions and create social capital, enhancing residents' mental state of residents (Van Cauwenberg et al., 2012).

Sidewalks can also impact the environmental condition of the urban areas. As discussed in chapter 4, the choice of surface materials can directly impact the microclimates within the city, which can lead to the creation of Heat Islands, a phenomenon associated with global warming and climate change (Estoque et al., 2017; Oke, 1982). Impervious surface materials are also found to be the primary cause of Combined Sewer Overflows (CSOs), which can lead to massive pollution of natural bodies of water and street flooding (Joshi et al., 2021).

1.2 The Challenges of Quality Assessment

Assessing the built environment is a challenging task facing several municipalities. An important mobilizing factor is the Americans with Disabilities Act (ADA) which was passed in 1990 and required public agencies to identify all barriers to access in publicly-owned streets and buildings. In the Chicago region, for instance, only 22 of the region's 200 municipalities with more than 50 employees had a plan (Metropolitan Planning Council, 2021). Even with clear federal requirements, the prohibiting cost and time of assessing the built environment, specifically pedestrian infrastructure, is a major constraint. In an area as large as NYC, with thousands of city blocks, manual auditing of the sidewalks is virtually impossible. To make this problem even more complex, sidewalks, like any built environment feature, change and evolve. It is then necessary to keep data inventories up to date, as a static and outdated dataset is arguably only marginally more useful than no dataset.

In recent years, online platforms made available by city agencies where residents can

report a wide range of problems (including defected sidewalks) have received significant attention. Analyzing complaints filed to NYC's 311 services can reveal ongoing concerns related to sidewalks. However, these complaints mostly come from certain segments of society, resulting in highly disproportionate and asymmetric data that do not paint a clear and precise picture of the state of the pedestrian infrastructure. On top of that, complaint-reporting services such as 311 can often be seen by communities (particularly disadvantaged ones) as non-responsive, leading to an increasing discredit in the capacity of the public agents to enact change.

A more recent alternative is to then *automatically* assess the state of the pedestrian infrastructure, with minimal on-site human intervention and leveraging data and machine learning techniques. While these approaches present challenges in and of themselves, from the cost of data collection to the accuracy and reliability of models, they point towards a horizon where municipalities and communities can paint a more vivid picture of the state of a city's infrastructure. Also importantly, new methods and techniques can cover large spatial regions of the city and, given the unbiased nature of many leveraged datasets, offer an opportunity to address concerns that heavily impact underrepresented communities.

1.3 Motivation

This dissertation is motivated by the lack of information regarding pedestrian infrastructures and facilities, despite their indispensable role in shaping the urban experience. Pedestrian infrastructure has a significant impact on the everyday life of people, specifically those with special needs, for whom such infrastructures are the primary means of accessing public spaces (Qin et al., 2018; M. Saha et al., 2019). Ironically, this data shortage exists in the face of all the advances in data collection, storage, and management, and in an era where "Big Data" and "data availability" has become the recurring theme of the majority of recent academic publications across various fields. The juxtaposition of the dearth of information about pedestrian facilities and the breadth of various data about vehicular infrastructure

draws a vivid picture of the state of disinvestment on this matter. As the findings of a recent study on the availability of pedestrian infrastructure data across 178 municipalities in the United States show, most municipalities do not collect and maintain data on environmental accessibility features (Deitz, 2021). The inadequacy of pedestrian infrastructures to meet the needs of different groups of people, including those using wheelchairs and other forms of mobility aids, cannot be properly addressed if the extent of the problem remains unknown, its changes untraceable, and any effort to address the issue being limited to the number of publicly available datasets. The existing datasets are collected mainly by resourceful cities, with substantial variation in the extent of data and attributes and inconsistent methods from place to place, which creates significant barriers to conducting comparative studies, or data integration, across administrative borders (Deitz, 2021; Hou & Ai, 2020; Louch et al., 2020).

Aside from the data availability issue, there is a gap between the real needs of researchers and practitioners directly studying urban problems, such as accessibility, and the urban analysis tools being developed. The current state of practice and research in urban science and analytics often suffers from a lack of understanding of urban systems' concepts and theories since most of such tools are developed by people who do not have proper training in the domain fields such as urban sociology, geography, urban planning, and design, and urban economics, hence, fail to account for the deeper connections between different forces shaping the cities (Boeing, 2020; Gahegan, 2018; Kontokosta, 2018).

Standing at the intersection of economics, urban planning, and computer science, my research aims at addressing this gap by providing a set of theory-rich tools and methods for large-scale assessment of the quality of urban sidewalks. Designing such tools and methods promotes equitable access to urban data and mitigates problems created by unequal distribution of investments and poor governance in introducing pedestrian-level data collection projects. It also enables urban planners and researchers, practitioners, and municipal decision-makers to have a more realistic image of the everyday challenges of

pedestrians with connectivity, accessibility, and walkability of sidewalks. Moreover, it can pave the way for monitoring the compliance of these infrastructures with the official codes and guidelines designed to serve diverse groups and adopting a complex system approach to tackle the pressing challenges by combining such information with various socio-demographic, environmental, or economic data.

My goal is to address two main challenges that have created significant barriers to the research and practice of pedestrian infrastructure assessment and planning.

1. Lack of scalable, easy to implement, and standardized method to map sidewalk networks (B. Kang et al., 2021; Rhoads et al., 2020)

2. Scarcity of fine-level datasets describing sidewalk features (Deitz, 2021; Louch et al., 2020; Pratt et al., 2012; M. Saha et al., 2019).

Even if the fine-level data scarcity is addressed, it cannot be used to its full potential without having a comprehensive map of pedestrian networks allowing us to study sidewalks and pedestrian infrastructures from a complex systems perspective in relation to other location-dependent factors (Rhoads et al., 2020).

I use state-of-the-art techniques in computer vision to design frameworks and tools for analyzing pedestrian facilities instead of relying on the general pre-trained models with sub-optimal performance in this domain. In doing so, I address some of the challenges of semantic segmentation models regarding the high cost of image annotation by employing different techniques, such as active learning to offer solutions tailored to the specific qualities of urban problems at hand.

1.4 Significance of the Study

The studies presented here were shaped around the dire need to create an inclusive, accessible, safe, and healthy environment for pedestrians and aided by the advent of new

techniques in data processing and management and computer vision that make it possible to analyze the pedestrian environment at human scale as well as city and global scales.

To this end, this work **1)** proposes a standard computer vision-based method for mapping sidewalk networks from high-resolution sub-meter satellite images; **2)** demonstrates a state-of-the-art computer vision approach to extract a select set of build environment features at large-scale and with significantly lower costs, using the publicly available data, and creates generalizable models for automated auditing of pedestrian facilities, specifically sidewalks; **3)** addresses some of the challenges related to the high cost of data annotation for semantic segmentation for both human-scale and city-scale analysis, and **4)** uses the developed models to create pedestrian level sidewalk data sets at scale.

1.5 Organization of the Dissertation

This dissertation is organized based on the three-article format dissertation. The three chapters following the literature review are dedicated to each of the three articles.

Chapter 1 gives an overview of the motivation, identified problems and gaps to address, and the significance of the study.

Chapter 2 provides a review of the conventional and emerging methods of measuring sidewalk attributes. The last part of this chapter investigates related works on semantic segmentation and active learning method, looking at emerging trends and how they have been applied in urban contexts. Aside from this, each article in the presented chapters includes a relevant literature review.

Chapter 3 focuses on the first challenge and proposes a scalable computer vision approach for generating sidewalk network datasets from aerial imagery. I detail the method used to construct the pedestrian network from aerial images, converting the predictions of the semantic segmentation model into georeferenced polygons, and finally, creating the network representation from the polygons. The whole study is formed into a journal article, with the same format presented here, and is submitted to *Computers, Environment, and*

Urban Systems journal and is now under review (Hosseini et al., n.d.).

Chapter 4 addresses the second identified challenge and investigates the application of computer vision to extract the fine-scale features of the sidewalks from street-level images. In this chapter, I detail the framework and method used to extract one of the most challenging features of sidewalks, surface material. I chose this feature due to its multi-pronged importance, specifically for urban sustainability planning and accessibility analysis. Due to the challenging nature of texture segmentation, automated methods to extract sidewalk surface data have remained fairly unexplored. The high within-class variability and between-class similarity of surface materials in an urban setting, present unique challenges requiring both technical and domain specific knowledge. This study is published in the *Sustainable Cities and Society* journal in January 2022 (Hosseini, Miranda, et al., 2022).

Chapter 5 Presents an extended version of the third paper, which draws upon the results of the previous two studies and offer a more comprehensive analysis of pedestrian infrastructure with a threefold understanding of *where* sidewalks are, *how* they are connected, and *what* their condition is. To do that, I used the sidewalk accessibility data from Project Sidewalk (M. Saha et al., 2019), together with the pedestrian network and surface material data for Washington DC, to map and assess sidewalks for people with disabilities and create different visualizations of sidewalk connectivity and accessibility patterns across neighborhoods with varying socioeconomic conditions. The work also addresses some of the challenges raised in Chapter 3 regarding topology correction. The paper was accepted to CVPR AVA (Accessibility, Vision, and Autonomy Meet) workshop as a poster presentation and short paper (Hosseini, Saugstad, et al., 2022).

Chapter 6 is the conclusion section, where I put all the pieces together and talk about the contribution of the presented studies, their potential implications in practice to solve real-world problems, their limitations, and the future direction of my research.

CHAPTER 2
LITERATURE REVIEW

In this chapter, I review the previous studies on measuring sidewalk attributes, both conventional and emerging methods, focusing on computer vision techniques, specifically semantic segmentation and active learning.

2.1 Empirical Studies on Sidewalks

As inter-disciplinary research examined how the built environment can influence health (B. J. Lee et al., 2009; Nickelson et al., 2013), safety (Aghaabbasi et al., 2018; Asadi-Shekari et al., 2015; Forsyth et al., 2008; Naik et al., 2014), social inclusion (Bise et al., 2018; Thornton et al., 2016), social capital (Rogers et al., 2013), objective measurement of the qualitative built environment features has become of key interest.

Systematic observations or *audits* have been designed and developed to measure different attributes of the built environment for a specific purpose, such as measuring accessibility, walkability, impact on physical activity (Pikora et al., 2002), or assessing the pedestrian streetscape (Cain et al., 2012), by providing a list of features to be assessed together with the objective measurements of each (Aghaabbasi et al., 2018; Clifton et al., 2007; S. Lee & Talen, 2014).

A review of 25 pedestrian indices identified sidewalk presence, width, paving materials, and running slope as important features for creating walkable neighborhoods (Maghelal & Capp, 2011). The design, quality, and accessibility of sidewalks are found to impact pedestrian fatalities (Retting et al., 2003), perception of safety (Ariffin & Zahari, 2013), willingness to walk (Katzmarzyk et al., 2018), physical activity (Forsyth et al., 2008; Williams et al., 2005), and risk of certain diseases such as cardiovascular or respiratory diseases (Diez Roux, 2003; Sallis et al., 2012).

Pedestrian Level of Services (PLOS) was one of the first attempts to quantify the quality of pedestrian infrastructure. Calculating PLOS dates back to 1974, when Lautso and Murole (Lautso & Murole, 1974) first introduced this concept to study the impact of the built environment on pedestrian facilities. PLOS is one of the simplest and easiest to calculate quality measurements. To calculate the PLOS for the sidewalk, only two measurements are required: counts of pedestrians per minute passing a given location and the effective width of the sidewalk. Although easy to calculate and interpret, PLOS fails to capture the complex nature of pedestrian experience in that space and reduces it to merely sidewalk width (Bloomberg & Burden, 2006; Jaskiewicz, 2000).

Walkability is probably the most recurrent theme among the empirical studies of sidewalks (Ewing & Handy, 2009; Ewing et al., 2006; Maghelal & Capp, 2011; McCormack et al., 2017). A walkable neighborhood in many studies refers to a dense, highly accessible area with essential destinations within walking distance from where people live or work. Transit-Oriented Developments (TODs) are designed with such an approach (Greenwald & Boarnet, 2001; McKibbin, 2011; Olaru & Curtis, 2015). Multiple audit tools have been developed to measure different features of sidewalks that are believed to be correlated with walkability (Aghaabbasi et al., 2018; Frackelton et al., 2013; S. Lee & Talen, 2014; Millington et al., 2009). But selection bias should also be considered in such analysis since it can very much be the case that people who are more active and generally tend to walk more choose to live in more walkable areas (Boone-Heinonen et al., 2010).

Some longitudinal studies accounted for this bias by observing the behavior of the same population over time (preferably pre and post-move to a more walkable neighborhood) and controlling for unmeasured characteristics (Krizek, 2003; McCormack et al., 2017).

Collecting comprehensive and fine-level sidewalk data using conventional methods is cost-prohibitive. According to Hou and Ai (2020), as of 2019, only 17 sidewalk inventories were created for cities in the United States, even though the Americans with Disabilities Act (ADA) requires all state and local transportation agencies to collect data on some spec-

ified key features of sidewalks, such as width, slope, and paving materials (Department of Justice, 2010).

2.1.1 Auditing the Built Environment

To analyze the impact of the built environment on any of the above-mentioned fields, its properties should be quantified and for this purpose, different systematic observations, or audits, have been designed and developed. Aside from measuring the measurable properties, audits try to quantify the qualitative ones, such as beauty, design, and safety as well. In doing so, each larger qualitative category is divided into smaller properties and then scored based on a different scale native to the specific tool. For instance, for measuring the safety, lighting, sky exposure, width of the sidewalk, obstructions, and many other properties are recorded. Then, based on the approach and theoretical framework of the study, each property will receive a respective weight, showing its importance in forming the overall score for that broader concept. It is often very difficult to compare the score resulting from different audits of the same location, the reason being that the data collection process can differ, the time frame mostly differs, the auditors are not the same, and in many cases, often the terminology used can be confusing as different audit tools often use different words for the same concept Marshall and Garrick, 2010 S. Lee and Talen, 2014.

Pedestrian Level of Services (PLOS) was one of the first attempts to quantify the quality of the built environment. Calculating PLOS dates back to 1974, when Lautso and Murole Lautso and Murole, 1974 first introduced this concept to study the impact of the built environment on pedestrian facilities. Highway Capacity Manual (HCM), which provides tools and guidelines to evaluate transportation facilities, suggests using PLOS to evaluate the condition of pedestrian facilities, such as sidewalks, and decide whether any actions should be taken about them. PLOS is one of the simplest and easiest to calculate quality measurements. To calculate the PLOS for sidewalks, only two measurements are required: counts of pedestrians per minute passing a given location and the effective width of the

sidewalk. HCM provides a chart where planners can evaluate how the flow rate of the sidewalk is ranked. Despite being extremely easy to calculate and interpret, PLOS fails to capture the complex nature of pedestrian experience in that space and reduces it to merely sidewalk width Bloomberg and Burden, 2006; Jaskiewicz, 2000.

2.1.2 Conventional Methods for Measuring Sidewalks

The most conventional and widely used method for assessment of the physical conditions of sidewalks is in-field auditing, and the majority of the popular auditing tools are developed for this audit method. As the name indicates, it requires trained auditors to be present in the field, recording their observations based on different auditing protocols and, in many cases, performing some level of on-site assessments (Nickelson et al., 2013; Sampson, 2012). The in-person visit requirement imposes major limitations on both the geographical coverage and time of data collection; the cost of training, quality control, or recalculation of erroneous data is quite high (S. Lee & Talen, 2014). Moreover, the number of features that can be measured and assessed should be limited since timely and tiring data collection can impact the quality of the data collected.

2.1.3 Emerging Methods for Measuring Sidewalks

With the advent of new computer vision techniques and the availability of street-level images from different cities worldwide, the research towards quantifying the urban built environment has taken a new direction to create semi or fully automated virtual auditing tools for different purposes. The ultimate goal in designing automated audit tools is automating the inference on urban built environment features, achieving higher scalability and more uniform analysis compared to manual or semi-automated audits. Due to the subjective nature of some assessment tasks, maintaining a constant and consistent rating scheme among different auditors can be challenging (Aghaabbasi et al., 2018; Frackelton et al., 2013).

Street-level images have gained popularity as a virtual audit tool due to being a free,

publicly available, and easy-to-use platform, which can overcome some of the limitations of the manual method (Kelly et al., 2013; Phillips et al., 2017; Rundle et al., 2011; Wilson et al., 2018). Using these tools, auditors can virtually explore the designated area and record the features of interest according to the auditing protocols. This would save travel time and costs, significantly expand the study's geographical area, and make quality control of the collected data much easier. Of course, only visual features can be recorded using this method. For features like noise level, odor, and as such, other datasets should be used . (Charreire et al., 2014; Rzotkiewicz et al., 2018; Shatu & Yigitcanlar, 2018).

However, specific problems are associated with virtual audits using tools such as Google Street View (GSV). Despite the extensive international coverage, the image capture frequency differs across different locations. The images are more frequently updated in developed countries compared to the developing ones, resulting in a more significant time gap in cross-country studies compared to intra-country cases (Charreire et al., 2014; Curtis et al., 2013; Rzotkiewicz et al., 2018). This inconsistency in image capture date can become more problematic in quality assessment studies, where the focus is on fine-level features such as sidewalk obstacles, litter, signage, or surface condition (Wilson et al., 2018).

Computer vision techniques have been widely applied to street-level imagery to map openness in cities (X. Li et al., 2017), assess street-level urban greenery (X. Li et al., 2015; Ye et al., 2019), extract land use information from the built environment (X. Li et al., 2017), measure the visual quality of street space (Tang & Long, 2019), visual enclosure (Yin & Wang, 2016) and sky exposure (Carrasco-Hernandez et al., 2015), and to detect traffic signs (Balali et al., 2015; Campbell et al., 2019), urban landmarks (Lander et al., 2017), pavement defects (Cao et al., 2020; Guan et al., 2021; Jenkins et al., 2018; Ma et al., 2017; Nolte et al., 2018; L. Zhang et al., 2016), and curb ramps (Hara et al., 2014).

Image classification and object detection have been frequently used in urban analysis (Campbell et al., 2019; Kharazi & Behzadan, 2021; Law et al., 2018; Miranda, Hosseini, et al., 2020; Nolte et al., 2018). However, semantic segmentation, which provides

pixel-level predictions for object classes, has remained relatively under-utilized. Pavement material classification has been used in safety and route-finding applications to alert pedestrians of upcoming obstacles (H. Kang & Han, 2020; C. Sun et al., 2019; K. Sun, Xiao, et al., 2019; Theodosiou et al., 2020) and to help the visually impaired in identifying street entrances based on the change in surface materials captured by cellphones (Jain & Gruteser, 2018).

2.2 Semantic Segmentation

The rise of autonomous vehicles and self-driving cars created significant demand for fast and efficient algorithms that can extract both high and low-level information from urban scenes, leading to notable improvements in the field of scene parsing, specifically pixel-wise classification, commonly referred to as *Semantic Segmentation*. This method makes dense predictions inferring labels for each pixel of an image, hence, giving each one a semantic meaning (Ess et al., 2009; Geiger et al., 2012).

Early work incorporated multi-resolution processing into segmentation architectures to improve performance over a static resolution approach (H. Zhao et al., 2017). This has been followed by rapid developments in multi-scale pyramid-style networks (Ding et al., 2018; J. He, Deng, & Qiao, 2019; J. He, Deng, Zhou, et al., 2019). In particular, HRNet (K. Sun, Zhao, et al., 2019; J. Wang et al., 2020) connects high-to-low resolution convolutions via parallel and repeated multi-scale fusions to better preserve low-resolution representations alongside the high-resolution ones in comparison to previous works (Y. Chen et al., 2018; Newell et al., 2016; Yu et al., 2018). A variant of HRNet, HRNet-W48, has shown superior performance across segmentation benchmarks such as Cityscapes (Cordts et al., 2016) and Mapillary Vista (K. Sun, Xiao, et al., 2019), which is used as a key component of this proposal's segmentation framework.

Attention-based mechanisms have been adopted in multiple semantic segmentation architectures (L.-C. Chen et al., 2016; Fu et al., 2019; Q. Huang et al., 2017; H. Li et al.,

2018). Instead of feeding multiple resized images into a shared network and merging the features to make the prediction, which can lead to sub-optimal results, the attention mechanism learns to assign different weights to multi-scale features at a pixel-level and uses the weighted sum of score-maps across all scales for the final prediction (L.-C. Chen et al., 2016). Q. Huang et al. (2017) proposed RAN, a reversed attention mechanism that trains the model on the features which are not associated with the target class. The network has three branches that simultaneously perform direct, reverse, and reversed-attention learning. Hierarchical multi-scale attention is a network architecture that learns to assign a relative weighting between adjacent scales (Tao et al., 2020). This method has shown to be four times more memory efficient and allows for larger crop sizes that can lead to more accurate results. We adopted this architecture in our network generation pipeline due to its superior performance in detecting both high and low-level features while benefiting from its memory-efficient design.

When applied to urban context (J. H. Kim et al., 2021; R. Wang et al., 2019; F. Zhang et al., 2018; H. Zhou et al., 2021), researchers often forego retraining or fine-tuning their models on their target datasets and rather rely only on publicly-available models pre-trained on datasets such as CityScapes (Cordts et al., 2016), Mapillary (Neuhold et al., 2017), and ADE20K (B. Zhou et al., 2017). This reliance on pre-trained models not specific to the desired task limits analysis of the pre-defined object classes included in those datasets (Ahn & Kwak, 2018). Further, pre-trained models not fine-tuned on domain-specific data can yield sub-optimal performance (Azizi et al., 2021).

Emerging work has explored sidewalk surface material classification via patch-level sampling classification (A. Ferreira & Giraldi, 2017; Ran et al., 2019). Another group of studies used images of sidewalk materials taken with cameras directed at the surface (Jain & Gruteser, 2018; Xue et al., 2020). This can significantly limit the scalability of the method. In contrast, CitySurfaces (section 4.3) uses publicly available street-level urban images *in the wild* such that the paving material is only part of the overall scene.

We note that recent progress in self-supervised learning (T. Chen et al., 2020; Grill et al., 2020; K. He et al., 2020) has led to dramatic gains in image classification with limited annotated data alongside large collections of unlabeled images. As the architectures and training strategies for self-supervised classification depart significantly from those used for supervised urban image segmentation, we instead use active learning to address limited data concerns in order to make use of high-performing urban segmentation frameworks.

2.3 Active learning

Deep network training requires a large number of annotated images to achieve generalization. In particular, semantic segmentation has the highest associated annotation cost as every pixel needs identification (Pathak et al., 2015) and modern cameras ubiquitously capture millions of pixels. The substantial cost of accurate annotation restricts the practicality of semantic segmentation on new datasets and tasks relevant to urban analysis (Montoya-Zegarra et al., 2014; Xie et al., 2020).

Active learning aims to achieve high accuracy with minimal labeled data by incorporating human supervision during training. By annotating images that the model struggles on or the most informative samples, fewer labeled instances are required to achieve similar performance when compared to supervised approaches where every image is densely annotated (Settles, 2009). Common methods to active learning for vision usually follow an uncertainty-based (Gal et al., 2017; K. Wang et al., 2016) or a representation-based (Gissin & Shalev-Shwartz, 2019; Sener & Savarese, 2017) approach.

Active learning for semantic segmentation requires measurements of unlabeled image informativeness for segmentation networks (Xie et al., 2020) and includes methods that use the entire image for sampling (Kuo et al., 2018; L. Yang et al., 2017) and region-level methods, which only require informative regions to query unlabeled data (Casanova et al., 2020; Mackowiak et al., 2018). We refer the reader to Settles (2009) for an extensive review of active learning techniques.

CHAPTER 3
PEDESTRIAN NETWORKS

3.1 Introduction

After a century of car-oriented urban growth (Walker & Johnson, 2016), cities around the world are implementing policies and plans that aim to make their neighborhoods and streets more walkable and transit oriented. Renewed attention to walkability is driven simultaneously by the impending climate crisis, public health concerns, and a strive for economic competitiveness. With more than a third of all CO_2 emissions attributable to the transport sector (EPA, 2021), it has become clear that climate goals will not be reached unless urban populations start driving less and relying more on walking and public transportation (Cervero, 1998; Speck, 2013). From a health perspective, more walkable cities have been found to have lower obesity and inactivity-related conditions, respiratory diseases, and lower overall public health expenditures (Frank & Engelke, 2001; Grasser et al., 2013; Zapata-Diomedi et al., 2019). Economically, walkable and transit-served city environments have also become an important draw for a competitive workforce (E. Glaeser, 2010; Moretti, 2012) and now command some of the highest-priced real estates in American cities (Leinberger & Lynch, 2014).

Despite the growing, multi-pronged importance of pedestrian-oriented city design, the necessary geospatial data for pedestrian infrastructure mapping and modeling remains far behind vehicular infrastructure data. Digital mapping of vehicular road networks expanded rapidly in the 1990s, led by Federal legislation (President Clinton 1994), municipal governments' investments, as well as private companies such as Navteq and TomTom that operationalized roadway mapping in cities across the world. Assembly and wide-scale dissemination of such data has been instrumental to numerous technologies that use road

network data as a key input: mapping and routing applications (e.g., Google Maps, TransitApp), transportation service technologies (e.g. Uber, Amazon Prime), urban transportation models and policies (e.g., metropolitan and urban Travel Demand Models, congestion charging systems in various of cities), as well as mobility data specification standards (e.g., Google's General Transit Feed Specification, and the City of Los Angeles' Mobility Data Specification).

Transportation debates are often skewed towards topics rich in data – vehicle throughput, for instance, which is monitored on individual streets in many cities, is a key parameter for new road design and investment. *Not only is comparable data describing pedestrian throughput on sidewalks typically unknown, the locations and types of sidewalks are also rarely mapped or updated, contributing to systemic underinvestment in the pedestrian realm.* When pedestrian accessibility is analyzed, it is often done using simplified road-centerline data, not the actual pedestrian infrastructure–sidewalks, footpaths, and road crossings (S. Liu et al., 2021). A number of studies have highlighted the inadequacy of using street-centerline networks for pedestrian routing (Cambra et al., 2019; Qin et al., 2018; C. Sun et al., 2019), which can lead to inaccuracies (e.g., streets with no sidewalks), simplifications (e.g., assumptions that buildings can be directly accessed on both side of a street centerline, while in reality crossing a street is only allowed at certain locations), and misrepresentation (e.g., assuming pedestrian connections based on vehicular routes, where there are none) (Chin et al., 2008; Ellis et al., 2016). Not only can road-network data be imprecise for pedestrian needs, it can also be hazardous for the more vulnerable street users, such as vision-, hearing- or mobility-challenged travelers, wheelchair-bound travelers, the elderly, and the young (M. Saha et al., 2019; H. Zhang & Zhang, 2019).

Aside from navigation purposes, in the absence of comprehensive pedestrian network data, researchers also used road networks to analyze different features of pedestrian infrastructure, which, as discussed, does not provide a true representation. The inadequacy of street centerline to represent the pedestrian network is mentioned in multiple studies (Cam-

bra et al., 2019; Chin et al., 2008; Qin et al., 2018; K. Sun, Xiao, et al., 2019). For instance, Chin et al. (2008) use examples from four metropolitan suburbs in Perth, Western Australia, to show the impact of using actual pedestrian network data instead of the street network in walkability analyses and finds a significant difference between the two networks in measuring connectivity and walkability of neighborhoods.

To address these challenges, we introduce TILE2NET –an end-to-end framework for automated mapping of pedestrian infrastructure using aerial imagery. TILE2NET enables users to download orthorectified sub-meter resolution image tiles for a given region from public sources and generate topologically interconnected, georeferenced sidewalk and crosswalk centerlines as well as sidewalk, road, and crosswalk polygons. Our goal is to map pedestrian networks "as they are" rather than trying to improve the network connectivity artificially. To achieve this, we use a semantic segmentation model that can detect sidewalk, footpath, and crosswalk polygons from orthorectified tiles. We then use the resulting polygons to create an interconnected network. We pilot tested the approach in Manhattan, NY, Washington, DC, Boston, and Cambridge, MA, and achieved high accuracy in each of these cities. The model can be finetuned based on the topological characteristics of different datasets and cities.

Our key contributions are as follows:

1. We provide an end-to-end, open-source framework to create large-scale pedestrian networks from orthorectified imagery(link omitted to satisfy double-blind review requirements).

2. The framework also generates georeferenced polygons of roads, sidewalks (including footpaths) and crosswalks.

3. We offer techniques for the automated creation of annotation masks, using publicly available or user input datasets to train the semantic segmentation models.

4. Our generalized pedestrian feature detection model–made publicly available–is trained

on a selected number of cities with varying street network geometries, building shadow densities, and tree covers (Cambridge, Washington, DC, and New York City parks), making it applicable for other cities with similar environments without any need for additional training.

5. Our solution is adjustable to different city environments, offering various settings to finetune the model on the new dataset, based on the local characteristics of the data.

3.2 Literature Review

3.2.1 Map generation

At least five different frameworks for mapping sidewalk infrastructure can be disguised in existing literature and practice, with additional combinations thereof. The main differentiating point between these five categories lies in the method used to detect pedestrian infrastructures such as sidewalks, footpaths, and crosswalks. Figure 3.1 offers an illustrative summary of these methods.

First, physical site surveys and manual aerial imagery surveys have been used in a number of cities to develop datasets on pedestrian facilities (e.g., in Melbourne, Singapore, and Boston). This involves tracing observable sidewalks and crosswalks from georeferenced aerial imagery, combined with on-the-ground observation and validation (Proulx et al., 2015). Such mapping efforts can produce accurate and high-quality results, but it can also be prohibitively labor intensive and difficult to scale across large regions. In a recent study, 6,400 intersections in San Francisco were manually reviewed and classified based on the crosswalk presence and condition, which took 90 hours for a researcher to complete (Moran, 2022). Some cities have relied on crowd-sourcing sidewalk mapping to a community of online users (Sachs, 2016). Custom-built mapping platforms, such as OpenSidewalks (TCAT, 2016), WalkScope (Placematters and WalkDenver, 2014), or global open-access platforms like OpenStreetMap, enable users to view and edit available

datasets collectively. How these open-sourced data is generated can vary, but can also include the methods described in this section.

Second, network buffering uses a geospatial road centerline network as a reference, which is offset on both sides to generate polygons whose boundaries approximate the right-of-way of the roadway. In this method, which is a widely used and a common approach in geo-information processing, the boundaries of the resulting polygons are considered as the approximate location of the sidewalks segments, assuming that (1) pedestrian path segments only exist along roads, (2) sidewalks exist along both sides of selected roads, and (3) crosswalks are located at every intersection. Buffer distances can include road right-of-way or road-width dimensions from the vehicular road centerline network dataset. After sidewalk segment geometries are generated, crosswalks can be added by linking the endpoints (i.e., intersections) of the assumed sidewalk intersections perpendicularly across road centerlines (Brezina et al., 2017; Karimi & Kasemsuppakorn, 2013). This approach has several shortcomings, first is the limited extent of the locations such a network can cover. A network constructed based on streets and roads does not include off-road footpaths, pedestrian bridges, skywalks, or underground tunnels. In other words, it is limited to only where roads can go and can generate arbitrary sidewalks and crosswalks, which can lead to inaccuracies (e.g., all streets will have sidewalks on both sides), simplifications (e.g., assumptions that buildings can be directly accessed on both sides of a street centerline, while in reality crossing a street is only allowed at specific locations), and misrepresentation (e.g., assuming pedestrian connections based on vehicular routes, where there are none) (Chin et al., 2008; Ellis et al., 2016), each of which can lead to potentially hazardous situations for pedestrians, specifically the more vulnerable population (M. Saha et al., 2019).

Third, pedestrian pathways have also been identified from Global Positioning System (GPS) trajectories of pedestrian movement. This can include data from designated GPS tracking devices that are handed out to consenting participants or collected from their smartphone tracking Apps (Cottrill et al., 2013). Third-party data aggregators, such as

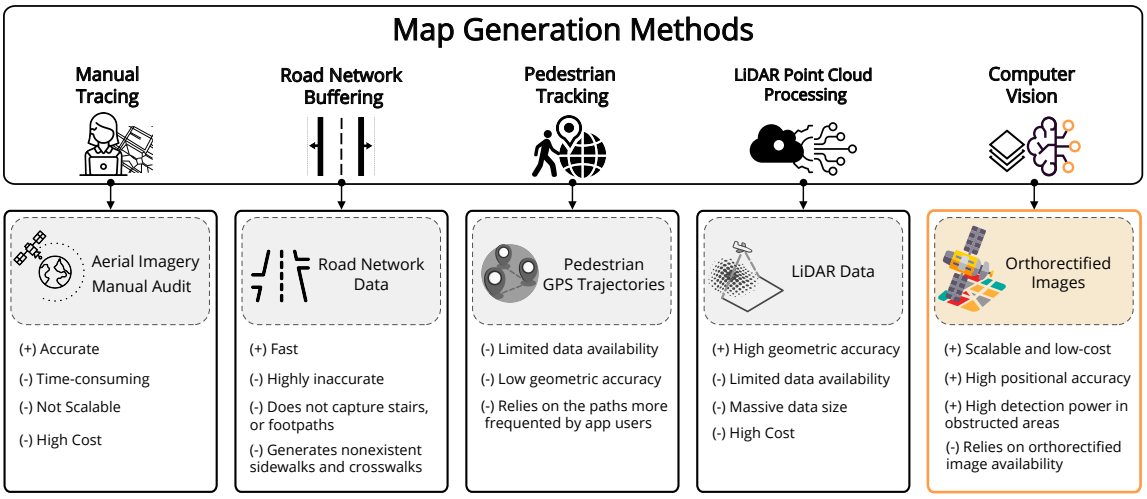

Figure 3.1: Different methods of map generation. Each box presents the main data sources (shaded parts), as well as the strengths (+) and weaknesses (-) of each method. The last box highlighted in orange denotes the method used in this paper.

StreetlightData and Cuebiq collect GPS trace data from hundreds of different Apps that track their users' location history. Once collected, GPS traces can be merged, simplified, and joined into contiguous network datasets (Kasemsuppakorn & Karimi, 2013). The results can effectively illustrate where people (or at least App users) actually walked, but they may ignore segments not frequented by smartphone or app users (X. Yang et al., 2020). Moreover, the accuracy of the final network relies heavily on the positional accuracy of the GPS trajectories, which can be noisy, specifically in locations such as the vicinity of high-rise buildings (Karimi & Kasemsuppakorn, 2013).

The fourth category is LiDAR point cloud processing, which utilizes airborne Light Detection and Ranging (LiDAR) point clouds data. LiDAR devices use active sensing and can be fixed or mounted on mobile objects such as planes and drones (Cura et al., 2018). In general, three main methods have been used for processing LiDAR point cloud data to extract road and sidewalk features. 1) Geometry-based methods, which uses prior knowledge of the unique geometrical shapes and measurements of urban ground elements. 2) Reflectance-based methods utilize the reflectance intensity of different classes of objects to classify the data. The classified points are then normalized based on the laser scanning

model, and distance projection is used to create a saliency map. These two methods are often combined to more accurately extract the streetscape features. 3) Scan-based methods take advantage of the scanning pattern to connect the results from consecutive scans into a continuous boundary and refine the segmentation (Ai & Tsai, 2016; Baker & Hou, 2019; Balado et al., 2018). In clustering feature classes, pedestrian path segments are typically assumed to be made of concrete, and parking lots of asphalt (Hou & Ai, 2020; Karimi & Kasemsuppakorn, 2013; Kasemsuppakorn & Karimi, 2013), which as shown by Hosseini, Miranda, et al. (2022), is not the case in many cities. The resulting data represents sidewalks as vector lines or polygons that can be both accurate and scalable (Horváth et al., 2022; Treccani et al., 2021). Unlike aerial imagery, LiDAR data can be acquired during different hours (day and night), and the data is already georeferenced. However, the lack of spatially dense, universal LiDAR data has limited this approach to relatively few cities overall.

Fifth, and in line with our work, different computer vision techniques have more recently been deployed in a limited number of studies to detect pedestrian infrastructure from aerial images (Ning et al., 2022). The detected features are then converted into georeferenced lines or polygons and go through topological corrections to produce the final network. Among computer vision techniques, semantic segmentation can result in highly accurate detection and localization of infrastructure elements. This method makes dense predictions inferring labels for each pixel of an image, hence, giving each one a semantic meaning (Ess et al., 2009; Geiger et al., 2012). To construct a pedestrian network, a segmentation model is first trained to detect different features of the streetscape, such as roads, sidewalks, and crosswalks, from aerial images. Although semantic segmentation has been broadly used to detect roads and building footprints from aerial images (Balali et al., 2015; Iglovikov et al., 2017; W. Li et al., 2019) and to create road networks (Bastani et al., 2018; Etten, 2020; Wei et al., 2019), it has not been widely implemented for sidewalk mapping so far, possibly due to several technical challenges. First, in order to achieve satisfactory

results, semantic segmentation algorithms need to be trained on densely annotated labels, which can be labor-intensive and costly to prepare. Consequently, in applying semantic segmentation models to urban context (J. H. Kim et al., 2021; R. Wang et al., 2019; F. Zhang et al., 2018; H. Zhou et al., 2021), researchers often forego retraining or fine-tuning their models on their target datasets and rather rely only on publicly-available models pre-trained on datasets such as CityScapes (Cordts et al., 2016), Mapillary (Neuhold et al., 2017), and ADE20K (B. Zhou et al., 2017). This reliance on pre-trained models, not specific to the desired task, limits analysis to the pre-defined classes included in those datasets (Ahn & Kwak, 2018). Further, pre-trained models not fine-tuned on domain-specific data can yield sub-optimal performance (Azizi et al., 2021). Second, compared to roads and buildings, detecting sidewalks, footpaths and crosswalks is more challenging since they constitute a small portion of the visual information of aerial images, and their detection can be further inhibited by occlusion from shadow, vegetation, and structures such as bridges or tall buildings (Hosseini et al., 2021). Hence, choosing the right network architecture that can preserve the fine local details while taking the global image context into account is crucial.

3.2.2 Semantic segmentation

The rise of autonomous vehicles and self-driving cars created significant demand for fast and efficient algorithms that can extract both high and low-level information from urban scenes, leading to notable improvements in the field of scene parsing, specifically pixel-wise classification, commonly referred to as *semantic segmentation*. Early work incorporated multi-resolution processing into segmentation architectures to improve performance over a static resolution approach (H. Zhao et al., 2017). This has been followed by rapid developments in multi-scale pyramid-style networks (Ding et al., 2018; J. He, Deng, & Qiao, 2019; J. He, Deng, Zhou, et al., 2019). In particular, HRNet (K. Sun, Zhao, et al., 2019; J. Wang et al., 2020) connects high-to-low resolution convolutions via parallel and repeated multi-scale fusions to better preserve low-resolution representations alongside the

high-resolution ones in comparison to previous works (Y. Chen et al., 2018; Newell et al., 2016; Yu et al., 2018). A variant of HRNet, HRNet-W48, has shown superior performance across segmentation benchmarks such as Cityscapes (Cordts et al., 2016) and Mapillary Vista (K. Sun, Xiao, et al., 2019), is used as a key component of this proposal's segmentation framework.

Attention-based mechanisms have been adopted in multiple semantic segmentation architectures (L.-C. Chen et al., 2016; Fu et al., 2019; Q. Huang et al., 2017; H. Li et al., 2018). Instead of feeding multiple resized images into a shared network and merging the features to make prediction, which can lead to sub-optimal results, the attention mechanism learns to assign different weights to multi-scale features at a pixel-level and uses the weighted sum of score-maps across all scales for the final prediction (L.-C. Chen et al., 2016). Q. Huang et al. (2017) proposed RAN, a reversed attention mechanism that trains the model on the features which are not associated with the target class. The network has three branches that simultaneously perform direct, reverse, and reversed-attention learning. Hierarchical multi-scale attention is a network architecture that learns to assign a relative weighting between adjacent scales (Tao et al., 2020). This method is shown to be four times more memory efficient and allows for larger crop sizes that can lead to more accurate results. We adopted this architecture in the sidewalk detection part of our pipeline due to its superior performance in detecting both high and low-level features while benefiting from its memory-efficient design.

3.3 Materials and Methods

In this section, we detail the datasets used for training the model, describe our methodology, and discuss how we have addressed the challenges of preparing labor-intensive annotation labels for training the algorithm and generalized it to detect pedestrian infrastructure in different urban environments. We also illustrate how initially detected polygon geometries can be converted into sidewalk centerlines, bringing the outputs closer to a topologically

interconnected network dataset that can be used for pedestrian routing and other network analysis procedures.

Table 3.1: Datasets used for training the model and their sources.

City	Dataset	Features	Date	Source
Cambridge, MA	Sidewalks	Sidewalk polygons	2018	(Cambridge GIS, 2018a)
	Roads	Roads polygons	2018	(Cambridge GIS, 2018d)
	Pavement Markings	Crosswalk polygons	2018	(Cambridge GIS, 2018b)
	Public Footpaths	paved & unpaved	2018	(Cambridge GIS, 2018c)
	Ortho-imagery	Image tiles	2018	(MassGIS, 2018)
Manhattan and Brooklyn	Sidewalk Inventory	Off-road footpaths inside parks	2018	(NYC DoITT, 2018)
	Roads	Road polygons	2018	(NYC DoITT, 2018)
	Ortho-imagery	Image tiles	2018	(NYC GIS, 2018)
Washington, DC	Sidewalk Inventory	Sidewalk and crosswalk polygons	2019	(DC GIS, 2019b)
	Road	Road polygons	2019	(DC GIS, 2019a)
	Ortho-imagery	Orthophoto SID	2019	(DC GIS, 2020)

3.3.1 Data description

The semantic segmentation model requires pairs of aerial images and their corresponding annotation labels to be trained. Two main data sources were used to create our training set: 1) High-resolution orthorectified imagery that is available across numerous U.S. (US Geological Survey, 2018) and international cities, and 2) Planimetric data that is created from orthorectified images. Next, we provide more details about each one and describe how they were used in creating the training data. Table 3.1 shows the datasets used to train the model and their delivery dates.

High-resolution orthorectified imagery

Raw aerial images inherently contain distortion caused by sensor orientation, systematic sensor and platform-related geometry errors, terrain relief, and curvature of the earth. Such distortions cause feature displacement and scaling errors, which can result in inaccurate direct measurement of distance, angles, areas, and positions, making raw images unsuitable for feature extraction and mapping purposes. Orthorectification removes these distortions and creates accurately georeferenced images with a uniform scale and consistent geometry (Tucker et al., 2004; G. Zhou et al., 2005). The orthoimagery tile system also makes it

possible to convert between positional coordinates of tiles in x/y/z (where z represents the zoom level) and geographical coordinates.

Aside from orthoimages provided by U.S. Geological Survey (USGS) (US Geological Survey, 2018), there are some state-wide programs dedicated to producing digital orthoimagery on different zoom levels, which may offer more recent data. For the purposes of this study, we used orthorectified images provided by Massachusetts (MassGIS, 2018), Washington, DC (DC GIS, 2020), and New York (NYC GIS, 2018) to train the model and pilot test the approach. TILE2NET is designed with the capability of automating the data preparation process. It can take as input, the textual name or geographic coordinates of the bounding box of a given region and download the tiles that fall within the bounding box, for the cities where orthoimagery is available.

To create the training data, using TILE2NET, we obtained 11,000 tiles from Washington, DC, 28,000 tiles from Cambridge, and 8,000 tiles from inside NYC parks. Except for Washington, DC, where the tiles are 512x512 pixels, the rest of the tiles come in 256x256 pixels. We choose zoom level 20 for the 256x256 pixel tiles, which corresponds to the zoom level 19 for 512x512 pixels tiles, where each pixel of the image represents 0.19 meters on the surface of the earth. Our experiments training the model with both sizes showed that the model would perform better using 512x512 pixel input images (an increase of roughly 12% in mIoU). Hence, we used the tool to stitch every four neighboring 256x265 pixel tiles to get 512x512 pixel images, creating a total of 20,000 tiles.

Planimetric GIS data

Planimetric mapping involves extracting features from orthoimagery to create maps that only capture the horizontal distance between the features irrespective of elevation (Quackenbush, 2004). Since planimetric data are created using orthorectified images, they are suitable for creating annotation masks–a priori known and accurate raster polygons that describe the features we seek to automatically detect using semantic segmentation mod-

els. An annotation label is like a reference map that corresponds to a given tile, where each pixel color represents the class to which the corresponding pixel in the image belongs (Figure 3.2(b,c,e,d)).

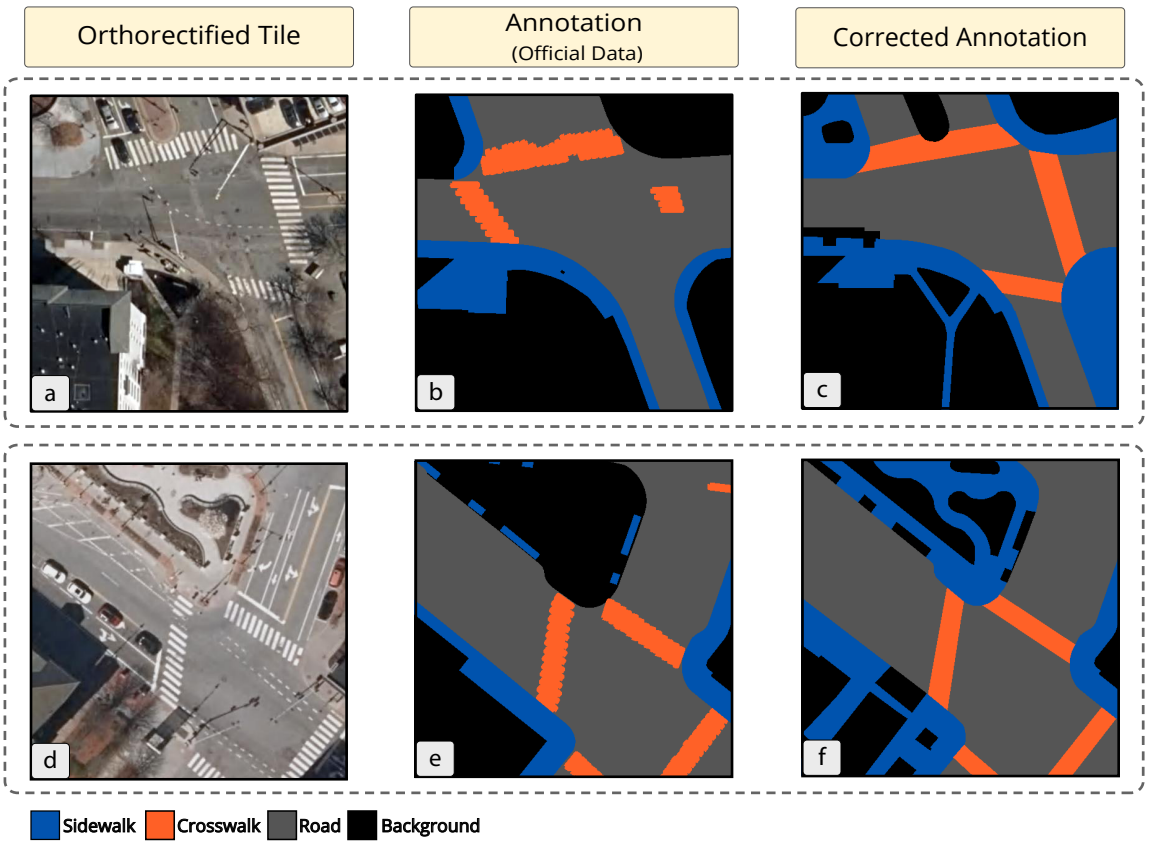

Figure 3.2: Examples of the mismatches between the aerial image and the annotation label created from the official data. The manually corrected annotation labels are shown in the last column.

To prepare the annotation labels, TILE2NET primarily relies on available GIS data on sidewalk, crosswalk, and footpath locations in select city environments. In this study, we used the publicly available planimetric data on sidewalks, footpaths, and crosswalks in parts of Cambridge, Washington, DC, and selected sites from inside the parks of New York City. Reliance on existing GIS datasets allows us to prepare large-scale annotation labels using available data rather than manually annotating a huge number of images. TILE2NET takes the bounding box of each tile, finds the corresponding sidewalk, footpath, crosswalk, and road polygons from the available planimetric GIS data, rasterized the GIS polygons

into pixel regions, and outputs annotated image tiles with four total classes: sidewalks (including footpaths), crosswalks, roads, and background, representing each class with a distinct color. These annotations are used as ground truth data for training the model.

However, challenges remain in creating accurate and consistent training data. The first challenge arises from the lack of consistency between the mapping standards used by different municipalities. Moreover, since GIS data on pedestrian infrastructure does not necessarily reflect the exact conditions that are represented in aerial images, there can be a temporal difference between tiles and GIS data as the creation of GIS data may have relied on a different underlying data source. As illustrated in Figure Figure 3.2, official GIS data can contain numerous errors. Human adjustment and correction may be necessary to bring ground truth annotation labels into alignment with the image data. To achieve that, our research team manually corrected 2,500 tiles of the 12,000 training set, 1,620 image tiles out of 4,000 tiles that were used as our validation set, and 1,500 tiles out of 4,000 test set tiles.

3.3.2 Methods

TILE2NET adopts a multi-scale attention model for detecting pedestrian infrastructure from aerial imagery: sidewalks, crosswalks, stairs, and footpaths that may be separated from streets and roadways (e.g., in parks and open spaces). We combine a semantic segmentation approach with a raster-to-polygon conversion process to generate vector shapefiles of pedestrian infrastructure elements and, separately, a polygon-to-centerline conversion process to produce a topologically interconnected network of pedestrian centerlines. The pipeline has two main parts: 1) Detecting street elements from aerial imagery (Figure 3.3 (a,b)), and 2) Network construction (Figure 3.3 (c,d)). In the following, we describe our methods in detail.

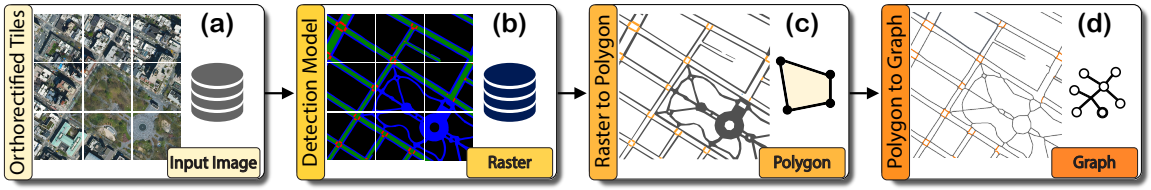

Figure 3.3: The proposed network generation pipeline. **a)** Unlabeled orthorectified tiles are passed through the semantic segmentation model for prediction, **b)**The model detected sidewalks (blue), crosswalks (red), and roads (green) in the input tiles, **c)** The sidewalks and crosswalks of the prediction results (raster format) are converted into georeferenced polygons, **d)** The line representation of the pedestrian network generated from polygons.

Detecting street elements from aerial imagery

To detect street elements from aerial imagery, TILE2NET allows users to train a pedestrian feature recognition model on custom, locally-specific data. The trained model can then be used to make inference on unlabeled data. For our semantic segmentation task, we adopted the Hierarchical Multi-Scale Attention model (Tao et al., 2020), and used HRNet-W48 K. Sun, Zhao, et al. (2019) and J. Wang et al. (2020) with Object-Contextual Representations (Yuan et al., 2019) as the backbone. The computed representation from HRNet-W48 is fed the OCR module, which computes the weighted aggregation of all the object region representations to augment the representation of each pixel. The augmented representations are the input for the attention model. For the primary loss function, we used Region Mutual Information (RMI) loss (S. Zhao et al., 2019), which accounts for the relationship between pixels instead of only relying on single pixels to calculate the loss.

The semantic segmentation model takes an input image, makes dense predictions inferring labels for each pixel, and outputs a feature map showing whether and where the objects of interest are recognized in the image tile. After the training phase is completed, the unlabeled orthorectified tiles are passed through the trained model, as shown in Figure 3.3 (a), the prediction model outputs a raster image where each pixel has a value corresponding to one of our four classes: sidewalk, crosswalk, road, and background (Figure 3.3 (b)).

Figure 3.4: Boston Commons: a) Aerial image, b) Detected sidewalk and footpath polygons (in orange) and detected crosswalks (in red), c) Fitted sidewalk, crosswalk, and footpath centerlines superimposed on the aerial image.

Network creation

After the pedestrian features were detected from the input images, TILE2NET takes the model's prediction in raster format and performs 1) raster to polygon conversion, which can save the output polygons in different formats such as GeoJSON and shapefiles, usable across multiple GIS tools; and 2) polygon to centerline conversion to create the final pedestrian network representation. Figure 3.4 shows the results of these two steps for Boston Commons, which was not part of the training data. Next, we will detail each of these steps.

Raster to polygon conversion

To obtain the vectorized, georeferenced sidewalks, crosswalks, and roads, the detected regions should be converted into polygons. To achieve that, we employed connected-component mapping algorithm (L. He et al., 2009; Rosenfeld & Pfaltz, 1966), in which the connected cells of the same category in the raster image form regions or *raster polygons*. These regions are then georeferenced, using an affine transformation, which preserves lines and parallelism and maps the raster pixels into the geographic coordinates.

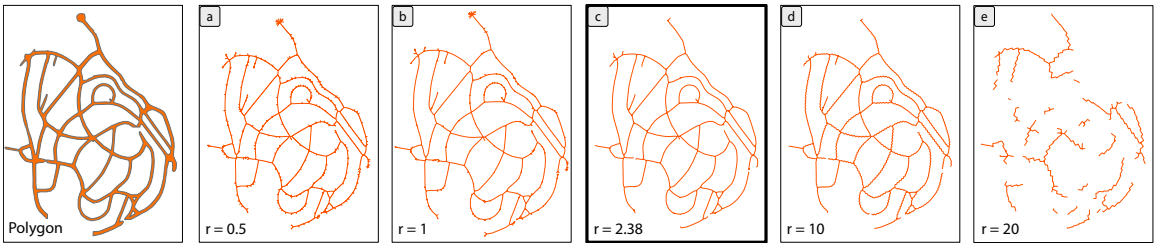

Figure 3.5: Impact of different interpolation distances on the resulting centerline created from the input polygon. Small values create extra branches (r=0.5 and r=1) and large values create zigzaggy (r=10) or disjointed lines (r=20). The middle centerline, highlighted with a thicker border, is computed using the interpolation distance computed using our heuristic approach.

Polygon to centerline conversion

In the third and final step, TILE2NET calculates the centerlines for each polygon. Given that the initially detected regions are pixel-precise, we first simplify the polygons using the Douglas-Peucker algorithm (Douglas & Peucker, 1973). Next, a dense Voronoi diagram is computed to extract the centerlines of the sidewalk polygons (Brandt & Algazi, 1992). The centerline is constructed by linking the internal diagram edges not intersecting with the boundary of the object. The border density parameter, called interpolation distance, densifies the input geometry's border by placing additional points at that given distance. If the interpolation distance is too small, the output will have many unwanted branches, while large values can lead to zigzaggy and disjointed centerlines (Lewandowicz & Flisek, 2020; Z. Li et al., 2021) as illustrated in Figure 3.5.

Finding the optimal interpolation distance is beyond the scope of the current work. To approximate a suitable parameter for each polygon, we used a heuristic approach and selected a sample of 400 polygons of varying areas and perimeters. Next, for each polygon, we tested different interpolation distances ranging from 0.5 to 20, using a 0.5 step (i.e., total of 40 different parameters) and chose the line with the highest connectivity and the least number of extra branches which best represents our irregular shapes. For each polygon, we record the interpolation distance that results in the best centerline, as well as the polygon area, perimeter, average width, number of vertices, area to minimum bounding box area

ratio, and area to perimeter ratio. We used a polynomial regression model and concluded that the area to perimeter ratio is a significant factor in choosing the interpolation distance. Using the derived coefficient, we compute the interpolation distance of each polygon for centerline creation. In Figure 3.5 the centerline highlighted with a thicker border is computed using the interpolation distance derived from our heuristic approach (r=2.38), having smooth lines which follow the form of the input polygons with very few extra branches compared to smaller values. The coefficient can be finetuned on new datasets. To clean and simplify the centerline, we trim branches shorter than an adjustable threshold, which is generally set to half of the average width of the polygon. Crosswalk centerlines were created by joining the centroids of the smaller edges of the minimum rotated rectangles for each polygon. The crosswalk centerlines are then connected to their nearest sidewalk lines. The resulting vector lines form the basis of our pedestrian network.

Following this step, the network goes through algorithmic post-processing operations to correct its topology: removing false nodes and removing the isolated lines. To close the small gaps, we used R-Tree (Guttman, 1984; Kamel & Faloutsos, 1993) and queried for gaps smaller than certain thresholds. Then we extrapolate both lines to meet in the center of the gap. These operations help refine the detected pedestrian centerlines into a topologically continuous network while avoiding undue corrections and additions where connections between sidewalk segments are lacking.

3.4 Implementation and Evaluation of Results

This section presents the implementation details and results of using TILE2NET to create city-scale pedestrian networks. We evaluate the performance of our proposed method in two parts. First, we evaluate the results of our semantic segmentation model based on ground truth masks (subsection 3.4.2). Next, we evaluate the accuracy of the constructed maps, both polygons, and centerlines, using the available official data (subsection 3.4.3). Table 3.2 presents an overview of the available ground truth data used in our

Table 3.2: Availability of the official data across different cities. Training: ○, Evaluation: •

City	Data type	Sidewalk	Crosswalk	Footpath
Boston	Polygon	○	-	-
	Centerline	•	-	•
Cambridge	Polygon	○	○	○
	Centerline	•	•	•
Washington DC	Polygon	○	○	○
	Centerline	-	-	-
Manhattan	Polygon	•	-	○
	Centerline	-	-	•

evaluation. The polygon data was partly used in our training process, denoted by a plain circle, as explained in subsection 3.3.1.

3.4.1 Implementation

The model was trained with a batch size of 16, SGD for the optimizer with polynomial learning rate (W. Liu et al., 2015), momentum 0.9, weight decay $5e^{-4}$, and an initial learning rate of 0.002. The multi-scale setting used 0.5, 1, 1.5, and 2, where a 0.5 scale denotes downsampling by a factor of two, and a scale of 2 denotes upsampling by a factor of 2 (Tao et al., 2020). We used color augmentation, random horizontal flip, random scaling (0.5x–2.0x), and Gaussian blur on the input tiles to augment the training data and improve the generalizability of the model. The crop size was set to 512x512. The image and annotation pairs were split into three parts: 60% of the tiles were used to train the model, 20% of the tiles to validate, and 20% were held-out to test the model in the final stage. To handle the class imbalance, we employed class uniform sampling in the data loader, which chooses equal samples for each class (Y. Zhu et al., 2019) (classes like road and background are present in almost all images, whereas crosswalks can appear less frequently) and the class uniform percentage was set to 0.5. The segmentation model was trained for 310 epochs using 4 NVIDIA RTX8000 GPUs with 48 GB of RAM each.

The trained model is then used to make inference to create the city-scale networks; I obtained the tiles corresponding to the bounding box of Boston, Cambridge, Manhattan,

and Washington, DC, on zoom level 20. Since smaller tiles result in more disjointed final shapes, I used 1024x1024 pixel tiles stitched using TILE2NET for the inference part. The hierarchical architecture of our semantic segmentation network made it possible to choose different scales during the inference. In our experiments using 512x512, 1024x1024, and 2048x2048 pixel tiles during inference, the best results were achieved using 1024x1024 pixel tiles, where the model had enough context to distinguish between different classes.

TILE2NET uses the Geopandas (Jordahl, 2014) and PyGEOS(Wel, Casper van der, 2019) libraries for performing different spatial operations. The raster to polygon conversion was done using the Rasterio library (Gillies et al., 2013). To create the centerlines, I used the Centerline library (Todic, 2016). Momepy (Fleischmann, 2019) was used to handle network cleanups, such as removing the false nodes.

3.4.2 Evaluation of the semantic segmentation results

Table 3.3: Evaluation metrics on the test set.

Label	IoU	Precision	Recall
Sidewalk	82.67	0.9	0.92
Road	86.04	0.91	0.94
Crosswalk	75.42	0.86	0.86
Background	93.94	0.97	0.96
mIoU	84.51		

The trained model outputs four classes in total, two of which were directly used to create the pedestrian networks, i.e., sidewalks and crosswalks, one was used to draw local attributes for finetuning the network creation parameters, and the background, which contains all other elements not used in this study. To evaluate the performance of the model, I used the Jaccard index, commonly referred to as the Intersection over Union (IoU) approach, which is a scale-invariant standard evaluation metric for semantic segmentation tasks. Class-specific accuracy measures are also calculated to assess the model's performance in classifying objects of different classes. I did not rely on the more biased pixel-level accuracy since sidewalks and crosswalks comprise a small portion of each image,

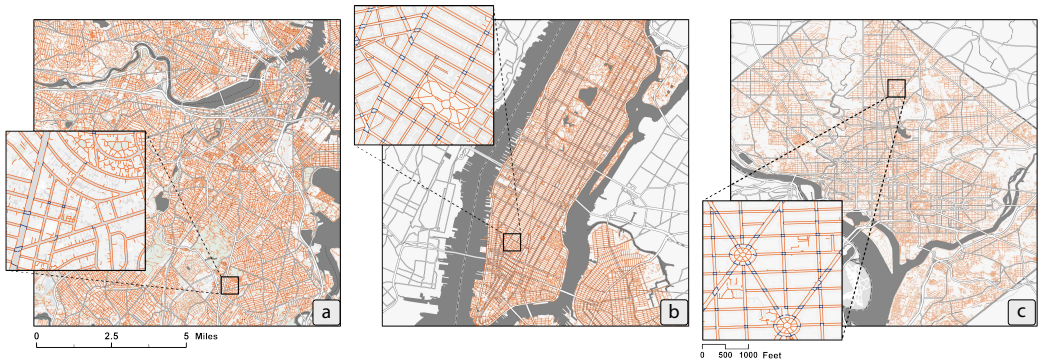

Figure 3.6: Model results showing detected sidewalk, crosswalk and footpath centerlines in a) Boston and Cambridge, b) Manhattan and parts of Brooklyn, c) Washington, DC. The maps are shown at the same scale for comparison.

resulting in a significant class imbalance and an arbitrary high pixel-level accuracy. Table 3.3 presents the average IoU (mIoU), as well as the class-wise IoU, precision, and recall. The model achieved 84.5% mIoU over all four classes, with sidewalks having 82.7% IoU and crosswalks having 75.42% IoU. The lower accuracy of the crosswalks can be attributed to the more temporal nature of the crosswalks and the fact that they can get faded and, in some cases, not even visible to human eyes.

3.4.3 Evaluation of the constructed maps

Figure 3.6 presents the model outputs in Boston and Cambridge, Manhattan, parts of Brooklyn, and Washington, DC. All cities are shown at the same scale for comparison. To evaluate the quality of the output vis-a-vis existing official GIS datasets available in each city. I compared both the detected polygons to corresponding city GIS polygons and the detected network segments to a priori known GIS sidewalk networks in each city. Table 3.2 summarizes the availability of official data across the four cities, and how they were used for both training and evaluation.

For polygon comparisons, comprehensive and public data for sidewalks, crosswalks, and footpaths, was available in Cambridge, and Washington, DC. In Boston, only sidewalk GIS polygons were available, and Manhattan's sidewalk data includes the footpath poly-

gons. Table 3.4 presents class-level evaluation metrics for detected polygons, showing the total count and the percentage of ground-truth polygons (from the cities' GIS data) that had a matching "detected" polygons spatially intersecting each element. In Cambridge, 98.9% of all polygons in official GIS data had overlapped with polygons detected by TILE2NET. In Boston, that number was 98.7%, in Washington, DC, 84.4%, and in Manhattan, 98.2%. Since most of the unmatched polygons were small in size, we also report the area-weighted overlap percentages in Table Table 3.4.

The last row of Table Table 3.4 reports the mean aerial overlap percent between official GIS pedestrian infrastructure polygons and polygons detected by TILE2NET (also weighted by size). This illustrates what percent of the area featured in the official pedestrian polygons overlaps with detected polygons. In Cambridge, 85.9% of the area of official GIS polygons was also covered by detected polygons, 77.9% in Boston, 73.8% in Washington, DC, and 87.5% in Manhattan. Figure 3.4 illustrates an overlay of detected polygons and network segments in a part of Boston covering the Boston Commons and some blocks around it.

To evaluate the accuracy of the networks extracted from the imagery, we compared them against the publicly available sidewalk, crosswalk, and footpath centerline shapefiles of each city, where available (Table 3.2). All three types of pedestrian infrastructure centerlines were available in Cambridge. In Boston, the sidewalk centerline dataset includes

Table 3.4: Comparison of polygon accuracy results in Cambridge, MA, Boston, MA, New York City, NY, and Washington, DC. The % detected indicates what proportion of polygons in the city dataset had a corresponding detected polygon that overlaps with it. Since many of the undetected polygons are small in area, we also report the % detected weighted by area. The mean area overlap % row reports how close in area (from 0-100%) the detected polygons are to the city dataset, on average (including those city polygons that remained undetected).

Measures	Cambridge, MA	Boston, MA	Washington, DC	New York City, NY
Official data polygon count	17,516	24,604	52,087	4,684
Match (overlaps with detected)	17,327	24,288	43,963	4,602
% Detected	98.92%	98.72%	84.40%	98.25%
% Detected (weighted by area)	99.62%	99.39%	97.48%	99.91%
Mean area overlap % (weighted by area)	85.9%	77.9%	73.8%	87.5%

Table 3.5: Comparison of network accuracy results in Cambridge, Boston, and Manhattan.

City	Measures	All	Sidewalk	Crosswalk	Footpath
Cambridge	Official element count	12,792	5,007	2,414	5,371
	Match (within 4m of centroid)	10,631	4,735	2,197	3,699
	% Match	83.1%	94.6%	91.0%	68.9%
Boston	Official element count	110,031	54,864	11,223	37,023
	Match (within 4m of centroid)	86,372	49,806	10,051	23,978
	% Match	78.5%	90.8%	89.6%	64.8%
Manhattan	Official element count	-	-	-	6,239
	Match (within 4m of centroid)	-	-	-	5,309
	% Match	-	-	-	85.1%

crosswalks, and in Manhattan, only footpath centerlines were available for comparison. However, in Cambridge and Boston, centerline data dates back to 2011.To investigate the reliability of the centerline data for evaluation, we analyzed the Cambridge data, where more recent polygon data (2018) are available for both sidewalks and crosswalks. we compute the percentage change of the sidewalk and crosswalk centerlines by intersecting the centerlines of each class with the more recent polygon data of that class. we manually examined all the mismatch cases and removed the false positives. Our analysis showed a 23% change from 2011 to 2018 in crosswalks, while sidewalks change was 9.2%, which shows the relative stability of the fixed features such as sidewalks over time. To perform the evaluation, we marked the centroid of each network segment from corresponding city datasets and buffered the centroid by four meters (corresponding to 95th percentile sidewalk width in Boston) to check how many ground-truth network segments have a detected segment within a 4-meter distance of their centroid. we relied on centroids rather than full segments or endpoints to avoid matching intersecting line segments around network nodes. The results are reported in Table 3.5.

In Cambridge, our model matched 83.1% of all segments, with notable heterogeneity among different types of elements. Among sidewalks, 94.6% of centerlines had a corresponding detected segment, among crosswalks, 91.0%, and among footpaths, 68.9%. The lower matching rates among footpaths were expected due to more frequent tree cover over footpaths in parks and green spaces. Network matching in Boston was fairly similar across

the same network types (Table 3.5). 90.8% of all sidewalk segments in city GIS data and 89.6% of all crosswalks were matched by our results. Footpath matching was again notably lower at 64.8%. In Manhattan, NY, we only had official footpath networks (in parks) available from the city's open data repository. Here, 85.1% of official footpath segments had a corresponding detected segment within a four-meter buffer of their centroid. In Washington, we did not find any official sidewalk centerlines.

For Washington, DC, the comparison could only be performed on more limited data. In Washington, DC, we did not find any official sidewalk centerlines and instead performed the comparison with the available OpenStreetMap sidewalk segments. The results are shown in Table 3.6. A somewhat lower matching rate with OSM networks was expected and confirmed by the 76.9% match across all categories since OSM sidewalk networks are not official data, following different standards than those prepared by city governments. Though our inspection of results confirmed that both sidewalks and crosswalks again matched more closely than footpaths in parks, no type attributes for such comparison were available in the OSM network.

Table 3.6: Network accuracy evaluation in Washington, DC.

City	Measure	All
Washington, DC	OSM swlk element count	11,317
	Match (within 4m of centroid)	8,703
	% Match	76.9%

3.5 Discussion

While the automated pedestrian infrastructure mapping methodology we explored was able to capture a 90% or higher share of sidewalks and crosswalks featured in city GIS datasets, and a notably lower share of footpaths in parks, green areas, and other public spaces, a few caveats need to be highlighted to interpret these results. First, the sidewalk, crosswalk, and footpath data available for validation in Cambridge, Boston, Washington, DC, and New York City are not necessarily temporally concurrent with the aerial imagery we used for

feature detection. This can lead to expected differences between ground truth and detected features. For instance, in Cambridge, the GIS data we used for validation was last updated to reflect the year 2010 flyover conditions according to the city's metadata, but the aerial image tiles we used as input for feature detection were captured in 2018. The Boston sidewalk and crosswalk centerline data were last updated to reflect 2011 conditions, while our Boston image tiles were captured in 2018. Some pedestrian elements in aerial views are therefore not featured in the cities' GIS data and vice versa, possibly because they were altered before or after the images were captured. As also explained in subsection 3.4.3, the percentage change between the data created based on the 2010 flyovers and the 2018 polygon data was 9.2% for sidewalks and 23% for crosswalks.

Second, we also noted errors in the cities' GIS datasets, where pedestrian infrastructure elements were missing or different from the Google Street View conditions dated to the same year. Given that the city datasets were likely prepared with a combination of automated feature detection and human correction, some error is expected. While these were the only data available to construct a quasi-official comparison of our results, these caveats are also partially responsible for the differences between detected and official pedestrian network elements.

The model can be improved with training and validation data that are both temporally and geometrically identical to the conditions captured in the image tiles used for feature detection. If city GIS data is versioned by year, the ground truth GIS data used for training the model could be dated back to an antecedent year that matches the image tiles and additionally humanly corrected to eliminate omissions and errors. This can ensure in future work that the detected polygons best match ground-truth polygons. The relatively lower detection accuracy of footpaths is attributable to several factors. On the one hand, feature detection from aerial imagery is hampered by significantly higher levels of tree cover and other vegetation obstructions over footpaths found in parks, courtyards, and campuses. Second, footpaths also tend to have more complex geometries with winding and non-gridiron

layouts, resulting in a much higher and more detailed segment count than on sidewalks and crosswalks. A complex curving footpath in a park made up of several segments may have a matching detected segments on some but not all of its segmented parts.

The polygon to centerline fitting part could also benefit from further improvement. The network geometry improvements can be categorized into three separate areas. First, as also mentioned in section 3.3, the Voronoi skeleton approach (Brandt & Algazi, 1992) used for converting polygons to centerlines is very sensitive to the interpolation distance parameter and is not optimized for extracting the centerline of elongated polygons. Moreover, the algorithm fits centerlines into discrete polygons and is not optimized for fitting the centerlines such that the endpoints of one skeleton topologically connect to the skeleton of another polygon, resulting in discontinuities between polygons. We were partly able to adjust this with automated post-processing routines, but further refinements would be desirable to output continuous centerline networks. There is an extensive body of literature on various skeletonization algorithms (P. K. Saha et al., 2016), with some focusing solely on creating the centerlines of the elongated polygons (Haunert & Sester, 2008; Lewandowicz & Flisek, 2020). However, finding the optimal interpolation distance value is beyond the scope of the current research, but as a future direction, we are planning to work on developing algorithms tailored for creating the centerlines of the pedestrian infrastructure.

Second, the resulting network segments are currently not optimized to form singular nodes or endpoints at intersections. Some detected line segments often converge near street corners, forming redundant intersections. This can be addressed in future work by improving the algorithmic procedures to join endpoints into a single overlapping endpoint located at the geometric centroid of the multiple nodes found within a given distance. This threshold distance would ideally be determined contextually, depending on the street widths in each area.

Third, though most computer vision solutions are fundamentally unable to detect sidewalk spatial elements where visual obstructions exist, lower detection accuracy in tree-

Figure 3.7: Mapping obstructed pedestrian facilities in different cities: a) Cambridge, MA. - sidewalks are mapped as continuous despite the heavy shadow, b) Manhattan - sidewalks and crosswalks obstructed by tree foliage and shadow are detected and mapped, c) Washington, DC. - crosswalks covered by vegetation are correctly detected and mapped.

covered regions was expected. Nevertheless, since our model was trained on planimetric GIS data, where pedestrian infrastructure elements were present regardless of obstructions, our model performed surprisingly well in occluded areas. Figure 3.7 shows examples of the created network in sample areas of Cambridge, MA, Manhattan, and Washington, DC. In each case, the detection model correctly classified sidewalks and crosswalks, creating a continuous network despite the heavy shadow concentration on sidewalks (a), shadow and vegetation obstructing sidewalks, and crosswalks (b), and vegetation obstructing curbs and crosswalks (c).

Future work could further examine ways to fill in missing gaps in the resulting networks using probabilistic techniques. For instance, if additional detection classes, such as "tree" or "shadow," are added to the semantic segmentation procedure, then these could be used in the network correction procedures to automatically connect gaps under trees and shadows. Yet, any automated correction for missing network links faces the hazard of erroneously creating pedestrian segments where they are not visible and hence may not exist. When networks are prepared for vulnerable street users (e.g., wheelchair users, mobility-impaired users, etc.), for whom network accuracy is critical, automated network correction procedures are likely futile, and improvements can only be made from ground surveys or

Google Street View images.

Moreover, in the future, we plan to add additional classes such as driveways, curbs, stairs, and separating public and private footpaths to our detection model. The model is presently limited to detecting only sidewalk and crosswalk elements, which may not be appropriate in cities, where considerable parts of the pedestrian infrastructure are invisible from aerial imagery–overground foot-bridges, under-ground pedestrian crossings, covered pathways, and public pathways inside buildings. Additional efforts will be needed to combine aerial sidewalk and crosswalk detection with invisible indoor elements in the contexts where the latter are significant (e.g., Hong Kong, Singapore, Minneapolis, and Montreal, to name a few).

The lack of standardized training data across different cities also posed challenges in our work. For instance, different cities have captured and mapped sidewalks with varying levels of detail. In Washington, DC, unpaved planter areas were excluded from sidewalk polygons, whereas in Boston and NYC, they were treated as parts of sidewalks. The same problem exists for curb extensions, medians, driveways, and curb-cuts. Moreover, the edges of the road and sidewalk polygons overlap and, in multiple instances, in GIS ground truth data. Crosswalk representation presented another source of variation among different cities. While they were mapped as part of sidewalk inventory data in Washington DC, in Boston, they were only presented in the sidewalk centerline dataset; hence, with no information available about their size and shape. In Cambridge, they were part of both the sidewalk centerline data and a separate dataset on road markings, where pedestrian zebras were represented as polygons.

Beyond heterogeneity in training data, the physical features, materials, and dimensions of sidewalks and crosswalks can also vary widely between cities. We observed multiple instances of faded crosswalks that made it challenging for semantic segmentation to detect. We also noted differences in both sidewalk materials and crosswalk materials across cities. Whereas very few crosswalks are paved in brick in NYC, they are common in Cambridge

and Boston. Had we trained the algorithm on NYC, it could have resulted in systemic underdetection in Boston and Cambridge. Such differences are bound to be much bigger between international cities, where construction materials, crosswalk marking conventions, and infrastructure dimensions vary more considerably than between the three East Coast cities included in our study. When extending the model to new contexts, especially outside the U.S., it is crucial to train the model specifically for each region.

3.6 Conclusion

In this paper, we presented TILE2NET, a solution that is able to create accurate pedestrian networks from aerial imagery in an end-to-end fashion. We pilot tested the approach in New York City, Washington, DC, Boston, and Cambridge, with varying street network geometries, building shadow densities, and tree covers and reported on the quality and accuracy of the approach. The resulting networks are created using the most recent orthorectified images, hence, more closely reflect the current urban form and pedestrian infrastructure. While the results are promising, we emphasize the need for expanding the work to additional cities and regions globally, where locally specific training may be needed to achieve high detection accuracy. However, the retraining for new regions can be done at much lower cost since our pre-trained model can be used for transfer-learning and domain adaptations with significantly less data compared to the initial training.

The resulting sidewalk and crosswalk dataset can be further combined with attribute information that may be useful for various pedestrian analytics. For instance, as shown by Hosseini et al. (2021), the captured sidewalk and crosswalk polygons can be used to measure the width of each sidewalk segment. Furthermore, using results by Hosseini, Miranda, et al. (2022), who developed a method for detecting sidewalk surface materials from Google Street View imagery, our sidewalk segments can be joined with corresponding geotagged material information, instead of having to aggregate the data from left and right sidewalks into road centerlines. Such measurable attributes can impact the quality and

attractiveness of sidewalks, and have been shown to affect pedestrian route choice and perceived route length (Basu et al., 2022; Erath et al., 2015; Sevtsuk et al., 2021).

Having pedestrian paths represented as continuous, topologically connected network datasets could open up new (and overdue) efforts for pedestrian routing, flow analysis, and potential location-based or delivery services. Transit-first policies, walkable-streets initiatives, step-free access for public transport, and vision zero goals represent but few planning and policy areas which could benefit from citywide sidewalk and crosswalk datasets.

CHAPTER 4

CITY SURFACES: CITY-SCALE SEMANTIC SEGMENTATION OF SIDEWALK MATERIALS

4.1 Introduction

As urban areas expand around the world, more impervious surfaces replace the natural landscape, creating significant ecological, hydrological, and economic disruptions (Arnold Jr & Gibbons, 1996; Chithra et al., 2015). Choosing the right material to cover city surfaces has become a critical issue in mitigating the adverse effects of increased anthropogenic activities. Historically, local availability, cost, strength, and aesthetics were the main factors influencing the choice of surface pavements (Lay et al., 2020; Tillson, 1900). The advent of asphalt and, later, concrete changed the face of cities. The longevity and durability coupled with relatively low production and installation costs made them the pavements of choice. However, as it was later revealed, these benefits came with huge environmental burdens (Van Dam et al., 2015).

One of the concerning environmental impacts of impervious surfaces is the sharp rise in urban temperature compared to its neighboring rural areas – a phenomenon called Ur-

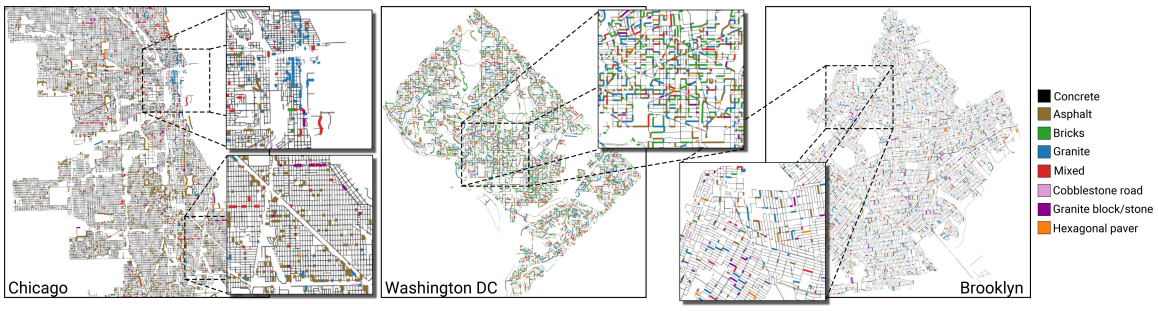

Figure 4.1: Using CitySurfaces to map the dominant surface material in Chicago, Washington DC, and Brooklyn (not part of our training data). Segments where the dominant material differs from concrete are drawn using a thicker line.

ban Heat Island (UHI) effect (Oke, 1982). UHI, which poses serious challenges to public health, ecological environment, and urban liveability (Estoque et al., 2017), is shown to be directly associated with surface characteristics, such as thermal performance and reflectivity. It can influence microclimates within the city by absorbing more diurnal heat and emitting that into the atmosphere at night (Nwakaire et al., 2020; Takebayashi & Moriyama, 2012; Wu et al., 2018). Natural surfaces and vegetation increase the amount of evapotranspiration and decrease the overall temperature and create a cool island effect (Amati & Taylor, 2010; Du et al., 2017). Reflective/high-albedo materials are also known to decrease UHI (Akbari et al., 2009; Santamouris, 2013; Santamouris et al., 2011; S. Zhu & Mai, 2019). Hence, the spatial distribution of land cover has a strong impact on the surface temperature (X. Chen & Zhang, 2017). Surface material also impact the water runoff and increase the risk of flooding. Sidewalks and roads form the main part of the urban ground surfaces. Today, the majority of the sidewalks are covered with impermeable materials which prohibit the infiltration of the water into the underlying soil, increase both the magnitude and frequency of surface runoffs (Bell et al., 2019; Shuster et al., 2005), reduce the groundwater recharge, and negatively impact the water quality. The excessive use of impervious surfaces is shown to be the primary cause of the Combined Sewer Overflows (CSOs), which can lead to massive pollution of natural bodies of water and street flooding (Joshi et al., 2021). Aside from the mentioned impacts, sidewalk pavements can also lead to public health hazards such as outdoor falls, or pose a barrier to walkability and accessibility of public spaces, specifically for the more vulnerable population and wheelchair users (Aghaabbasi et al., 2018; Clifton et al., 2007; Talbot et al., 2005). Studies show that uneven surfaces, indistinguishable surface colors, and low-friction materials contribute to the high incidence of outdoor falls in elderly populations (Chippendale & Boltz, 2015; Thomas et al., 2020a).

Despite the substantial economic, environmental, public health, and safety implications of sidewalk pavements (Estoque et al., 2017; Muench et al., 2010; Van Dam et al., 2015),

most cities, even in industrialized economies, still lack information about the location, condition, and paving materials of their sidewalks (Deitz et al., 2021). The lack of data creates barriers to understanding the real extent of the environmental and social impacts of using different materials and inhibits our ability to take a complex system approach to sustainability assessment (Van Dam et al., 2015). For instance, studies show a significant intra-urban variability of the urban thermal environment due to the street-level heterogeneity of paving materials (Agathangelidis et al., 2020). However, the data scarcity makes it challenging to measure this variability across different neighborhoods and consequently, impedes the development of a sustainable and resilient mitigation response plan (Akbari & Rose, 2008; X. Li et al., 2013; J. Yang et al., 2019). In the absence of fine-scale data, studies mainly rely on remote sensing images; however, the high-resolution aerial images are both spatially and temporally sparse (Y. Zhang et al., 2009), requiring researchers to use a variety of data aggregation and extrapolation techniques to fill in the missing data, which can lead to high bias and hurt the validity of the final results.

Collecting comprehensive and fine-scale sidewalk data using conventional methods is time-consuming and cost-prohibitive. Recent technological innovations in data collection opened new frontiers for research on public space and pedestrian facilities, creating opportunities to track features of interest at higher temporal frequencies and more granular geographic scales (Doraiswamy et al., 2018; E. L. Glaeser et al., 2018; Miranda, Hosseini, et al., 2020). The use of street-level images in urban analysis has gained popularity since the introduction of Google Street View (GSV) (Anguelov et al., 2010) and Microsoft Street Slide (Kopf et al., 2010), services that provide panoramic images captured by cameras mounted on a fleet of cars. Concurrently, developments in machine learning and computer vision applied to these new datasets have enabled novel research directions to measure the "unmeasurable" in urban built environments (Ewing & Handy, 2009), including sidewalks (Ai & Tsai, 2016; Frackelton et al., 2013).

In this work, we address this data gap and take a step towards exploring the surface of

our cities through CitySurfaces, a framework aimed at generating city-wide pavement material information by leveraging a collection of urban datasets. We combine active learning and computer vision-based segmentation model to locate, delineate, and classify sidewalk paving materials from street-level images. Our framework adopts a recent high-performing segmentation model (Tao et al., 2020), which uses hierarchical multi-scale attention combined with object-contextual representations. To tackle the challenges of high annotation costs associated with dense semantic label annotation, we make use of an iterative, multi-stage active learning approach, together with a previously acquired sidewalk inventory from Boston, which lists the dominant paving material for a given street segment. We demonstrate how the trained segmentation model can be extended with additional classes of materials with noticeably less effort, making it a versatile approach that can be used in cities with varying urban fabrics and paving materials. To show the generalization capabilities of CitySurfaces, we employ our framework in the segmentation of street-level images from four different cities: Brooklyn, Chicago, Washington DC, and Philadelphia, none of which were included in the training process. Figure 4.1 highlights how different pavement materials are spatially distributed in three cities.

Our contributions can be summarized as follows:

- We present CitySurfaces, a deep-learning-based image segmentation framework for large-scale localization and classification of sidewalk paving materials.
- We adopt an active learning strategy to significantly reduce pixel-level annotation costs for training data generation, and yield increased segmentation accuracy.
- We conduct extensive experiments using street-level images from six different cities demonstrating that our model can be applied to cities with distinct urban fabrics, even outside of the domain of the training data.
- We make publicly available our model as well as the results of our material classification in the six selected cities [1].

[1] https://github.com/VIDA-NYU/city-surfaces

Figure 4.2: The eight classes of surface materials used in our study. Top: standard and prevalent materials, Bottom: materials with distinct use.

This paper is organized as follows: section 4.2 describes the main data sources of our framework; section 4.3 describes the CitySurfaces framework; section 4.4 summarizes our results; section 4.6 highlights challenges and limitations; and section 4.7 presents our conclusion.

4.2 Data Description

Aware of the fact that manually labeling images is a time-consuming task, our proposed framework leverages a unique dataset that describes the material of sidewalks in Boston. We combine that data with the street-level images to create the training data for our semantic segmentation model. Next, we describe both data sources.

4.2.1 Boston sidewalk inventory

The sidewalk inventory (Boston PWD, 2014) is part of the Boston Pedestrian Transportation Plan (Loutzenheiser, Felix, 2010) and describes sidewalk features, including geographic coordinates and paving materials collected via manual field visits. The material

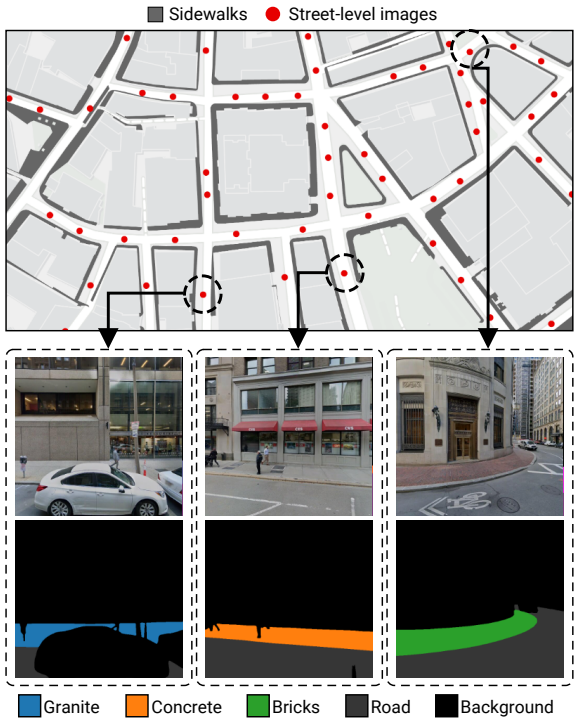

Figure 4.3: Examples of sampled points in Boston to obtain street-level images. Three different sampling locations are highlighted and for each location, the street-level image as well as the prediction result of the model is depicted.

attribute describes the dominant surface material of each street segment (either concrete, brick, granite, a mix of concrete and brick, or asphalt). Figure 4.2 illustrates patches of these five materials; the other three extra materials (granite block, cobblestone, hexagonal pavers) shown in the image were not recorded in the Boston dataset but were later manually added to our classes, as we will discuss in subsection 4.3.3. We grouped the street segments by materials, using the geographic coordinates of the paving materials in the Boston inventory, and used it to assign an overall image class to the street-level images to guide the annotation process.

4.2.2 Street-level images

Street-level image usage in urban analysis has gained popularity with the introduction of Google Street View (GSV) (Anguelov et al., 2010) and Microsoft Street Slide (Kopf et al., 2010), services that provide panoramic images captured by specifically designed cameras

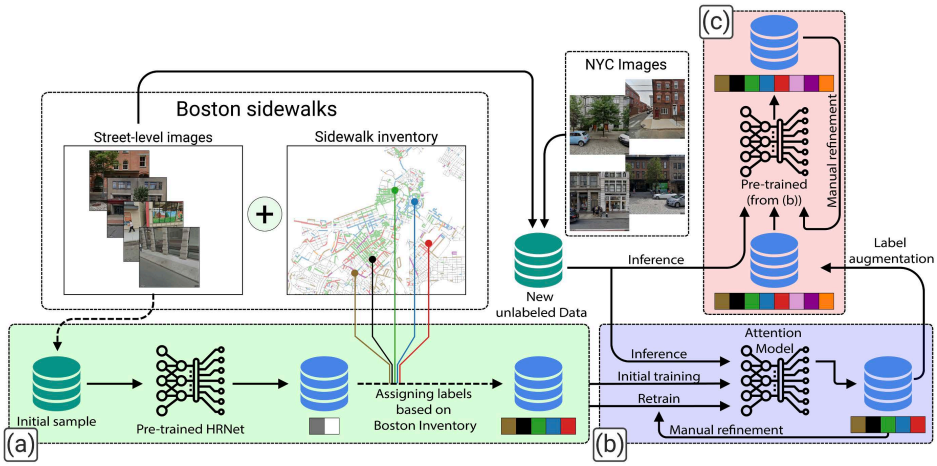

Figure 4.4: CitySurfaces workflow. Block (a): Creating the initial ground truth labels using the Boston sidewalk inventory and GSV images. A sample of unlabeled images is fed to a pre-trained HRNet, which outputs annotation labels containing two classes of interest: roads and sidewalks. The labels are manually refined to represent the five sidewalk paving classes, forming our ground truth set; Block (b): Training the base model to classify five classes of surface materials, plus roads. The data from block (a) is used for the first stage of training. The model is then iteratively retrained for multiple stages on new samples. In each stage, the most representative and informative samples are chosen, and the annotations are manually refined and added to the training set to retrain the network; Block (c): Introducing three new classes of materials. The pre-trained model from block (b) is retrained on the newly annotated image with three new classes. The final model can classify eight classes of different materials.

mounted on a fleet of vehicles. These new data sources enable new questions and study designs for urban planning and design, urban sociology, and public health (Griew et al., 2013; Mooney et al., 2016; Yin et al., 2015). The GSV API can retrieve street-level images via geographic coordinates and allows users to adjust camera settings such as the heading, field of view (FoV), and pitch.

We use the OSMnx library (Boeing, 2017) to obtain the Boston street network and query the GSV API for street-level images at a fixed interval (5 meters), excluding major highways and tunnels. We acquire the compass bearing of each street to set the camera heading to be perpendicular to the street, thus looking directly at left and right sidewalks. The pitch was set to $0°$ with an FoV of $80°$. To create a more diverse training set, for 35% of the training data, we use different combinations of headings (pitch $\in [-10°, -20°]$,

and FoV$\in [60°, 70°]$), to have sidewalk images taken at varying angles and perspectives. Figure 4.3 illustrates sampled street segments in Boston, together with their image-level annotations. In order to train our framework, 3,500 Boston images were obtained, and later 2,000 images from New York City (NYC) were added to the pool of initially unannotated data. We excluded images with no sidewalks as well as those where more than 80% of the sidewalks were occluded. The final set had a total of 4,300 images.

4.3 CitySurfaces

CitySurfaces adopts an active learning approach for the semantic segmentation of sidewalk paving materials. Using this framework, we aim to: 1) Train a model that can classify five different paving materials plus asphalt roads; 2) Extract information about sidewalk materials of a city for which no ground truth sidewalk inventory exists (e.g., NYC); and 3) Extend the model to classify additional classes of materials so that it can be applied to a more general set of cities.

Active learning aims at achieving high accuracy while minimizing the amount of required labeled data. The main hypothesis is, if we allow the model to choose the training data, it will perform better with fewer labeled instances (Settles, 2009). Through iteratively selecting the most informative or representative images to be labeled, fewer labeled instances are required to achieve similar performance compared to randomly selecting a large sample as training data and annotating all of it at once (Bloodgood & Vijay-Shanker, 2014; S.-J. Huang et al., 2010).

In general, our multi-stage workflow is different from previous works in active learning for semantic segmentation in two important ways: First, our sample selection method is not fully automated; we use the uncertainty measure to filter the pool of unlabeled data in each stage, but we also use domain expertise for selecting a sample of images to be annotated and added to the training set in the next stage. Second, our query frequency is ten epochs (each epoch is a pass through all training data). The conventional approach in active learning is

to select new samples (query) every iteration, which can work in cases where the cost of annotation is not high or in experimental studies that work with already annotated images to advance the field and develop new query algorithms, as is the case with most of the already published works in active learning for semantic segmentation, where they use datasets such as Cityscapes (Cordts et al., 2016) or ADE20k (B. Zhou et al., 2017). However, since no annotated dataset exists for sidewalk materials, we have to annotate every new sample we choose during the training process, and it is impractical to annotate a new sample for every iteration (T. Kim et al., 2020). To overcome this, we adopt a multi-stage framework and annotate a new sample at the end of each stage, where each stage consists of ten epochs.

Our workflow has three major blocks as illustrated in Figure 4.4: Block (a) creating initial training labels; Block (b) training a material segmentation model and; Block (c) extending the model to segment three additional classes from NYC standard materials. In this section, we first describe the different blocks of the workflow in detail, followed by a description of the semantic segmentation model. The training process and experiments were executed on 4 NVIDIA P100 GPUs with 12 GB of RAM each.

4.3.1 Block (a): Initial image annotation

To start the training process, we need a set of annotated images. To obtain the annotated data, we randomly sample 1,000 images from a pool of unlabeled Boston street-level images and feed that sample into HRNet-W48 (K. Sun, Zhao, et al., 2019; J. Wang et al., 2020) model pre-trained on Cityscapes (Cordts et al., 2016) and get the initial segmentation results (Figure 4.4(a)). The model outputs 19 classes from which we only keep roads and sidewalks. To generate an initial set of labeled data, we make use of the Boston Sidewalk Inventory (detailed in subsection 4.2.1). We first query for the street segments of the images in our initial sample and modify the label to match the audited pavement from the inventory. Effectively, we are ensuring that, instead of having a general *sidewalk* class outputted by the pre-trained HRNet, our image set will have annotations according to the

ground truth inventory data (e.g., concrete, bricks). We then manually refine them to account for the pre-trained model's prediction errors. In the initial training set, we restrict our sampling to images where the sidewalks mainly consist of a single material and eventually move to more complex material configurations in later stages. The final annotated images were split into 80% training and 20% validation to train the model in block (b).

4.3.2 Block (b): Model training on Boston and NYC

In the second block of the framework (Figure 4.4(b)), we train an attention-based model (detailed in subsection 4.3.4) using the labeled images from block (a). Our training step initially uses 800 images for training, and 200 images for validation, with a batch size of 8, SGD for the optimizer, momentum 0.9, weight decay $5e^{-4}$, and an initial learning rate of 0.002. We train the model in a multi-stage framework, where each stage consists of ten epochs. In each stage, we choose the epoch with the highest mIoU on the validation set. At the end of each stage, we make two decisions: 1) we select the best model considering all epochs of the current stage; and 2) we analyze the quantitative and qualitative results of the model to guide sampling the *new* addition to the training data. In particular, we analyze the confusion matrix, similarity matrix, as well as the top 10% of predictions with the highest mIoU and the top 20% of failures, obtained from the validation phase of the best epoch. The weights of the best model in the current stage are then used to initialize the model in the next stage with more training data. This restating scheme of SGD with the best solution of the previous stage is useful in increasing the chances of finding better solutions in the current stage.

To sample new images, we employ two strategies: i) Uncertainty in predicting unlabeled images: We make use of the model's uncertainty estimations on unlabeled data and select the images that were most challenging for the model to predict; and ii) Performance on validation set: By examining per-image IoU, uncertainty, and error rates of the images from failure and success cases together with confusion matrices, we construct a set of sam-

ple images to be used as inputs for finding similar unlabeled images. A more detailed explanation of these two techniques is provided in Appendix A.

Following the sample selection strategies, we retrieve 300 unlabeled images, apply the current model on these new unlabeled images to generate a prediction, and then modify the predicted labels to add them to the overall training set, such that the segmentation model is trained on more samples of hard-to-segment images. To improve model generalization, in the third stage, we begin including images from Manhattan, which has a different urban fabric and more diverse forms and types of paving materials, in the pool of unlabeled data. Since no ground truth data exists for Manhattan, to create the ground truth label, we need to have a model with reliable performance to create the base annotation. We chose the third stage since the model reached a reliable performance (83% mIoU) in detecting the main classes, and outputs had clear borders compared to the other two stages. The selected images from Manhattan were fed to the model, and the results were corrected

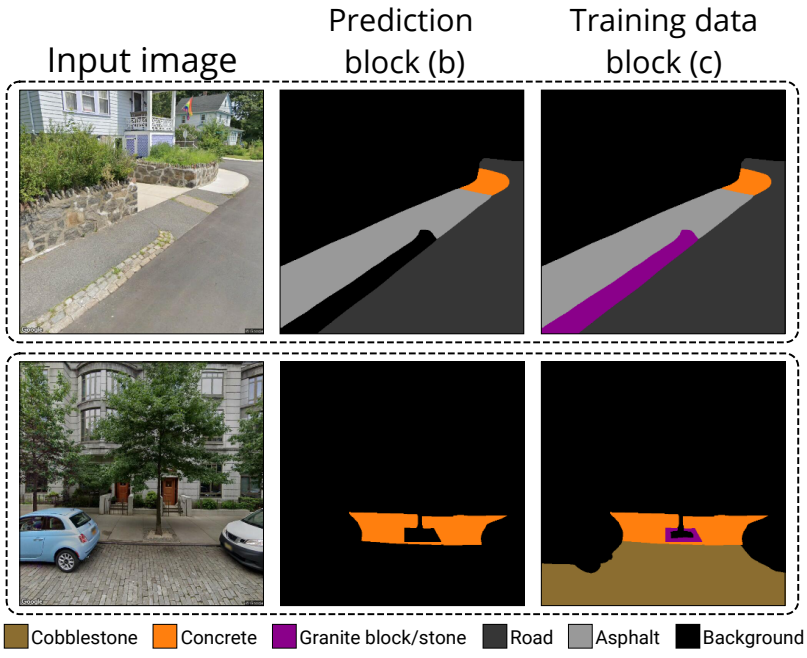

Figure 4.5: Examples of how the annotation labels with additional classes were created from the output of the model in block (b) of our framework. The model trained in block (b) classified granite blocks and cobblestone as background, leaving smooth and clear boundaries, which helps to augment the labels with new classes during manual refinement and train a model that can classify eight different materials (block (c) of the framework).

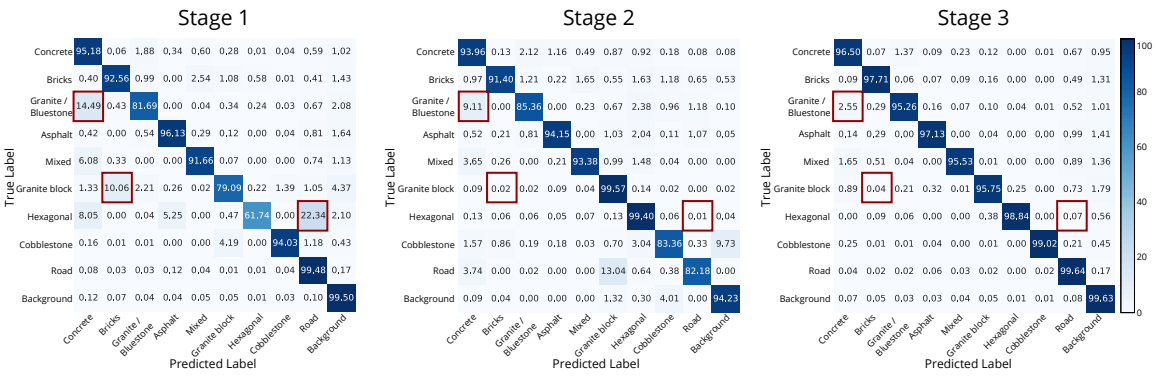

Figure 4.6: Confusion matrices for the three stages of the extended model. These results guided sample selection and signaled which type of images should be included in the training data for the next stage. Notice the improvement of the predictions for hexagonal pavers, granite block, and granite/bluestone (highlighted in red).

and refined using feedback from the domain expert and added to the training dataset. The segmentation model is then trained on the combined set of the initial and newly annotated data (1,100 images), initialized with the weight from the best epoch of the previous stage. This procedure is iterated for five stages (at which point we observe no further notable improvements). The model at the final stage was trained on 2,500 images (Figure 4.4(b)), and achieved 88.6 % mIoU on the held-out test set.

4.3.3 Block (c): Including additional materials from NYC

Once the model in block (b) attains sufficiently accurate segmentation performance, we extend it by adding three additional classes (Figure 4.4(c)). The three additional classes are granite blocks, hexagonal pavers, and cobblestone. These materials are standard sidewalk materials in the NYC street design manual (NYC DOT, 2020). While granite blocks and cobblestones were also observed in Boston, they were not included in the Boston sidewalk inventory. Since the original model in block (b) was not trained to detect these materials, they are initially either classified as background (mostly granite blocks and cobblestones) or misclassified (mostly hexagonal pavers) as other visually similar materials. To collect street-view images that have these new materials, we follow the NYC and Boston street design manuals (NYC DOT, 2020; Thomas M. Menino, 2013) to filter unlabeled data from

the locations in which these materials can be found. For example, hexagonal pavers (NYC only) are typically used on sidewalks adjacent to parks and open spaces, and cobblestones are used in historic districts.

We select a total of 800 images that contain these new classes to be iteratively sampled for training, 150 additional images for the validation set, and 200 images for the held-out test set. Annotating the new image set consumed fewer resources as compared to the initial annotations since smooth model predictions typically leave clear boundaries, which only needed to be assigned the appropriate label (see Figure 4.5). The newly generated set of labels was used to train the model by initializing the architecture with model weights in block (b) and only replacing the final softmax layer instead, to produce ten output channels (corresponding to eight paving materials, plus the road, and background). At the end of each stage, we select a new sample of unlabeled images following the same process explained in subsection 4.3.2, run them through the model, obtain segmentation predictions, refine the results, and retrain the model. In total, 726 additional images were added to the training set, and in the final stage, the model was trained on 3,226 images (2,500 from block (b) + 726). We halt the training in stage 3 after 30 epochs, and test the model on the held-out test set (Figure 4.4(c)). Figure 4.6 shows the confusion matrices for all three stages of our extended model, illustrating model performance as a function of the amount of training data. These matrices were also used in part to guide the sampling of images to annotate.

Using the described method, model performance increases from 74.3% mIoU to 88.6% for the base model (block (b)) and to 90.5% in the extended model (block (c)), with the manual refinement time decreasing from 25 to 4 minutes per image. Figure 4.7 depicts the evolution of the segmentation results of block (c) through the active learning stages. The model outputs more refined boundaries and significantly less noise in later stages; thus, significantly less time is needed to modify the newly annotated data as the stages go on. In each stage, the model is initialized with the weights from the previous stage.

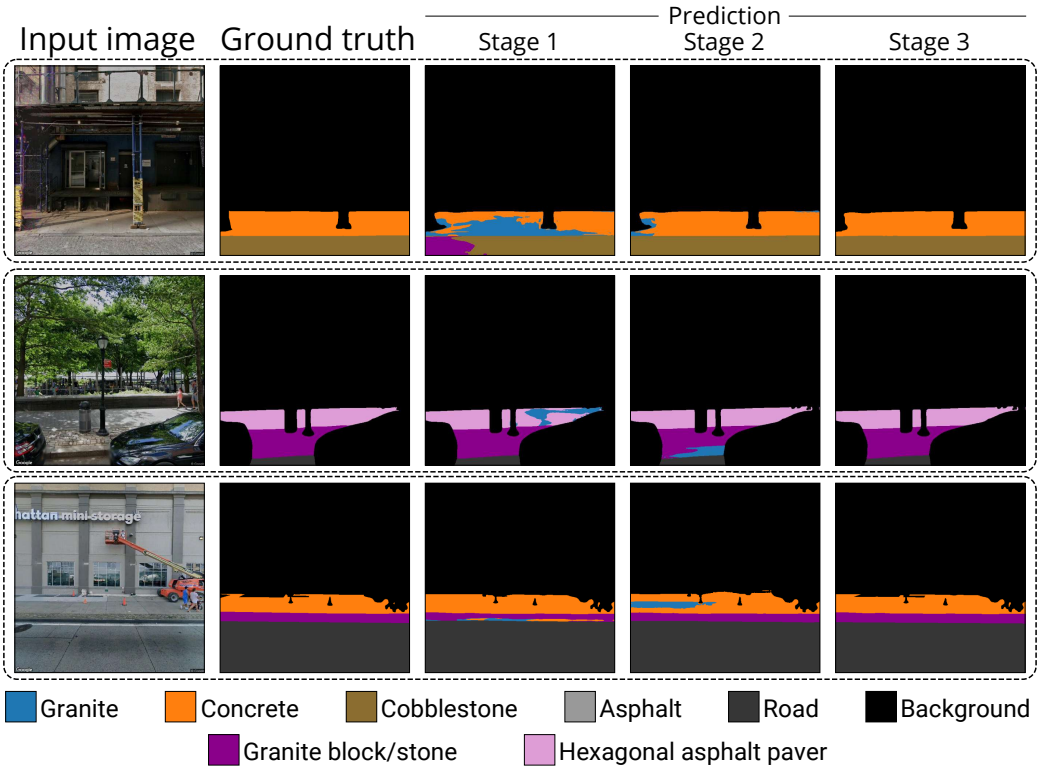

Figure 4.7: Evolution of the block (c) extended model's inference through different training stages.

4.3.4 Semantic segmentation model

For the semantic segmentation task (blocks (b) and (c)), we adopt the Hierarchical Multi-Scale Attention (Tao et al., 2020) and fine-tune the parameters on our dataset. To train the model, following Y. Zhu et al. (2019), we employ class uniform sampling in the data loader, which chooses equal samples for each class for handling the class imbalance, since some classes like road and background are almost present in all images, whereas classes like cobblestone and hexagonal pavers are not that prevalent. The Region Mutual Information (RMI) loss (S. Zhao et al., 2019) was employed as the primary loss function. RMI takes the relationship between pixels into account and uses the neighboring pixels around each pixel to represent it instead of only relying on single pixels to calculate the loss. We run different experiments with and without the RMI loss function for the main segmentation head. In the absence of RMI, standard cross-entropy loss was used instead. The model under the same

setting, but without RMI loss, performed slightly worse (89.84) compared to the one where RMI loss was used (90.51). Figure 4.8 presents an overview of the architecture. Next, we describe the network's architecture in more detail.

Backbone

We chose HRNet-OCR (Yuan et al., 2019) as the backbone. The network comprises HRNet-W48 (K. Sun, Zhao, et al., 2019; J. Wang et al., 2020) and adds Object-Contextual Representations (Yuan et al., 2019) to further augment the representation extracted by the HRNet. The final representation from HRNet-W48 works as the input to the OCR module, which computes the weighted aggregation of all the object region representations to augment the representation of each pixel. The weights are calculated based on the relations between pixels and object regions. The augmented representations are the input for the attention model described next.

Attention model

The model is mainly based on Share-Net (L.-C. Chen et al., 2016). Suppose an input image is resized to several scales, i.e., $s \in \{1, ..., S\}$. Each scale is passed through the back-

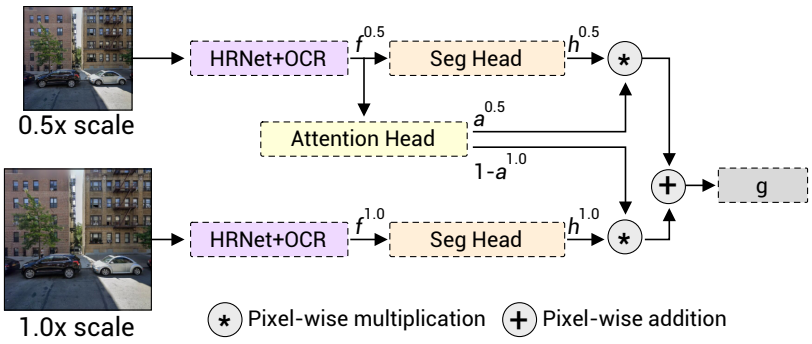

Figure 4.8: The general architecture of the hierarchical multi-scale attention (HMSA) based semantic segmentation method (Tao et al., 2020). The inputs are images from two scales. The network learns the relative attention between scales and hierarchically applies the learned attention to combine the results from two segmentation heads and make a prediction.

bone part (HRNet-W48+OCR), and we can get the output feature $f^s_{i,c}$. For the feature, $c \in \{1, ..., C\}$ (C is the number of classes of interest, and i ranges over all the spatial positions). As shown in Figure 4.8, the features then go through two heads, one for attention generation and the other for segmentation. The features $f^s_{i,c}$ are resized for different scales to have the same resolution (with respect to the finest scale) using bilinear interpolation before passing the model heads. For the attention head, we generate the learned weights for $f^s_{i,c}$ which is represented by $a^s_{i,c}$. This weight is integrated into the initial output $h^s_{i,c}$ from the segmentation head, and we have:

$$g^s_{i,c} = a^s_{i,c} * h^s_{i,c} \tag{4.1}$$

in which $g^s_{i,c}$ is the final output score map for scale s, and $*$ here represents the pixel-wise multiplication.

In the model, the combination of score maps is similar to (Tao et al., 2020) to make the flexible scales during inference time possible and improve the training efficiency. During the training, we only need to train with two adjacent scales (as shown in Figure 4.8). During testing, weights for the network are shared for each adjacent scale pair.

To be more specific, suppose the two selected adjacent scales are $1x$ and $0.5x$ (the final selected scales during training in the model are $0.5x$, $1x$, and $2x$) to obtain the pair of scaled images for the model input. For inference, we can hierarchically and repeatedly use the learned attention to combine N scales of predictions together. Precedence is given to lower scales since they have a more global context and can choose where predictions need to be refined by higher scale predictions. The final combination principle for these adjacent scales is defined as:

$$g_{i,c} = a^0_{i,c}.5 * h^0_{i,c}.5 + (1 - a^0_{i,c}.5) * h^1_{i,c} \tag{4.2}$$

The hierarchical mechanism used in the model coupled with the powerful HRNet-OCR backbone provides a robust architecture for the challenging task of material classification

Figure 4.9: Predictions of the model on the held-out test set. Fine details and boundaries of objects like poles, plants, wooden sticks, and fire hydrants are very precisely predicted. The model also segmented curb cuts (line 1 - column 2), different instances of the same material (3-1), (3-3), and visually similar materials of different classes (1-4).

in the wild.

4.4 Results

In this section, we present the results of applying our trained model on the held-out test set. We do not rely on pixel-level accuracy in evaluating the model since sidewalks comprise a relatively small portion of each image, while road and background can occupy more than 70% of most images, resulting in a significant class imbalance. This class imbalance creates an arbitrary high pixel-level accuracy, which is not a fair representation of the model's performance.

4.4.1 General evaluation metrics

Table 4.1 presents class-level evaluation metrics, the mean Jaccard index (IoU), precision, and recall for the final model. The model outputs ten classes in total, seven classes of sidewalk pavings, one extra class of street pavings - cobblestone - plus road and background. Excluding road and background, the model achieved 88.37% accuracy, with hexagonal asphalt pavers and asphalt sidewalks having the highest accuracy measures. Overall, half of the pavement classes have IoU above 90%. Concrete, the most prevalent and versatile

Table 4.1: Evaluation metrics on the held-out test set.

Label	IoU	Precision	Recall
Concrete	88.69	0.95	0.93
Brick	91.79	0.95	0.96
Granite/Bluestone	81.09	0.85	0.95
Asphalt	92.58	0.96	0.97
Mixed	86.11	0.93	0.93
Granite Block/Stone	82.92	0.94	0.88
Hexagonal Asphalt Paver	92.81	0.98	0.95
Cobblestone	90.95	0.94	0.96
Road	99.01	0.99	1
Background	99.16	1	1
mIoU	90.51		
mIoU (eight main classes)	88.37		

material, can be classified with 88.7 accuracy. A robust result considering the high within-class variation (i.e., it comes in various colors and textures). Granite/bluestone and granite block have the lowest accuracy (81.09 and 82.92 respectively). This can be partially explained by their visual similarity to dark concrete (or wet concrete), potentially leading to more false positive predictions.

Figure 4.9 illustrates some examples of the model's prediction, highlighting its performance in detecting boundaries between fine objects, like poles and plants, even in shadowed scenes (line 1 - column 1, 1-3, 2-1). The model can also detect curb ramps in most scenes, even though it was not specifically trained with such a goal (1-1 and 2-2). Figure 4.9 (1-2) shows an example in which the model accurately classified a sidewalk segment with patches of different materials. We can also see the model performance in distinguishing between visually similar materials (1-4, 3-2), as well as different variation of the same material such as (3-1) where two visually distinct concrete slabs are classified correctly.

4.4.2 Evaluating the generalization capabilities of CitySurfaces

To demonstrate the generalization capabilities of CitySurfaces, we tested the performance of our approach on samples from Chicago, Washington DC, Philadelphia, and Brooklyn

Table 4.2: Evaluation metrics on samples from the selected cities (outside of training domain).

City	mIoU	Mean Per-Segment Accuracy
Brooklyn	86.12	87.09
Chicago	84.31	86.52
Washington DC	82.61	84.27
Philadelphia	82.81	83.46

(NYC borough), which were not part of the training data. We randomly sampled 200 street segments from each city, and obtained their corresponding street-view images, at every five meters of each segment, from the left and right sides of the sidewalks. After data cleaning and pre-processing, we were left with roughly 600 images per city; these images were annotated using the model in block (b), then *manually* checked and refined to create the test set. Table Table 4.2 shows the results of applying CitySurfaces on these test sets. We report mIoU and mean per-segment accuracy. Mean per-segment is a simple and practical metric that measures whether the model correctly detected the dominant materials in each street segment and report the average accuracy over all images in the test set. All tested cities had an accuracy greater than 82%. Brooklyn achieved the highest accuracy, since the borough's paving materials follow the same street design regulation as Manhattan, which was part of the training data.

CitySurfaces enables generating city-wide sidewalk material datasets, as illustrated in Figure 4.9. This allows us to compare the distribution of different paving materials in various cities. Figure 4.10 shows the result of this comparison. We can see that Manhattan and Washington DC use more diverse and balanced material types. Concrete is the dominant material in all of the cities. Chicago has the highest number of asphalt sidewalks among the selected cities; Boston, Washington DC, and Philadelphia have a similar number of asphalt sidewalks, which come second to Chicago. Asphalt sidewalks are mainly used in suburban neighborhoods; that is why dense urban areas like Manhattan and Brooklyn have the lowest number of sidewalks paved with asphalt. Another interesting observation is the higher usage of granite/bluestone in Manhattan compared to Brooklyn, two boroughs of

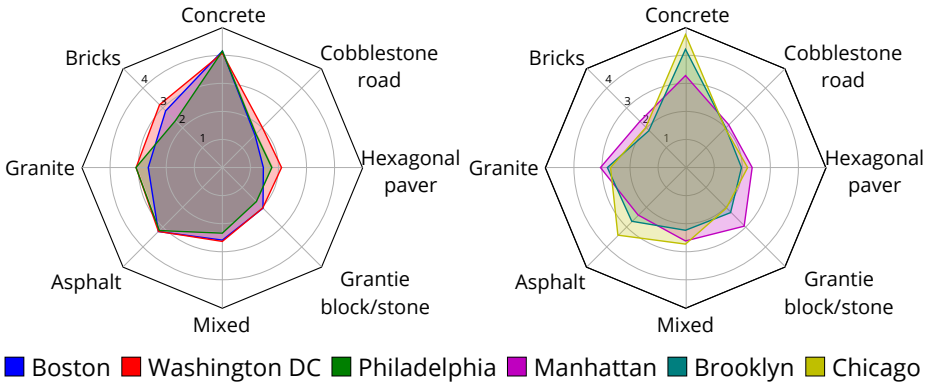

Figure 4.10: Comparison of the distribution of detected materials in six different cities. The star plots show the log of the number of sidewalk segments identified as having a given material.

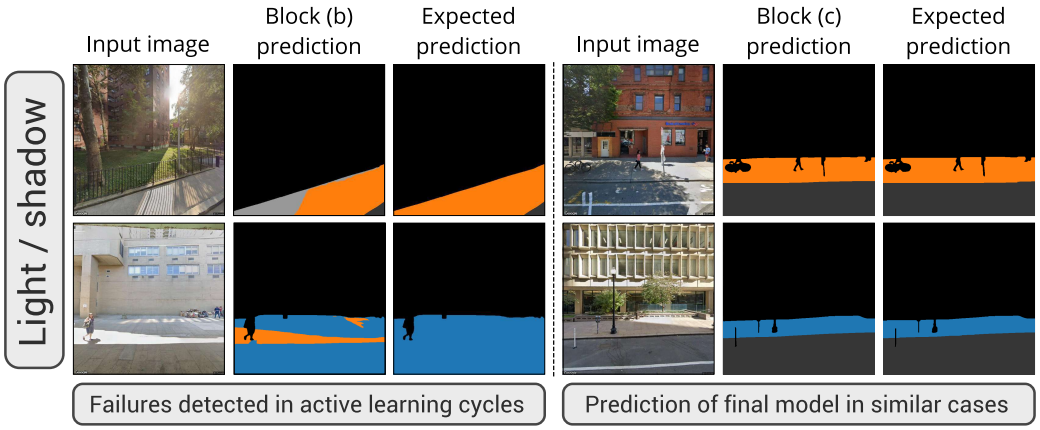

Figure 4.11: **Left**: Exposure to direct sunlight changed the appearance of colors and texture of the paving material, **Left top**: Part of a concrete sidewalk under the shadow was classified as asphalt. **Left bottom**: Part of a granite surface under direct sunlight was classified as concrete. **Right**: The correct predictions of the final model in the same settings.

the same city. Granite is considered an expensive and decorative material, used mainly in commercial streets or historic neighborhoods, which signals Manhattan's higher land value and income level, since maintenance and installation of decorative pavings are the owner's responsibility.

4.5 CitySurfaces Use Case: Plan a Safe Stroll in Downtown

Sidewalk surface is among the most important factors in determining the risk of outdoor falls (S. Lee, 2018; Twardzik et al., 2019). The type of surface material has a significant

impact on the walking experience (Ferreira & da Penha Sanches, 2007), in particular, for older adults and people using mobility aids (Chippendale, 2020). In cities with historic fabric, such as Boston or Philadelphia, fired bricks are the iconic surviving historic paving for sidewalks, aging back to 1795 (Archipedia New England, 2019), hence, considered an asset for many cities to preserve. Since bricks were often installed on a bed of stone dust and sand to allow surface water to pass through, the surface became uneven over time (Loutzenheiser, Felix, 2010) which can lead to major tripping hazards. In this section, we highlight how CitySurfaces can be helpful to domain experts engaged in route planning and fall prevention programs for seniors.

Slippery surfaces pose a major challenge to pedestrians navigating the outdoor environment in cold and snowy weather (Chippendale & Boltz, 2015). In the absence of sunlight, and when the temperature suddenly drops, transparent and slippery form of thick coating ice, known as *black ice*, can form on top of pavements. In general, the average temperature, the relative surface temperature and the type of surface material are key contributing factors to black ice formation (Aljuboori, 2014; Houle, 2008; Monroy Licht, 2015). Moreover, the form, bulk, and height of different surrounding structures determine the amount of light that can crawl in from between the tightly standing buildings (Miranda et al., 2019). Sidewalks located in places that get very little sunlight or much shade are usually lower in temperature. According to the study by Druschel (2020), the shadow under bridges caused a 28°F temperature difference in the asphalt pavement, which highlights the impact of sunlight obstruction on surface temperature, ultimately increasing the risk of black ice formation. Aside from the shadow and temperature, impermeable pavements, such as granite, are more susceptible to ice formation and getting slippery since they cannot let the water pass through.

We collaborated with an occupational therapist who runs fall prevention programs for older adults. One of the key practices of the fall prevention program is route planning, in which they assess the condition of different routes as well as various risk factors such

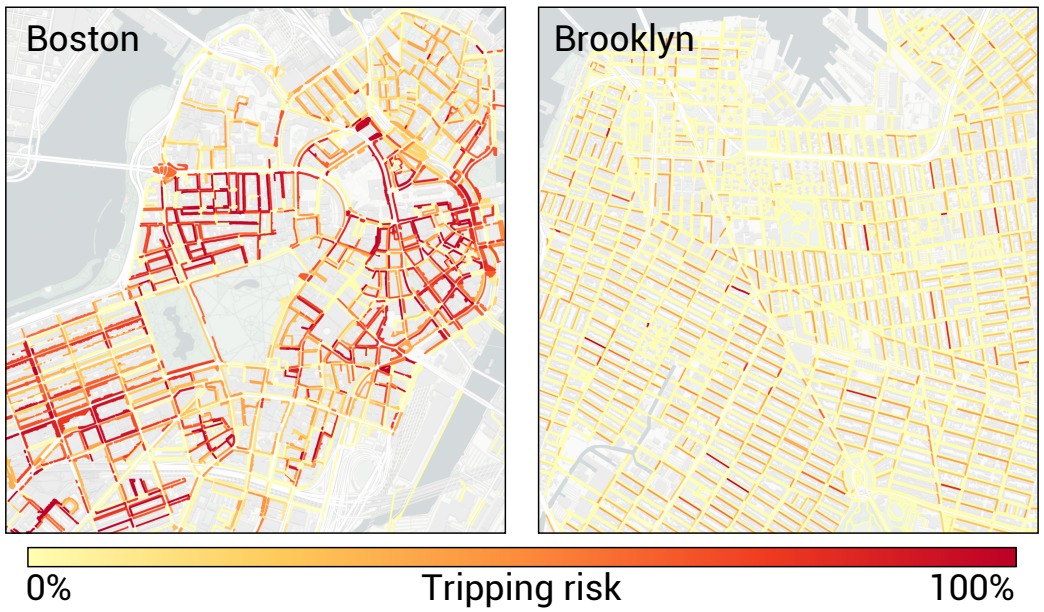

Figure 4.12: The risk of tripping based on the percentage of brick and granite and the accumulated shadow for each street segment.

as paving materials, street enclosure, lack of sunlight, and obstacles on the way to inform their participants about the condition of the walking environment along the route and equip them with strategies to prepare for each condition before the strolling starts (Chippendale, 2020). But the lack of comprehensive data describing the conditions of pedestrian facilities at a fine, human-scale level significantly limits the geographic coverage of their practice, specifically when the locations are distant or unfamiliar.

Motivated by our ongoing collaboration, we use the data generated by CitySurfaces, together with shadow accumulation data (Miranda et al., 2019, 2020) aggregated on sidewalk segment level to calculate a simple tripping risk measure for identifying sidewalks that can pose a higher risk of falls in cold seasons. The measure is calculated using three factors that our collaborator was interested in: 1) the percentage of bricks; 2) the percentage of granite; and 3) the accumulated shadow data for a day in December (winter solstice) for each sidewalk segment. Other contributing factors such as slope and width of the sidewalks can be taken into consideration, but in this use case, our goal was to highlight the role of surface materials; hence, we did not incorporate them. For simplicity, we gave all three

factors the same weight, but the weights can be adjusted based on the preferences as well as regional / city-scale fall risk assessment models, which is outside of the scope of this work.

For this study, we chose the downtown and North End neighborhoods of Boston, as well as downtown and Prospect Park areas in Brooklyn, for their historic fabric and the fact that these neighborhoods are considered popular destinations. Figure 4.12 shows our tripping risk map. The highest risk is posed by most shadowed streets with the highest cumulative percentage of bricks and granite; these are places where the risk of black ice formation is considerably higher. The heat map ranks the streets based on their safety level in cold and snowy weather. Moreover, the map informs pedestrians of the type of surface they would face in each street. As can be seen in Figure 4.12, downtown Boston has a higher concentration of streets with high tripping risk compared to downtown Brooklyn. The historic red brick sidewalks in Boston, coupled with less winter sunlight, creates a riskier environment for walking. In Brooklyn, downtown has a higher concentration of granite and bluestone, but the wider street and sidewalk design, specifically around the newly redeveloped areas of upper downtown (top center of the map), provide more sky exposure, more sunlight, and lower risk of ice formation and slippery surfaces.

Using the information about the percentage of different materials in each street segment and the amount of shadow accumulation, we can inform pedestrians about the condition of the walking environment they are planning to visit. An important and distinguishing aspect of the dataset generated through this work is how it goes beyond the dominant materials on each segment and provides the percentage of different surface materials used in a given sidewalk, which proves critical for cases related to safety and health since even a tiny patch of slippery surface can lead to tripping hazards.

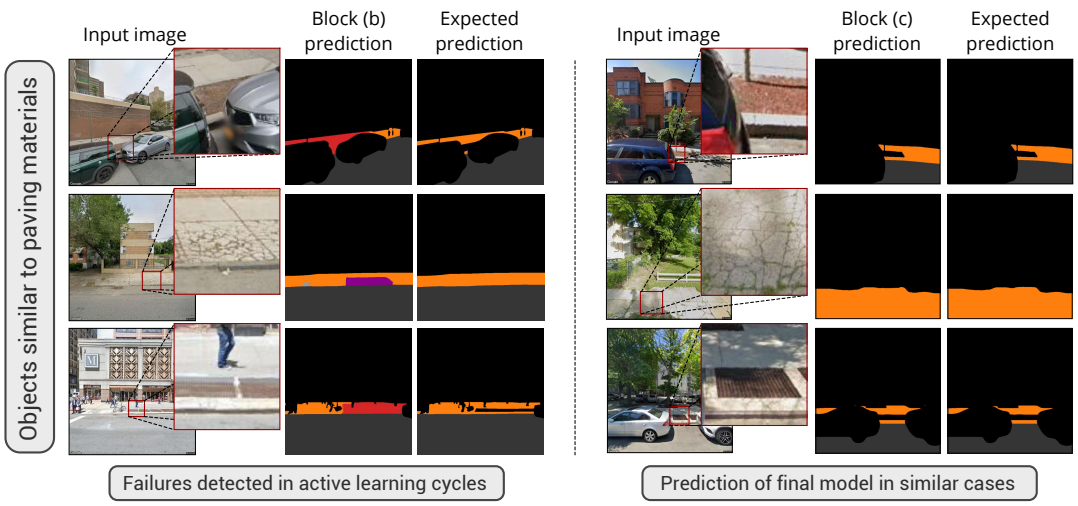

Figure 4.13: Objects with patterns similar to different materials. **Left**: Classifying failures caused by different patterns. **Left top**: Concrete alongside a furnishing zone was misclassified as mixed class since plant pit was detected as bricks, **Left middle**: Broken concretes were misclassified as granite blocks, **Left bottom**: Concrete was misclassified as mixed class due to the presence of brownish metal covers. **Right**: Correct prediction of the model for the similar pattern in the final cycle of active learning.

4.6 Discussion

The specific characteristics of computing the spatial distribution of sidewalk pavement materials require experts to oversee the performance of the model and ensure that the network is correctly classifying the pavement materials. Through active learning process, we identified certain elements of the urban scenes that can create higher prediction confusion and lead to misclassification. Two main categories of patterns repeatedly observed among the failure cases were shadow/light contrasts (Figure 4.11) and distinct objects such as metal gratings and plant pits that resemble brick from a distance (Figure 4.13). The texture and color of different materials can appear different under shadow or extreme light, showing a higher resemblance to another material. For instance, under the shadow, concrete is classified as asphalt (Figure 4.11 - left top). Moreover, some patterns or objects can look similar to certain materials. For example, the model initially classified certain plant pits (Figure 4.13 - left top) or brownish metal covers (Figure 4.13 - left bottom) as bricks alongside the concrete pavement and would incorrectly predict mixed materials for that part of the

sidewalk, or even small pieces of broken concrete or granite were classified as cobblestones (Figure 4.13 - left middle). Adding more images with these patterns to the training data improved the model's performance in the next stage. Some examples of the correct predictions for similar patterns are shown on the right side of Figure 4.13. The active learning strategy significantly helped with choosing the right data at each stage. Having an expert in the loop to review the results in each stage enabled identifying specific patterns that were not evident by merely analyzing the quantitative metrics of the model.

4.6.1 Challenges

One of the key challenges of this study was handling different textures of the same object (sidewalk). Objects have defined boundaries that are easier to classify (Jain & Gruteser, 2018). However, similar textures can appear on multiple objects. For instance, red bricks are used in both building facades and sidewalk pavings (although different types of bricks are used for each purpose, they possess very close visual characteristics). Our goal is to have a model that can detect *sidewalks* of certain materials from street-view images.

Another challenging aspect of this task is the high degree of within-class variation and between-class similarities. For instance, NYC designated five different types of concrete as standard materials for sidewalk pavings, while Boston uses three different types of concrete. Each of these types has distinct visual features that, in some cases, can resemble materials of other classes, which pose further challenges to the classification task. Distinguishing between dark concrete and bluestone in some cases is very difficult, even for humans. When wet, some concretes with aggregates can look very similar to granite, and under the shadow, asphalt and worn-off concrete can look very similar. Having a model that can accurately handle the within-class variability with between-class similarity calls for smartly selected training datasets with a good distribution of different classes as well as multiple variants of the same material under different conditions.

4.6.2 Limitations

Even though CitySurfaces can provide city-scale sidewalk material classification, some challenges remain unaddressed. For instance, in the absence of proper sidewalk network data, it can be challenging to map the materials to their corresponding locations accurately. The maps in Figure 4.1 are based on the road centerlines where GSV cars traveled to capture images, depicting the dominant materials for each street segment by taking an average over the materials observed in each image from both the left and right sides of the street. However, knowing the exact location of certain materials is critical for urban designers, planners, and those working with safety and ease of walk for people with special needs. Although our model produces this result at a highly fine level, we cannot depict this variety in detail without proper sidewalk network data. Having separate maps for left and right sidewalks can be one solution, but the intersections where more than one street is captured pose a challenge for assigning the correct materials to each segment.

Also, street-level images have some inherent limitations. Since the images are taken by cars moving alongside streets, in many instances, specifically in dense urban areas, the cars parked on the sides blocked the sidewalk view, as shown in the first street-view image of Figure 4.3. The issue can be mitigated to some extent by adjusting the heading and pitch of the camera, but that solution fails in images with large vehicles like trucks, or when the car with mounted cameras is too close to the sidewalks.

4.7 Conclusion

We present CitySurfaces, a scalable, low-cost approach towards the automatic computation of the spatial distribution of pavement materials at the sidewalk segment level. Our model can detect a diverse range of materials, which to our knowledge, were not covered by any existing dataset. For instance, hexagonal pavers or granite blocks were not reported in any sidewalk inventories reviewed in this study. CitySurfaces produces accurate segmentation

considering multiple cities both within and outside the domain of the training data, demonstrating generalization capabilities across varying urban fabrics. CitySurfaces can detect, delineate, and classify eight standard surface materials used throughout most US cities. As shown in subsection 4.3.3, the framework can be extended to include additional surface materials with less effort than building a city-specific model from scratch, which makes it possible for almost any city or government agency that has spatially dense street-level image data, to create a similar dataset. Moreover, since we have covered the standard materials, such as concrete, asphalt, granite/bluestone, and brick, the model can be applied to a wide range of cities without any further annotation effort or with substantially less effort using our pre-trained model. The models as well as the datasets generated for the six selected cities are publicly available in a GitHub repository.

This work has addressed some challenges in data annotation and accurate classification of different materials with high between-class similarities and within-class variation. The active learning framework utilized in this study helped reduce the annotation costs by choosing the most informative set of data to be annotated and incrementally decreasing the manual modification time. By offering the first comprehensive dataset of sidewalk surface materials at the city scale, this study goes beyond reporting the dominant material of each segment and provides information on the percentage distribution of all detected materials per sidewalk segment. The material classes in this study were selected based on the standard surface materials listed by Boston sidewalk inventory (Boston PWD, 2014), to use it as our baseline ground truth. That list is not extensive and does not distinguish between various types of the same class of material, such as concrete. However, for some more in-depth analysis, such as measuring UHI, we may need to classify the materials differently, and distinguish between different variations of the same material within one class. For instance, reflective granite and dark matte bluestone should have two distinct classes, same goes with the dark and light concretes since they have distinctively different albedo values. The CitySurfaces framework can be easily extended to detect more classes of ma-

terials as illustrated with the Manhattan example in subsection 4.3.3, given the availability of the images corresponding to each class of interest to create the initial ground-truth set. In future work, we plan to take these differences into account and combine the generated data with shadow accumulation (Miranda et al., 2019) to generate a city-scale UHI map.

To facilitate designing automated audit tools, we are going to extend our model to detect surface problems such as potholes, significant breakage, and obstacles on pedestrian paths for accessibility analysis (Miranda, Hosseini, et al., 2020). We also aim to address the walkability and active design of sidewalks by developing a model to detect the relevant features of the sidewalks wall plane and furnishing zone, such as window-to-wall ratio. As another line for our future work, we would like to explore automated sample selection procedures and self-supervised learning techniques and tailor them to sidewalk and pedestrian facility analysis. We chose a simple (yet effective) uncertainty measure and coupled it with the analysis of the model's performance on the validation set and used expert's feedback to refine the annotations and check whether the model is predicting correctly since, on many instances, it is difficult to distinguish between visually similar materials.

CHAPTER 5
CROWD+AI TECHNIQUES TO MAP AND ASSESS SIDEWALKS FOR PEOPLE WITH DISABILITIES

5.1 Introduction

Sidewalks form the backbone of cities. At their best, they offer sustainable transit, help interconnect mass transportation services, and support local commerce and recreation. For people with disabilities, sidewalks support independence, physical activity, and overall quality of life (Christensen et al., 2010; Eisenberg et al., 2017; Harris et al., 2015; Mitchell, 2006a). Despite decades of civil rights legislation, however, city streets and sidewalks remain inaccessible (United States Department of Justice, 1990). As the UN notes, "*[there is a] widespread lack of accessibility in built environments, from roads and housing to public buildings and spaces*" (Nations, 2020).

The problem is not just a lack of accessible sidewalks but also a lack of reliable data on where sidewalks exist and their quality (Deitz, 2021; Eisenberg et al., 2020a; Froehlich et al., 2019). In a sample of 178 US cities, Deitz *et al.* found that only 36 (20%) published sidewalk data, 18 (10%) had curb ramp locations, and even fewer included detailed accessibility information like sidewalk condition, obstructions, and cross controls (Deitz et al., 2021). This lack of data fundamentally limits how sidewalks can be studied in cities, the ability of communities, disability advocacy groups, and local governments to understand, transparently discuss, and make informed urban planning decisions, and how sidewalks and accessibility are incorporated into interactive maps, navigation, and GIS tools (Froehlich et al., 2019; Miranda, Hosseini, et al., 2020).

We argue that any comprehensive analysis of pedestrian infrastructure must include a threefold understanding of *where* sidewalks are, *how* they are connected, and *what* their

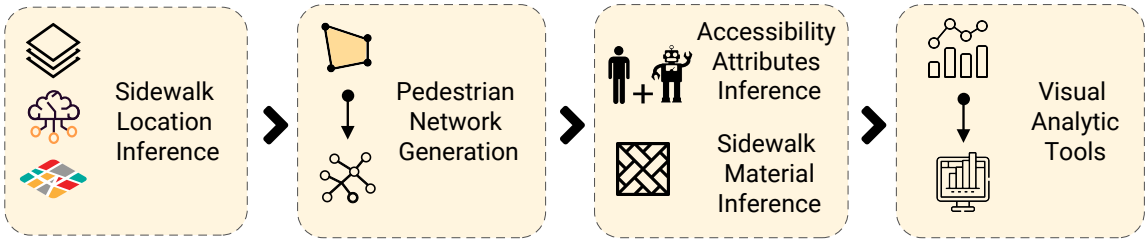

Figure 5.1: We introduce a four-stage Crowd+AI sidewalk pipeline that combines computer vision and crowdsourcing to *locate* sidewalks, build a *network topology*, infer *surface material*, and *assess accessibility*. The resulting output can support accessibility-aware pedestrian routing and new urban science analyses centered on equity and access.

condition is. In this paper, we introduce an initial semi-automatic pipeline that maps *sidewalk locations*, infers *surface materials*, and applies an *accessibility rating* using a combination of crowdsourcing and machine learning techniques (Figure 5.1). To demonstrate its potential, we apply our pipeline to Washington DC and create different visualizations of sidewalk connectivity and accessibility. We close with a discussion of key areas of open research that intersect computer vision, HCI, accessibility, and urban informatics.

5.2 Crowd+AI Sidewalk Pipeline

At the core of our contribution is the threefold integration of sidewalk *location*, *connectivity*, and *condition*. All three are critical to assessing pedestrian infrastructure and building pedestrian-oriented routing tools. To achieve this, we propose a four-stage *Crowd+AI Sidewalk Pipeline* that leverages vision and crowdsourced techniques and aerial and street-level imagery to enable network-level sidewalk assessments. We describe each pipeline stage below.

5.2.1 Extracting Sidewalks from Aerial Imagery

Our pipeline begins with the extraction of *pedestrian pathways*—including sidewalks, footpaths, and crosswalks—from aerial imagery using semantic segmentation. Although semantic segmentation has been broadly used to detect roads and buildings from aerial images (Balali et al., 2015; Iglovikov et al., 2017; W. Li et al., 2019) and to auto-generate

road network topologies (Bastani et al., 2018; Etten, 2020; Wei et al., 2019), it has not been widely applied to pedestrian infrastructure—perhaps due to two key challenges. First, semantic segmentation algorithms require large-scale, high-quality training datasets, which can be labor-intensive and costly to prepare. Thus, researchers often rely on pre-existing publicly available models pre-trained on datasets such as *CityScapes* (Cordts et al., 2016), *Mapillary* (Neuhold et al., 2017), and *ADE20K* (B. Zhou et al., 2017), which historically underemphasize pedestrian-related features. Second, compared to roads and buildings, detecting sidewalks, footpaths, and crosswalks is more challenging due to their comparatively smaller visual footprints and occlusion from shadows, vegetation, and tall structures (Hosseini et al., 2021).

To detect pedestrian infrastructure, we used the model described in chapter 3, which is trained on the open-government datasets drawn from three US cities: Cambridge, MA; Washington DC; and New York City (Hosseini et al., n.d.). The segmentation model outputs a labeled raster image. Each pixel is labeled with one of four classes: *sidewalks* (including footpaths), *crosswalks*, *roads*, and *background*. It uses a hierarchical multi-scale attention model (Tao et al., 2020)— which scales the input image up and down, by a 0.5 factor, during the training and learns at which scale the model is performing better for a certain class. The model then uses that scale during the inference to make better predictions (L.-C. Chen et al., 2016). The detection model is also shown superior performance in occluded scenes and shadows and can distinguish between visually similar classes such as asphalt roads and sidewalks.

5.2.2 Creating Sidewalk Network Topologies

The Stage 1 detection model outputs labeled pixels in rasterized format, fed into our Stage 2 pedestrian network creation algorithm. This algorithm has two key parts: first, we convert the labeled rasters to georeferenced polygons using connected-component labeling (L. He et al., 2009; Rosenfeld & Pfaltz, 1966)—which finds contiguous pixel groups within the

same class to form regions or raster polygons. We then map these polygonal elements into geographic coordinates. Second, to create a node-network diagram of sidewalk connectivity, we use computational geometry techniques to convert the polygons into polylines (the centerline of the polygon).

Here, we addressed some of the challenges described in section 3.5 regarding the TILE2NET network generation algorithm. As it was explained in subsection 3.3.2, the complex shapes of the georeferenced polygons, together with the sensitivity of the dense Voronoi diagram algorithm to the interpolation distance parameter, made it difficult to create a clean network representation. One challenge was filtering polygons based on their shapes to choose a more suitable interpolation distance. Our previous solution mainly relied on basic geometrical attributes such as the area to perimeter ratio, which, while useful in many cases, could not distinguish between elongated, circular, compact, convex or concave, and simple or complex polygons. Hence, we added different measures describing the polygon's shape Boccalatte et al., 2022; Fleischmann et al., 2021, such as fractal dimension (McGarigal, 1995), to describe the complexity of the polygon, circular compactness to describe how close a polygon's shape is to a circle by comparing the polygon's area to the area of the minimal enclosing circle (Dibble et al., 2019), square compactness (Feliciotti, 2018), convexity and rectangularity (Dibble et al., 2019), and elongation (Gil et al., 2012). We also added the polygon's azimuth, defined as the orientation of the longest edge of the polygon.

Next, we selected long, convex, and close to rectangular shape polygons with relatively low fractal dimensions, which often represent sidewalks on the sides of the roadways. We query for not complex (low fractal dimension), convex (high convexity value), and elongated polygons (low elongation index value), which are not compact (low circular compactness, low square compactness) and are close to a rectangular shape (high rectangularity value). Since these polygons do not have complex shapes and are very close to rectangles, we use the minimum bounding box to create their centerlines by connecting the centroids

of the shortest sides of the box. For the rest of the polygons, we continue filtering based on the mentioned shape descriptors, and using the method described in subsection 3.3.2, we choose the interpolation distance to create centerlines and set trim parameters accordingly to remove unwanted branches.

We applied this method to the network generated for DC, which was our lowest performing city. The generated network using our updated algorithm has 34% fewer dangle lines and branches. We used the same evaluation method described in subsection 3.4.3, and our results matched 80.2% of the OSM sidewalk networks in DC, which has improved compared to our previous 76.9%. Overall, these results are promising and demonstrate the potential of automatically creating pedestrian networks from aerial imagery but also suggest opportunities for crowdsourced review and refinement.

5.2.3 Inferring Sidewalk Surface Material

Sidewalk Surfaces and Accessibility

The surface material is critical in the accessibility assessment of sidewalks (Ferreira & da Penha Sanches, 2007; Maghelal & Capp, 2011). The type, color, texture, size, design, cut, and chamfer of pavers can influence the ease and comfort of the walk, the frequency of required maintenance, and the risk of accidents. Uneven surfaces, indistinguishable surface colors, and low-friction materials contribute to the high incidence of outdoor falls in elderly populations (Chippendale & Boltz, 2015; Talbot et al., 2005; Thomas et al., 2020a). Thomas et al. (2020b) found uneven and bumpy surfaces to strongly correlate with walking unsteadiness by comparing various surfaces, including bricks, cobblestone, and flagstones. Sidewalk pavements can create public health hazards such as outdoor falls or pose a barrier to walkability and accessibility of public spaces, specifically for the more vulnerable population and wheelchair users (Aghaabbasi et al., 2018; Chippendale & Boltz, 2015; Clifton et al., 2007; Talbot et al., 2005; Thomas et al., 2020a). For example, porous and high-traction materials should be deployed in regions with extreme climates to prevent the formation of

thick black ice and decrease the risk of falling. These characteristics are critical for at-risk populations such as the elderly and people with mobility or visual impairments (Aghaabbasi et al., 2018; Clifton et al., 2007; Kasemsuppakorn & Karimi, 2008). By examining the interaction between poured concrete sidewalk pavements and varying chamfer widths and wheelchairs, Cooper et al. (2003) concluded that the quality, maintenance, and repair of surface materials have a higher impact on navigability for wheelchair users than the size and number of joints of surface pavers. According to Ferreira and da Penha Sanches (2007), pavement condition, material composition, and effective width are key metrics to determine the sidewalks' accessibility for wheelchair users. Pavement material classification has been used in safety and route-finding applications to alert pedestrians of upcoming obstacles (H. Kang & Han, 2020; C. Sun et al., 2019; K. Sun, Xiao, et al., 2019; Theodosiou et al., 2020) and to help the visually impaired in identifying street entrances based on the change in surface materials captured by cellphones (Jain & Gruteser, 2018).

Material Extraction from Street-level Images

While Stages 1 and 2 produce a sidewalk network topology, they do not include an assessment of sidewalk surface composition (Stage 3) or its accessibility (Stage 4). Thus, in Stage 3, we examine techniques to automatically infer sidewalk surface materials, such as concrete, brick, and cobblestone, which, as explained, can have varying impacts on pedestrian safety and accessibility. We employed *CitySurfaces* (Hosseini, Miranda, et al., 2022), an active-learning-based framework for the semantic segmentation of surface materials that automatically classifies sidewalk materials using omnidirectional streetscape imagery—specifically, *Google Street View* (GSV).

Phase 1. To start the training process, we randomly sampled 1,000 streetscape images from Boston, MA, fed our sample into HRNet-W48 J. Wang et al., 2020 pre-trained on the Cityscapes dataset Cordts et al., 2016, and obtained initial segmentation results. While HRNet outputs 19 classes including *sky*, *trees*, and *buildings*, we filter only to *roads* and

sidewalks. To generate an initial set of labeled surface material data, we use the *Boston Sidewalk Inventory* Boston PWD, 2014—a unique open dataset that describes the dominant surface material of each sidewalk segment collected via manual field surveys: *concrete*, *brick*, *granite*, *concrete/brick mix*, and *asphalt*.

Phase 2. We iteratively train an attention-based model using the labeled images from Phase 1. We begin with 800 images for training and 200 for validation with a batch size of 8 and similar hyperparameters to Stage 1. We train the model in multiple stages. At each stage (10 epochs), we choose the epoch with the highest average IoU on the validation set and qualitatively analyze the results to guide new training data sampling. The weights from the best epoch are used to initialize the next stage's model with more training data. We examine the model's uncertainty estimates to sample new images and select images that performed worst. Following this sampling strategy, we retrieve 300 unlabeled images, apply the current model, correct the predicted labels and add them to the overall training set. To improve model generalization, we begin to include streetscape images from a second city: Manhattan.

5.2.4 Crowd+AI Accessibility Assessments

The above stages produce sidewalk topologies and surface classifications—both of which impact human mobility and people with disabilities—but neither focus specifically on *accessibility*. Thus, in Stage 4, we introduce Crowd+AI techniques to semi-automatically find, label, and assess sidewalk accessibility features in the built environment, such as *curb ramps*, *surface problems*, and *obstacles*. In previous work, we demonstrated that online streetscape imagery is an accurate source for assessing accessibility infrastructure (Hara, Azenkot, et al., 2013, 2015) and that with our custom labeling tools, minimally trained crowdworkers could accurately and quickly find street-level accessibility problems (Hara et al., 2015; Hara, Le, et al., 2013; Hara et al., 2012). However, relying solely on human labor limits scalability. We then explored how to effectively combine automated methods

Figure 5.2: Stage 4 uses Crowd+AI techniques to label accessibility features/barriers in the pedestrian environment. Above, a user labeled a *curb ramp* (in green) and an *obstacle* (in blue) in Project Sidewalk (M. Saha et al., 2019)

with crowd work (Hara et al., 2014). Our first hybrid Crowd+AI system, *Tohme*, infers the difficulty of a sidewalk scene using a trained SVM and allocates work accordingly to either a computer vision-based pipeline or human users (Hara et al., 2014). In a study of 1,000 street intersections across four North American cities, Tohme performed similarly to a purely human labeling approach but was more efficient. While promising, Tohme was limited to a small training dataset and only supported one sidewalk feature (curb ramp recognition).

Thus, we began to explore more scalable approaches, culminating in *Project Sidewalk*—an interactive online tool that allows anyone with a laptop and Internet connection to remotely label accessibility problems by virtually walking through city streets in GSV, similar to a first-person video game (Figure 5.2). In a 2018 pilot deployment, 1,400 users virtually audited 2,934 km of D.C. streets, providing 250,000 sidewalk accessibility labels (M. Saha et al., 2019). With simple quality control mechanisms, we found that remote users could find and label 92% of accessibility problems, including *missing*

curb ramps, *obstacles*, *surface problems*, and *missing sidewalks*. To qualitatively assess reactions to our tool, we also conducted a complementary interview study with three stakeholder groups (N=14)—government officials, people with disabilities, and caregivers. All felt that Project Sidewalk enabled rapid data collection, allowed for gathering diverse perspectives about accessibility, and helped engage citizens in urban design. Key concerns included data reliability and quality, which are ongoing research foci in our group.

Building on this D.C. pilot and working closely with local government partners and NGOs, we have deployed Project Sidewalk in ten additional cities, including Mexico and the Netherlands. Thus far, we have collected over 700,000 geo-located sidewalk accessibility labels and 400,000 validations—to our knowledge, the largest and most granular open sidewalk accessibility dataset ever collected. This large, ever-growing labeled dataset of images paired with advances in computer vision has enabled new deep learning methods for automatic sidewalk assessment. In Weld *et al.* (Weld et al., 2019), we showed how a modified version of *ResNet-18*—which incorporates LIDAR depth, scene position, and geography features in addition to pixels—could achieve state-of-the-art performance in automatically validating human labels (average precision/recall: 81.3%, 77.2%). We also presented the first examination of cross-city model generalization showing that one city's labels (D.C.) could be used to pre-train model weights for two other test cities (Seattle, WA, and Newberg, OR).

5.3 Demonstrating Proof-of-Concept

To demonstrate the potential of our approach, we apply our four-stage pipeline to Washington DC and create sidewalk visualizations of topology, surface material, and accessibility. D.C. provides an interesting testbed: it has over 1,100 miles of city streets, diverse and historic urban designs, and is a popular tourist destination; however, no official pedestrian network data exists for the city.

First, we use our Stage 1 algorithm to detect pedestrian pathways from Washington

DC orthorectified aerial images. Then, in Stage 2, we converted the auto-labeled sidewalk, footpath, and crosswalk rasters into georeferenced polygons and centerlines. Finally, to compute accessibility metrics, we incorporated surface material inference data (from Stage 3) and crowdsourced accessibility information (from Stage 4).

To extract the pedestrian pathways, we fed 73,000 orthorectified satellite image tiles obtained from Washington D.C. open data (DC GIS, 2020) into our detection model, described in subsection 3.3.2. We then use TILE2NET to convert the raster results into georeferenced polygons. Using our improved algorithm, we construct the network of sidewalks for Washington D.C.

We fed 78,786 street-level images from D.C. to our CitySurfaces model to create the surface material data. The model outputs a tabular dataset of the number of pixels belonging to each class per image. To compute the percentage of sidewalk materials per street/sidewalk segment, we need to know the percentage of each material per *detected* sidewalk, not per image, since some images may only contain some part of a sidewalk due to occlusion. The number of pixels belonging to a certain class for a given sidewalk in an image can change based on how much sidewalk is captured in that image. To do that, we first filter images where the sum of all pixels equals at least 0.97 of the sum of class road and background, meaning we have no sidewalks in the image. Then, for each image, we calculate *sidewalk material composition* of each sidewalk material, as the ratio of the number of pixels belonging to any of the seven classes of sidewalk surface materials (concrete, bricks, mixed, asphalt, granite/bluestone, granite block/decorative stone, and hexagonal asphalt pavers.), to the sum of the total pixels belonging to these seven classes detected sidewalks.

We compute a similar measure for the detected road surface with our two classes, asphalt roads, and cobblestones. Next, having the geographic coordinates of each GSV image, we will do a spatial join between image-wise *sidewalk material composition* and the road centerlines, as well as the sidewalk network data. Our aggregation method takes the

average of each class, so the sum of sidewalk material composition for all sidewalk classes equals one in each segment.

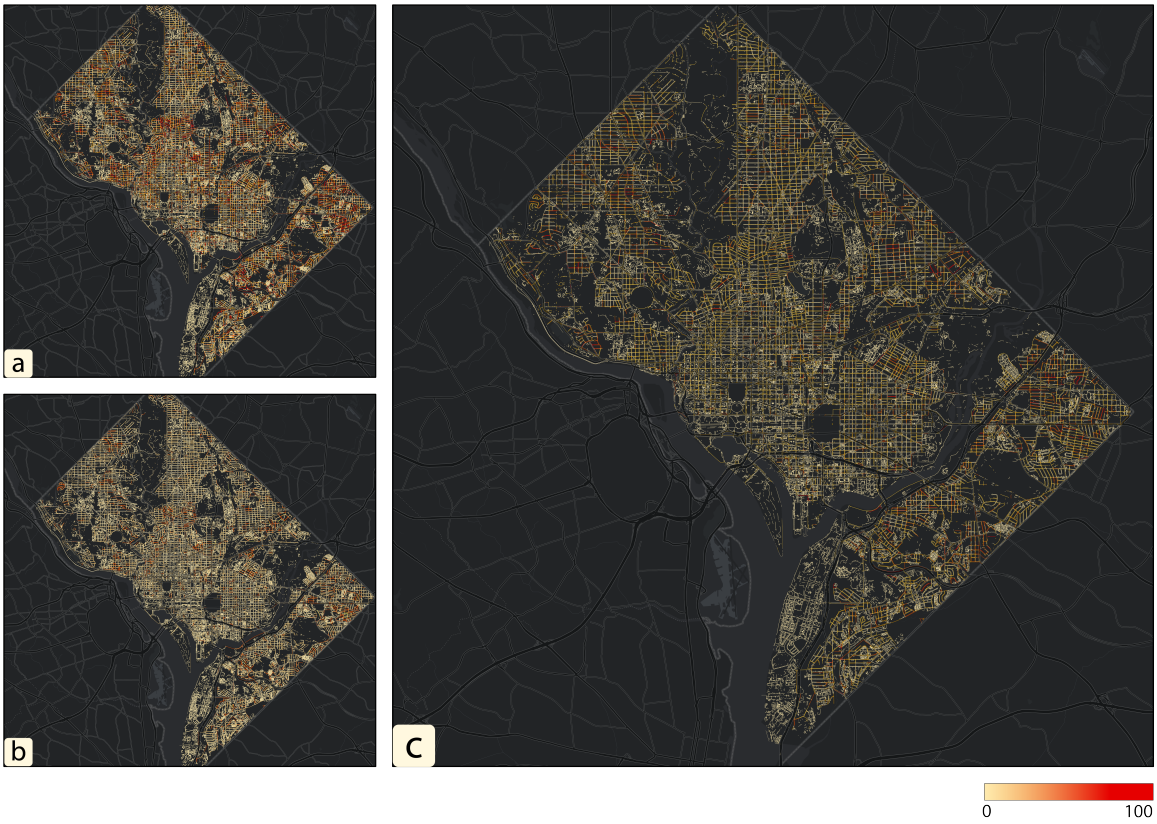

Figure 5.3: Proof-of-concept of our pipeline in Washington DC.

We produce three proof-of-concept visualizations based on computed sidewalk-accessibility scores—an open area of research (A. Li et al., 2018; M. Saha et al., 2022). First, we created a sidewalk heatmap visualization using Stage 4 accessibility data (Figure 5.3a; red is worse). We differentiate between street crossings, which connect sidewalk segments, and the sidewalk segments themselves. For the street crossings, we associate *curb ramps* and *missing curb ramps* with intersections and compute a *crossability score*. We calculate a severity-weighted sum of all accessibility problems over each sidewalk segment for the sidewalk segments. Second, we created a sidewalk heatmap visualization that incorporates Stage 3's surface material inference data (Figure 5.3 b). Here, we apply higher weights to bricks and cobblestone surfaces, which pose higher tripping hazards to people using mobility aids and bumpy, uncomfortable surfaces for wheelchair users. Finally, we created

a hybrid visualization that incorporates both surface material (Stage 3) and accessibility (Stage 4) shown in Figure 5.3 c.

5.4 Discussion and Future Work

Our overarching vision is to develop scalable Crowd+AI techniques capable of *mapping* and *assessing* every sidewalk in the world. In this paper, we introduced a preliminary four-stage pipeline that extracts sidewalk *locations*, infers *surface materials*, and applies an *accessibility* rating using a combination of computer vision and crowdsourcing. While prior work has examined each in piecemeal, we offer the first comprehensive pipeline towards addressing the grand challenge of identifying *where* sidewalks are, *how* they are connected, and *what* their condition is (Froehlich et al., 2019). All three are needed to create accessibility-aware pedestrian routing algorithms, interactive maps of neighborhood accessibility, and to enable equity analyses examining sidewalk infrastructure availability/condition and key correlates such as race, real-estate pricing, and socio-economics.

Towards future work, we would like to examine: (1) how the crowd and A.I. can work together in each stage to improve efficiency and accuracy; (2) how our methods perform across varying urban fabrics and geographic contexts; (3) and advance understanding of the underlying biases in our methods—where do they fail and why?

Finally, we call on this cross-disciplinary community to create a database of high-quality, labeled satellite and streetscape scenes for sidewalks and sidewalk accessibility problems along with computer vision benchmarks, which has been so critical to innovation in other ML-based areas.

CHAPTER 6
CONCLUSION

At a high level, the goal of this dissertation has been to address the lack of fine-level data describing pedestrian infrastructure and to provide techniques and tools rooted in domain theory to answer the requirements and needs of urban planners and researchers. Towards this goal, we designed and developed a set of models and tools for large-scale assessment of the quality of urban sidewalks. This dissertation extends the research on walkability and accessibility of sidewalks, assessment of the urban built environment, and human-scale design of the cities by proposing a set of novel computer vision-based approaches to collect the built environment data at a large scale systematically.

6.1 Summary of Contributions

In this section, we restate the contributions listed in the Introduction chapter and summarize how each of these contributions was achieved.

6.1.1 Models and tools to analyze sidewalks at different scales

In chapter 5 we argue that any comprehensive analysis of pedestrian infrastructure must include a threefold understanding of *where* sidewalks are, *how* they are connected, and *what* their condition is and thus, requires data at human scale as well as city-scale. The research done in this dissertation tackled the problem of pedestrian infrastructure assessment at the *human scale* by proposing CitySurfaces (chapter 4), which uses *street-level* images, and at the *city scale* by introducing TILE2NET –an end-to-end framework for automated mapping of pedestrian infrastructure using *orthorectified aerial imagery* (chapter 3).

CitySurfaces leverages computer vision techniques for classifying sidewalk materials using widely available street-level images. It was designed to have more insights into how

the model correlates different visual features during the training process. Through an active learning scheme with expert-in-the-loop, we help reduce the faulty and false associations which often happen when the machine learning models are treated as a black box. Although it can make the training process longer, by examining how the model performs after each stage (10 epochs in our case), we could detect some interesting patterns of errors (Figure 4.11, Figure 4.13) and choose the right training data to teach the model to make better predictions. We trained the framework on New York City and Boston images, and the evaluation results show a 90.5% mIoU score. Furthermore, we evaluated the framework using images from six different cities, demonstrating that it can be applied to regions with distinct urban fabrics, even outside the domain of the training data. The models and codes are made publicly available through our GitHub repository.

TILE2NET was motivated by the lack of open source tools tailored for sidewalk analysis and the fragmented available tools to handle each part of the pipeline. Urban datasets are numerous, loosely linked, and can be laborious to sift through. TILE2NET connects various pieces, from orthorectified aerial tiles and planimetric data to handing tile system and its associated computations, sidewalk detection through semantic segmentation, vectorizing and georeferencing the raster data, and finally, creating the topologically correct pedestrian network. We use a semantic segmentation model that can detect sidewalk, footpath, and crosswalk polygons from orthorectified tiles and then use the resulting polygons to create an interconnected network. The approach was pilot tested in Manhattan, NY, Washington, D.C., Boston, and Cambridge, MA, and achieved high accuracy in each of these cities. The tool and the semantic segmentation model will be made publicly available to facilitate research on urban sidewalk analysis.

Designing such tools can mitigate problems created by unequal distribution of investments and poor governance in introducing and conducting pedestrian-level data collection projects. It also enables urban planners, practitioners, and municipal decision-makers to gain new insights into the sidewalks' current condition, monitor their compliance with the

official codes and guidelines designed to serve diverse sidewalk users, and combine such information with various socio-demographic data to form new hypotheses. While each of these works has approached the sidewalk assessment problem from one scale, in chapter 5 we offer the first comprehensive pipeline towards addressing the grand challenge of identifying *where* sidewalks are, *how* they are connected, and *what* their condition is. All three are needed to create accessibility-aware pedestrian routing algorithms and interactive maps of neighborhood accessibility and to enable equity analyses examining sidewalk infrastructure availability/condition, and key correlates such as race, real-estate pricing, and socio-economics.

6.1.2 Addressing annotation challenges with two different techniques

The quality of the training data has a significant impact on the quality of the model's inference. Even the most robust architecture would fail if trained on noisy and inaccurate data. As explained in section 2.2, the substantial cost of accurate annotation restricts the practicality of semantic segmentation on new datasets and tasks relevant to urban analysis (Montoya-Zegarra et al., 2014; Xie et al., 2020). Hence, it is often difficult to find annotated datasets for sidewalk-related tasks. To ensure quality and accuracy, we had to annotate a large number of images to train models in both chapter 3 and chapter 4. We used two different techniques to overcome the annotation bottleneck based on the nature of the task, the type of images used, and the availability of external datasets. This section describes our two strategies and how each can contribute to pushing pedestrian analysis forward.

1. *CitySurfaces:* To our knowledge, no annotated street-level or in-the-wild image dataset exists for sidewalk materials. To address this gap while ensuring the quality of our segmentation task, I annotated more than 3000 images from scratch for this project. Using this framework, the annotation time decreased from approximately 25 minutes to 4 minutes per image for the last set of images. Although the data cannot be pub-

licly shared due to Google's restrictions on using street-level images, the resulting model has a strong performance that can be relied on for generating similar datasets in much less time. Moreover, the paper provides a framework for applying active learning with unlabeled data that needed to be annotated live during the training process; while the majority- if not all- of the studies exploring active learning techniques in urban scene parsing and semantic segmentation use available, annotated datasets and only deal with choosing the right data (Colling et al., 2020; Golestaneh & Kitani, 2020; Kasarla et al., 2019; Mackowiak et al., 2018; Xie et al., 2020). We proposed a stage-wise model, which would require new samples after ten epochs instead of feeding new samples at each iteration, which would not be feasible with live data annotation.

2. TILE2NET*:* The semantic segmentation model was trained on annotation labels created using official data provided by city municipalities. The official data come as georeferenced vector polygons and cannot be directly used as training data. As explained in section 3.3, TILE2NET is designed with the capability to automate the creating of these annotation masks. It takes the bounding box of each tile, finds the corresponding sidewalk, footpath, crosswalk, and road polygons from the available planimetric GIS data, rasterized the GIS polygons into pixel regions, and outputs annotated image tiles with four total classes: sidewalks (including footpaths), crosswalks, roads, and background, representing each class with a distinct color. These annotations are used as ground truth data for training the model. As discussed in detail in chapter 3, the lack of consistency between the mapping standards used by different municipalities, and numerous instances of mismatch between the official data and the satellite images, posed another set of challenges in training a generalizable model. To overcome this issue, we created a set of "rules of capture" to accommodate different mapping standards across municipalities. With the help of a team of undergraduate planning students from MIT, we modified and addressed temporal

differences between the GIS data and the aerial images. Appendix B presents our rules of capture and editing strategies which were shared with our team to guide their annotation modification efforts. The rules are designed with careful consideration of urban planning and accessibility concepts. In total, our research team manually corrected 2,500 image tiles of 12,000 in the training set (20.8%), 1,620 of 4,000 in validation (40.5%), and 1,500 of 4,000 in test (37.5%). Consequently, The model trained on this data outputs more uniform and consistent results that can be used for sidewalk analysis across different cities.

6.1.3 New datasets describing sidewalks for multiple cities

The analysis done through this dissertation research led to the creation of new city-scale datasets described below.

1. Surface material dataset for five cities. Except for Boston, none of these cities had any data describing their sidewalk materials at street level detail. The Boston dataset is limited to five standard materials and only reports the dominant material used in each segment. At the same time, CitySurfaces offer the percentage of materials used in each segment.

2. Pedestrian network dataset for four cities. As discussed in section 5.2, complete pedestrian network data was only available in Boston and Cambridge. Even resourceful cities like New York and Washington, DC, did not have pedestrian network data. New York City does not have any public datasets of georeferenced crosswalks.

6.2 Implication

This section reflects on the implications of this dissertation research across different domains.

6.2.1 Creating sidewalk inventories for different cities

City agencies can create new or update sidewalk inventories at a much lower cost by using the models provided in this research. CitySurfaces can detect, delineate, and classify eight distinct standard surface materials used throughout most US cities. The framework can be extended to cities with additional surface material types with less effort than building a city-specific model from scratch, which empowers almost any other city or government agency with spatially dense street-level image data to create a similar dataset. Since CitySurfaces covers standard materials, such as concrete, asphalt, granite/bluestone, and brick, our model can be applied to a wide range of cities. Moreover, since the model focuses mainly on sidewalks, it works better at detecting sidewalks and their boundaries than many publicly available datasets.

Aside from detecting the materials, CitySurfaces can be used to detect the existence of sidewalks in street-level images and create annotation labels for semantic segmentation of other objects on the sidewalks, such as obstacles, poles, planters, and as such. The model can be used to make inference on unlabeled images from different cities with similar streetscape features, and the results can be used, with minimal effort, to create new classes and train new models since it can already detect such objects and classifies them as background with clear boundaries around them. The multi-scale architecture in both parts of this work makes it possible to work with images of different sizes, creating more flexibility for applying them to other datasets.

The semantic segmentation model of TILE2NET makes it possible to detect sidewalks from high-resolution orthorectified tiles to generate pedestrian network data with high accuracy. It makes handling tile computation easy and provides automated methods to create annotations from publicly available datasets. TILE2NET is designed with the capability of automating the data preparation process. It can take as input the textual name or geographic coordinates of the bounding box of a given region and download the tiles that fall within the bounding box for the cities where orthoimagery is available. Currently, it offers automatic

downloads of NYC, Boston, MA., Washington DC, and Seattle, WA.'s most recent aerial tiles.

6.2.2 accessibility aware routing apps

Having comprehensive network data is essential for safe and easy navigation. Today, GPS-based navigation apps have become an integral part of our daily travels. The underlying data for navigation applications is a map database, which existed for vehicular road networks far before their wide-scale use in navigation applications (Y. Zhao, 1997). Nevertheless, comparable data describing pedestrian paths seldom exists, and the locations and types of sidewalks are rarely mapped or updated. Instead, the majority of the existing navigation apps rely on road networks to guide pedestrians through different streets, which can lead to several problems. One of the common problems of relying on road networks for pedestrian navigation is the limited extent of the locations it covers. A network constructed based on streets and roads does not include any off-road footpaths. In other words, it means being limited to only where roads can go, which can lead to inaccuracies (e.g., streets with no sidewalks), simplifications (e.g., assumptions that buildings can be directly accessed on both sides of a street centerline, while in reality crossing a street is only allowed at certain locations), and misrepresentation (e.g., assuming pedestrian connections based on vehicular routes, where there are none) (Chin et al., 2008; Ellis et al., 2016), each of which can lead to potentially hazardous situations for pedestrians, specifically the more vulnerable population (M. Saha et al., 2019).

Given this lack, sidewalk mobility has not benefited from a wave of technological innovation in routing applications, pedestrian-centered location-based services (e.g., deliveries and services using active modes), or evidence-based infrastructure investments that would channel scarce tax-payer dollars into sidewalks and public spaces that likely impact the greatest number of constituents. Having pedestrian paths represented as continuous, topologically connected network datasets could open up new (and overdue) efforts for pedes-

trian routing, flow analysis, and potential location-based or delivery services. Transit-first policies, walkable-streets initiatives, step-free access for public transport, and vision zero goals represent but few planning and policy areas which could benefit from citywide sidewalk and crosswalk datasets.

6.2.3 Fall prevention programs

The demand for age-in-place among older adults increases the need to audit the built environment and ensure its safety. Existing tools are limited in their geographic scope, focusing on specific neighborhoods and not making use of the technological opportunities available today (Eisenberg et al., 2020b). Fall prevention programs are aimed at assisting the elderly population in navigating safely in their neighborhood streets. They assess the condition of different routes, accounting for risk factors such as paving materials, street enclosure, lack of sunlight, and obstacles on the way to inform their participants about the condition of the walking environment along the route and equip them with strategies to prepare for each condition before the strolling starts (Chippendale, 2020). But the lack of comprehensive data describing the conditions of pedestrian facilities at a fine, human-scale level significantly limits the geographic coverage of their practice, specifically when the locations are distant or unfamiliar. As discussed in section 4.5, using the data about the percentage of different materials in each street segment and the amount of shadow accumulation, we can inform pedestrians about the condition of the walking environment they are planning to visit.

6.2.4 Water Runoff Management

The type of surface material and its porousness can impact the water runoff and increase the risk of flooding. Sidewalks and roads form the main part of the urban ground surfaces. Excessive use of impermeable materials, which prohibit the infiltration of the water into the underlying soil, increases both the magnitude and frequency of surface runoffs (Bell et al.,

2019; Shuster et al., 2005), reduces the groundwater recharge, and negatively impacts the water quality. Having detailed information about the type of materials used in different parts of cities can help design appropriate disaster management plans and choose more effective strategies.

6.3 Limitations and Directions for Future Research

In this section, we cover the primary limitations of this dissertation and highlight opportunities for future work.

6.3.1 Creating databases of labeled images for sidewalks

One of the challenges in semantic segmentation of pedestrian facilities is the lack of standardized, high-quality labeled image datasets. One direction for future research could be to create publicly available, standardized datasets with clear rules of capture for annotating different classes. The dataset can be accompanied by an active learning-based tool with capabilities to add new features and classes to incorporate incremental learning strategies.

6.3.2 Improving the network generation algorithm

Although our model performs quite well in detecting sidewalks and crosswalks in many challenging scenes, the network generation algorithm is still far from perfect. One of the biggest challenges is to apply post-processing techniques that close the gaps in the network or treat some irregularities without adding arbitrary or fake connections. Unlike roads, pedestrian networks do not guarantee connectivity; sidewalks can unexpectedly end where they should have been continuous, they may not even exist in many areas, and pedestrian footpaths can have irregular forms. None of these problems exist in street networks, and the body of research on pedestrian network representation is very limited. For the road networks, the job that was done by NavTech in the 1990s and TomTom, who did a lot of mapping in the US initially and then globally. To create an accurate and reliable network,

they ended up driving the routes, going through every single path and validating whether that path exists and is derivable or not, and hence, in a way, the whole network was ground-truthed to every single segment that ended up in the map. For the pedestrian network, we are not able to do the ground truthing the same way as the vehicles were able to do. The algorithmic approach can work as a scalable solution to clean up as many areas as possible and filter the regions in need of inspection. Nevertheless, to be used as a basis for pedestrian and wheelchair users' navigation, there is still a necessity for cities to do manual field visits in places where the gap in the network exists to distinguish between places where the gaps are real and those created due to the shortcomings of the model.

One future direction for this research is to build algorithmic ways to handle the complexities of pedestrian networks by 1)improving the polygon to line conversion method and 2)improving the post-processing and network simplification methods. The former mostly requires surface reconstruction and computer graphics techniques, while the latter requires computational geometry and morphological analysis methods.

6.3.3 Extending the sidewalk detection model

The sidewalk detection model presented in chapter 3 could benefit from adding more classes, such as driveways, and stairs, to create richer network data. Moreover, separating footpaths and sidewalks into separate classes can be helpful for connectivity and continuity analysis and network generation algorithms. Due to their more complex shapes, drawing the centerline of footpath polygons is often more challenging. It requires different parameters for both the Voronoi algorithm and post-processing steps.

6.3.4 Global scale analysis of pedestrian facilities

The three studies presented here were limited to US cities. An interesting avenue for research can be extending these models to cities worldwide and exploring the challenges of applying models trained on US data to cities on other continents. Considering the high vari-

ations and inconsistencies in the datasets of US eastern cities, we expect to see even sharper differences in a global context. Given the sharper differences in the global context, we can investigate whether having country/region-specific models can provide more accurate and reliable predictions compared to models trained on multiple cities worldwide. The global scale presents some unique challenges in different stages of the process, from handling a large number of high-resolution aerial images, pre-processing and optimizing the method and algorithms for speed and accuracy, to incorporating varying design elements and urban fabrics of different countries in the analysis. Using the computational geometry and network science (Wasserman, Faust, et al., 1994) methods, we can look into 1) designing automated models to construct a richer and more diverse pedestrian network representation that can be applied to various cities around the world and 2) using this network as a basis for a comparative, cross-country analysis on the measures of accessibility and urban morphology and their correlation with different socio-demographic indicators.

6.4 Final Remarks

Although new data sources and knowledge discovery systems can help us wrestle with tricky questions, we impoverish our ability to reason with computers if we do not center theory when we create computational representations of the real world—even if we must rethink or advance our technologies and tools to do so (Boeing, 2020).

Current research in developing tools for urban data analysis and visualization shows a considerable lack of understanding of the domain field concepts and theories. This is mainly due to the fact that the majority of these tools are not developed by people who have proper training in the domain fields such as urban geography, urban planning, or Urban social theories Boeing, 2017 and conversely, students with sound theoretical backgrounds do not have the required skills to develop the tools they need.

Due to their different outlook and nature, it is not easy for experts in humanities and engineering disciplines to communicate effectively through published research. Hence, the

majority of the published works are circulated within the research community of each field. Articles written by urban scholars are often not read by researchers working on developing new tools, and papers introducing new urban planning support tools often don't reach their target audiences, and when they do, they are often considered too "technical" or missing the grounded theoretical framework. This becomes more evident when such papers target long debated, complex concepts such as "walkability," where knowing *how* and *from what perspective* walkability is defined makes a huge difference in *how* and *for which purpose* the tool can be utilized by urban researchers and practitioners.

This dissertation reflects my journey and efforts to make this conversation possible and to help create the foundations for bridging this gap and raising awareness about the critical concepts and concerns of designing inclusive and accessible pedestrian facilities.

Appendices

APPENDIX A

APPENDIX FOR CHAPTER 4: SAMPLING STRATEGIES

A.1 Uncertainty in predicting unlabeled images

Uncertainty sampling is one of the most frequently used query methods to select a new sample of training data in active learning (Settles, 2009). To measure the uncertainty, we use softmax probability, which has been commonly used in active learning as a strategy for choosing the next training sample (Settles, 2009). We use the outputs of the softmax layer as part of the sampling strategy, which can partly reveal the most challenging instances for the model to predict. We apply multi-class uncertainty sampling known as margin sampling (MS) (Scheffer et al., 2001), which calculates the difference between the two highest prediction probabilities on softmax to produce uncertainty maps. The smallest margin in each map is then chosen as the image-level uncertainty. The MS measure is defined as:

$$x^*_{MS} = argmin_x P_\theta(\hat{y}_1|x) - P_\theta(\hat{y}_2|x) \tag{A.1}$$

where $\hat{y}_1$ and $\hat{y}_2$ are the class labels for pixel x, with the first and second highest probability, respectively, under the model θ. The lowest margin gives us the highest uncertainty, which is used as an image-level uncertainty measure.

To select new samples, we feed the pool of unlabeled images to our network, obtain the segmentation and calculate image-level uncertainty to select images with the highest uncertainty. We start by selecting 10% of the images using this strategy. As the training proceeds, we increase the share of images selected through this strategy at each stage by 10%.

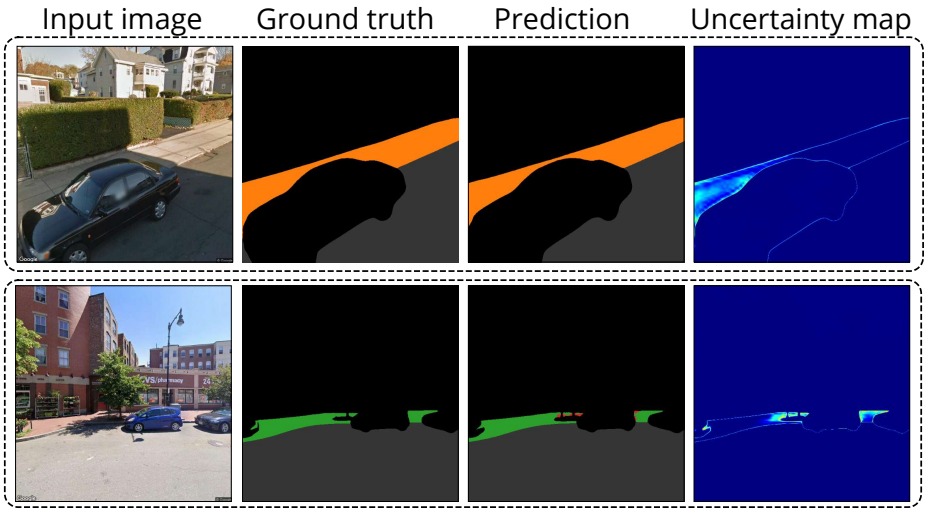

Figure A.1: Two different scenarios of using the model's output and uncertainty map in sample selection. The warmer colors in the uncertainty map represent areas where the model was less confident in its prediction. **Top**: the model correctly predicted the class in a previously identified challenging setting (shadow) but was less certain in predicting the shadowed areas; **Bottom**: The model classified the parts in shadow as concrete alongside brick and outputted mixed class for that region. The uncertainty map shows that the model was least certain in its prediction for that area.

A.2 Performance on validation set

Since softmax probabilities do not necessarily represent the true *correctness* likelihood, a problem known as "confidence calibration" (Guo et al., 2017), we need other strategies as well to select an informative sample for the model. To this end, at each stage, we examine the performance of the best epoch on the validation set and select 10% of the best predictions and 20% of the top failures. Images from failure and success cases are then clustered using K-means (Cover & Hart, 1967; Fix, 1985) with the Euclidean distance to investigate potential common patterns in each group. In each cluster, we rank images based on the average IoU of all classes, excluding road and background. We then select images with the highest error rate. The error rate is defined as the sums of false positive and false negative predictions of the model in each image. Aside from the described method, we examine the clusters of images to detect common error-causing patterns. Figure A.1 (bottom row) depicts a brick sidewalk that the initial model incorrectly segmented the part next

to shadowed regions as the “mixed” class. Its associated uncertainty map reveals prediction difficulty near the edge of the car and the plant pit, which are incorrectly classified as mixed. Uncertainty maps of the success cases are examined to find regions where the model is least confident while making a correct prediction. Figure A.1 highlights a set of uncertainty maps. After we find the most error-prone images, we use them to find similar unlabeled images. We extract their feature maps using the backbone HRNet-W48 (K. Sun, Zhao, et al., 2019; J. Wang et al., 2020) and employ cosine similarity distance to retrieve similar images from the pool of unlabeled data.

APPENDIX B

PEDESTRIAN NETWORK ANNOTATION MODIFICATION PROJECT

The goal of this short term project is to modify the inaccuracies of the annotation labels that will be used to train a semantic segmentation model to detect pedestrian infrastructure from satellite imagery. The annotations are created using the official, public data published by different city agencies (NYC, Cambridge and DC in this case). As you will see, such data is not always accurate and comes with various errors and incorrect classifications.

You will be given image files in PNG format. Each file contains a satellite image (left) and an annotation label (right). The annotation label is an image where each pixel's value (color) represents the class that pixel belongs to. In this study, we have four total classes: sidewalk, crosswalk, roads, and any other object in the image will be assigned the class background (Figure B.1). Table B.1 shows these four classes and their associated color codes.

Table B.1: Classes and RGB color codes

Class	Color	RGB code
Road	Green	(0, 128, 0)
Sidewalk feature (including pedestrian footpaths)	Blue	(0, 0, 255)
Crosswalk	Red	(255, 0, 0)
Background	Black	(0, 0, 0)

In general, you should identify the incorrectly annotated areas as well as missing features. The incorrect annotations can range from dislocated features to cases of total error where an annotated feature is not present in the satellite image.

B.1 General rules of capture

1. If the annotated area is not visible in the image and there is no way of guessing the shape of the feature, that part should be set as background (black) (Figure B.2).

Figure B.1

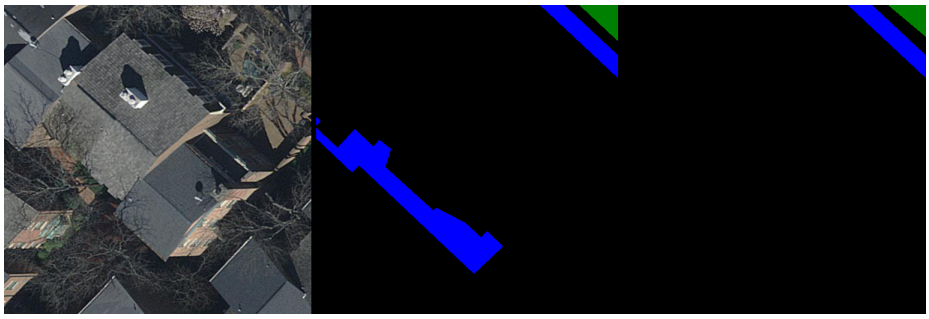

Figure B.2

2. Raised medians should be annotated as background

3. Large planter areas should be annotated as background

4. Any objects other than the ones belonging to class sidewalk, road, or crosswalk should be classified as background (e.g. planter areas, buildings, etc.)

B.2 Sidewalks

This class includes a wider range of pedestrian infrastructure, such as off-road footpaths, medians that connect crosswalks, plazas, footpaths to buildings, and stairs.

B.2.1 Rules of capture

The rules of capture for sidewalks are mainly similar to the NYC planimetrics rules of capture.

1. In areas where equipment is stored or installed on the sidewalk, the full extent of the sidewalk was approximated.

2. Sidewalks were annotated when crossing large medians or traffic islands (Figure B.3).

Figure B.3

3. Sidewalks overlay the exit and entrance portion(s) of parking lot features and driveways (b) and alleys overlay the sidewalks (a) (Figure B.4).

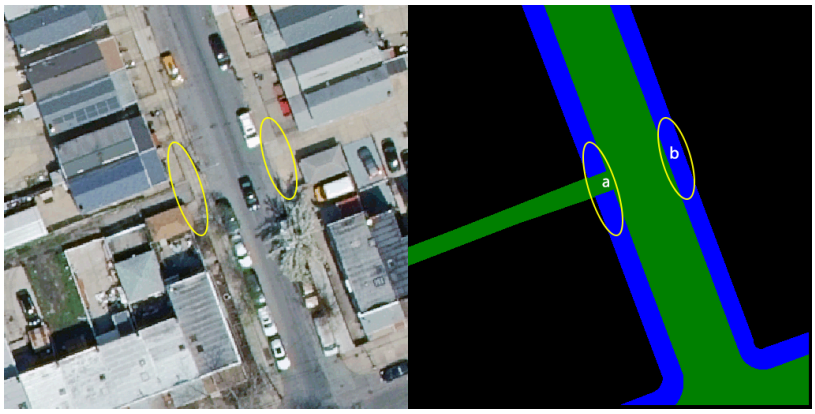

Figure B.4

4. Sidewalks should be annotated as continuous regions when they are obstructed by different objects such as trees, bridges or shadows, if they are visible on both sides of the obstruction (Figure B.5 left) or if the obstruction is extended to the edges of the image (Figure B.5 right).

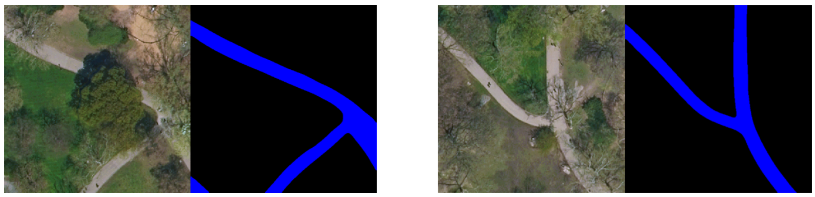

Figure B.5

5. Furnishing zones of the sidewalks should be annotated as sidewalks unless we have large planter areas in the furnishing zones.

6. large public spaces adjacent or connected to sidewalks are annotated as sidewalks

B.2.2 Examples of common incorrect cases that should be edited

1. Sidewalks obstructed by buildings are not annotated while in real-world the sidewalks continues through the obstruction (Figure B.6). In the corrected annotation the gap should be filled and the sidewalk should be continuous in that area.

Figure B.6

2. Missing pedestrian footpaths (Figure B.7)

3. Highway shoulders annotated as sidewalk features.

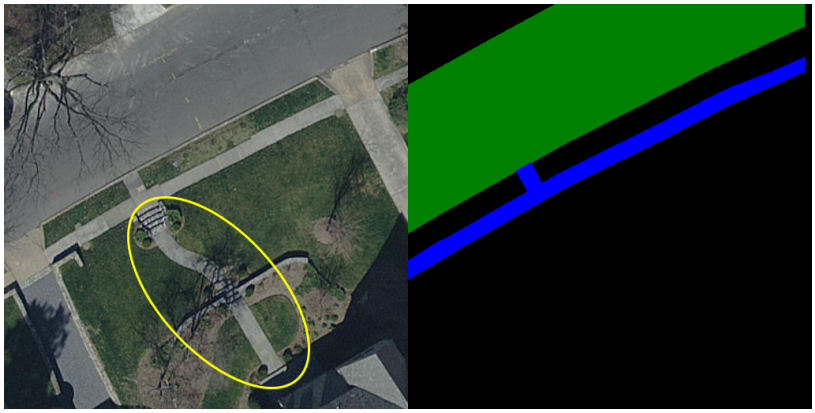

Figure B.7

B.3 Crosswalks

This class includes the marked crosswalks of different types such as standard, mid-block or controlled crosswalks, and enhanced crossings. All different markings of crosswalks should be annotated as solid rectangular polygons attached to both ends of the pedestrian-dedicated spaces.

B.3.1 Rules of capture

1. Crosswalks should be continued up to the edges of the sidewalks, medians or curb cuts.

2. Crosswalks should be annotated as one solid polygon even if they are demarcated as two parallel lines.

B.3.2 Examples of common incorrect cases that should be edited

1. Crosswalks are not marked (we don't see them in the image) but they appear in the annotation mask (a), are visible in the image but they are not annotated (b), are detached from the sidewalks/curbs or the medians connecting them (c) (Figure B.8).

2. Crosswalks demarcated by two parallel lines and annotated as two parallel narrow polygons should be changed to a larger polygon covering the whole area between the

Figure B.8

two lines.

Figure B.9

3. Crosswalks annotated with wiggly edges - the selected part should be removed and replaced with the green color since [in this case] it is part of the roadbed (Figure B.10).

B.4 Roads

The rules for roads are mainly based on NYC capture rules NYC planimetrics. Read the section on roads carefully to be able to distinguish sidewalks from road shoulders. Driveways should not be annotated as roads.

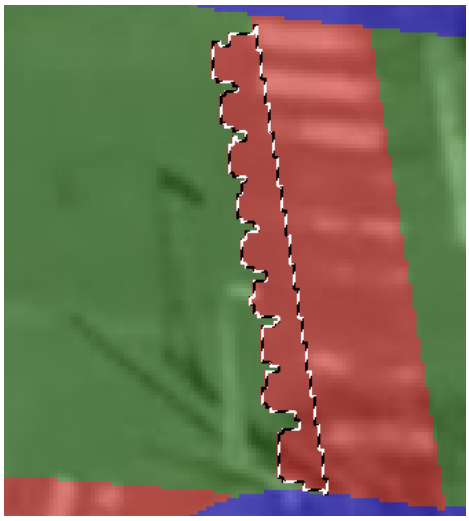

Figure B.10

B.4.1 Examples of common incorrect cases that should be edited

1. Parking lots annotated as roads and should be changed to background (black)

2. Large planter areas annotated as roads (Figure B.11)

Figure B.11

3. Railroads annotated as roads

B.5 Adobe Photoshop recommended settings

In the package, you can find a Photoshop action file called copyhalf.atn . Load the action in your action panel. You can use it to automatically create an overlay of satellite image

over the annotation part.

We have also provided the swatch library specific to this project in a file called Ped-Net.aco. The four colors included in this library are the only colors you should use for the whole project. These four colors correspond to the four classes described before.

1. Set the image mode to RGB color - 8 Bits/channel (Figure B.12)

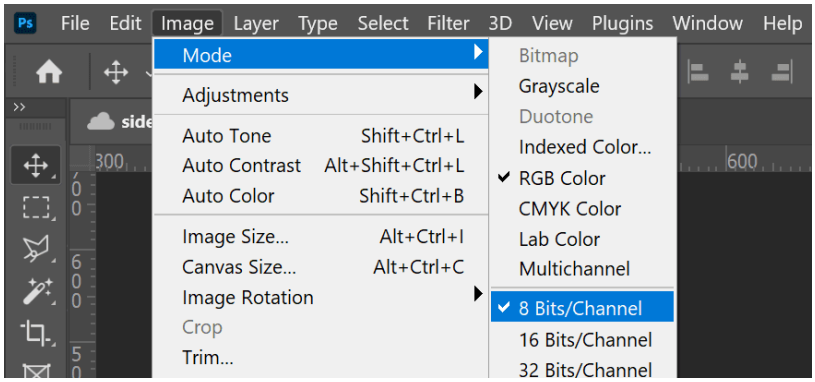

Figure B.12

2. Make sure the anti-alias is not selected in any of your selection tools. Figure B.13 and Figure B.14 shows the recommended setting.

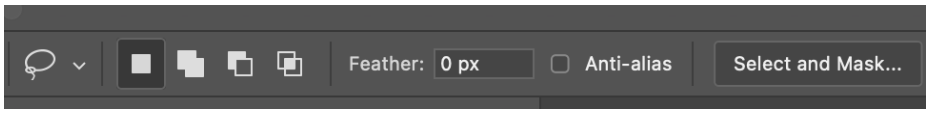

Figure B.13

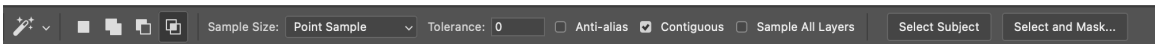

Figure B.14

3. Unless you know how to handle the color ranges created by Object Selection and Quick Selection tools, do not use them in your work since they can result in the selected area getting filled with a range of colors instead of one solid color you wanted to replace as shown in the image below Figure B.15.

Tips:

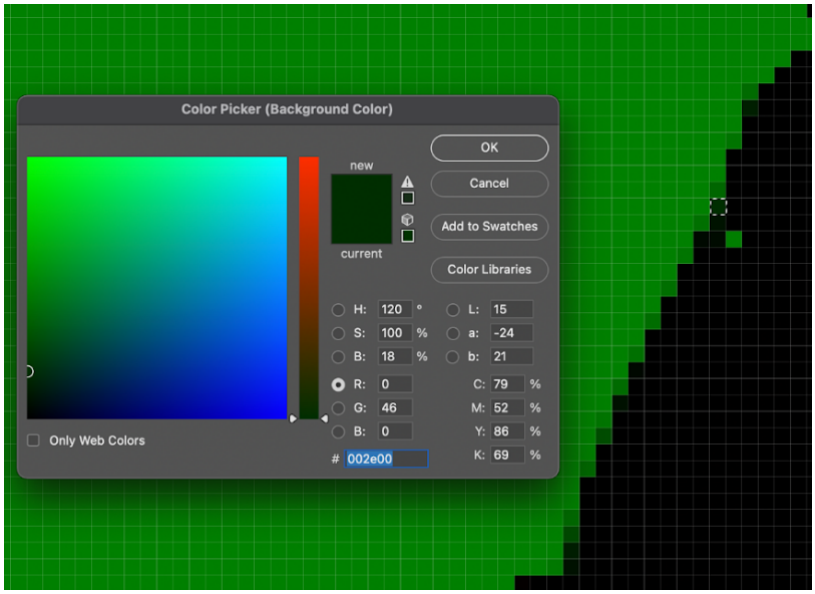

Figure B.15

1. Run the copy half action on each file to see the actual state of the annotation. The annotation label may look correct but when you do the overlay, you may find significant differences between the image and annotation.

2. When you are happy with the result (last step) you should delete the overlay layer and save as png.

3. Keep a backup on the cloud (google drive, dropbox, etc.)

4. Do not change the name/format of the files

APPENDIX C

KEY TECHNICAL CONCEPTS AND TERMS IN THIS DISSERTATION

Most of the definitions are taken from (GoogleDevelopers, n.d.), except for the ones identified by a star (*).

1. Active Learning: A training approach in which the algorithm chooses some of the data it learns from. Particularly valuable when labeled examples are scarce or expensive to obtain. Instead of blindly seeking a diverse range of labeled examples, an active learning algorithm selectively seeks the particular range of examples it needs for learning.

2. Attention: Any of a wide range of neural network architecture mechanisms that aggregate information from a set of inputs in a data-dependent manner.

3. Batch: The set of examples used in one iteration (that is, one gradient update) of model training.

4. Batch Size: The number of examples in a batch. Batch size is usually fixed during training and inference.

5. Class: One of a set of enumerated target values for a label. In a multi-class classification model that identifies dog breeds, the classes would be poodle, beagle, pug, and so on.

6. Classification Model: A type of machine learning model for distinguishing among two or more discrete classes. E.g. a natural language processing classification model could determine whether an input sentence was in French, Spanish, or Italian.

7. Computer Vision*: A field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs.

8. Confusion Matrix An NxN table that summarizes how successful a classification model's predictions were; that is, the correlation between the label and the model's classification.

9. Deep Model: A type of neural network containing multiple hidden layers.

10. Epoch: A full training pass over the entire dataset such that each example has been seen once. Thus, an epoch represents N/batch size training iterations, where N is the total number of examples.

11. Feature: An input variable used in making predictions.

12. Fine Tuning: Perform a secondary optimization to adjust the parameters of an already trained model to fit a new problem. Fine tuning often refers to refitting the weights of a trained unsupervised model to a supervised model

13. Framework : A basic structure underlying a system, concept, or text

14. GitHub*: GitHub, Inc. is a provider of Internet hosting for software development and version control.

15. Ground Truth The correct answer. Reality. Since reality is often subjective, expert annotators(raters) typically are the proxy for ground truth.

16. Holdout Data Examples intentionally not used ("held out") during training. Holdout data helps evaluate your model's ability to generalize to data other than the data it was trained on.

17. HRNet* High-Resolution Network is a general purpose convolutional neural network for tasks like semantic segmentation, object detection and image classification.

18. Hyperparameter: The ”knobs” that you tweak during successive runs of training a model.

19. Image Recognition: A process that classifies object(s), pattern(s), or concept(s) in an image. Image recognition is also known as image classification.

20. Inference: In machine learning, often refers to the process of making predictions by applying the trained model to unlabeled examples. In statistics, inference refers to the process of fitting the parameters of a distribution conditioned on some observed data.

21. Inference In machine learning, often refers to the process of making predictions by applying the trained model to unlabeled examples. In statistics, inference refers to the process of fitting the parameters of a distribution conditioned on some observed data.

22. Intersection over Onion (IoU) The intersection of two sets divided by their union. In machine-learning image-detection tasks, IoU is used to measure the accuracy of the model’s predicted bounding box with respect to the ground-truth bounding box. In this case, the IoU for the two boxes is the ratio between the overlapping area and the total area, and its value ranges from 0 (no overlap of predicted bounding box and ground-truth bounding box) to 1 (predicted bounding box and ground-truth bounding box have the exact same coordinates).

23. Iteration A single update of a model’s weights during training. An iteration consists of computing the gradients of the parameters with respect to the loss on a single batch of data.

24. Label In supervised learning, the ”answer” or ”result” portion of an example. Each example in a labeled dataset consists of one or more features and a label.

25. Learning Rate A scalar used to train a model via gradient descent. During each iteration, the gradient descent algorithm multiplies the learning rate by the gradient. The resulting product is called the gradient step.

26. LiDAR* Light Detection and Ranging, is a remote sensing method that uses light in the form of a pulsed laser to measure ranges

27. Loss A measure of how far a model's predictions are from its label. Or, to phrase it more pessimistically, a measure of how bad the model is. To determine this value, a model must define a loss function.

28. Machine Leaning(ML) A program or system that builds (trains) a predictive model from input data. The system uses the learned model to make useful predictions from new (never-before-seen) data drawn from the same distribution as the one used to train the model. Machine learning also refers to the field of study concerned with these programs or systems.

29. Model: The representation of what a machine learning system has learned from the training data.

30. Model training: is the phase in the data science development life cycle where practitioners try to fit the best combination of weights and bias to a machine learning algorithm to minimize a loss function over the prediction range.

31. Neural Network: A model that, taking inspiration from the brain, is composed of layers (at least one of which is hidden) consisting of simple connected units or neurons followed by nonlinearities.

32. OSMnx*: OSMnx is a Python package that lets you download geospatial data from OpenStreetMap and model, project, visualize, and analyze real-world street networks and any other geospatial geometries.

33. Overfitting: Reusing the examples of a minority class in a class-imbalanced dataset in order to create a more balanced training set.

34. Parameter: A variable of a model that the machine learning system trains on its own. For example, weights are parameters whose values the machine learning system gradually learns through successive training iterations. Contrast with hyperparameter.

35. Precision: A metric for classification models. Precision identifies the frequency with which a model was correct when predicting the positive class. That is: precision= True Positive ÷ (True Positive - False Positive)

36. Prediction: A model's output when provided with an input example.

37. Prediction bias: A value indicating how far apart the average of predictions is from the average of labels in the dataset. Not to be confused with the bias term in machine learning models or with bias in ethics and fairness.

38. Pre-trained Model: Models or model components (such as embeddings) that have already been trained. Sometimes, you'll feed pre-trained embeddings into a neural network. Other times, your model will train the embeddings itself rather than rely on the pre-trained embeddings.

39. Python (Py)* An interpreted high-level general-purpose programming language.

40. Recall: A metric for classification models that answers the following question: Out of all the possible positive labels, how many did the model correctly identify? That is: Recall = True Positive ÷ (True Positive - False Negative)

41. Self-supervised Learning: A family of techniques for converting an unsupervised machine learning problem into a supervised machine learning problem by creating surrogate labels from unlabeled examples. Self-supervised training is a semi-supervised learning approach.

42. Semantic Segmentation* A computer vision technique to do pixel-wise classification for images so every pixel will be assigned a class.

43. Semi-supervised Learning: Training a model on data where some of the training examples have labels but others don't. One technique for semi-supervised learning is to infer labels for the unlabeled examples, and then to train on the inferred labels to create a new model. Semi-supervised learning can be useful if labels are expensive to obtain but unlabeled examples are plentiful.

44. Supervised Machine Learning: Training a model from input data and its corresponding labels. Supervised machine learning is analogous to a student learning a subject by studying a set of questions and their corresponding answers. After mastering the mapping between questions and answers, the student can then provide answers to new (never-before-seen) questions on the same topic.

45. Support Vector Machines: are a set of supervised learning methods used for classification, regression and outliers detection.

46. Test Set: The subset of the dataset that you use to test your model after the model has gone through initial vetting by the validation set. Contrast with "training set" and "validation set".

47. Training: The process of determining the ideal parameters comprising a model.

48. Training Set/Training Data: The subset of the dataset used to train a model. Contrast with "validation set" and "test set".

49. Unsupervised Machine Learning: Training a model to find patterns in a dataset, typically an unlabeled dataset. The most common use of unsupervised machine learning is to cluster data into groups of similar examples. The resulting clusters can become an input to other machine learning algorithms.

50. Validation Set: A subset of the dataset—disjoint from the training set—used in validation. Contrast with “training set” and “test set”.

51. Validation: A process used, as part of training, to evaluate the quality of a machine learning model using the validation set. Because the validation set is disjoint from the training set, validation helps ensure that the model’s performance generalizes beyond the training set.

52. Weight: A coefficient for a feature in a linear model, or an edge in a deep network. The goal of training a linear model is to determine the ideal weight for each feature. If a weight is 0, then its corresponding feature does not contribute to the model.

REFERENCES

Agathangelidis, I., Cartalis, C., & Santamouris, M. (2020). Urban morphological controls on surface thermal dynamics: A comparative assessment of major european cities with a focus on athens, greece. *Climate*, *8*(11), 131.

Aghaabbasi, M., Moeinaddini, M., Shah, M. Z., Asadi-Shekari, Z., & Kermani, M. A. (2018). Evaluating the capability of walkability audit tools for assessing sidewalks. *Sustainable Cities and Society*, *37*, 475–484.

Ahn, J., & Kwak, S. (2018). Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4981–4990.

Ai, C., & Tsai, Y. (2016). Automated sidewalk assessment method for americans with disabilities act compliance using three-dimensional mobile lidar. *Transportation Research Record: Journal of the Transportation Research Board*, (2542), 25–32.

Akbari, H., Menon, S., & Rosenfeld, A. (2009). Global cooling: Increasing world-wide urban albedos to offset co 2. *Climatic change*, *94*(3), 275–286.

Akbari, H., & Rose, L. S. (2008). Urban surfaces and heat island mitigation potentials. *Journal of the Human-environment System*, *11*(2), 85–101.

Aljuboori, M. (2014). Development and verification of an intelligent sensor system for roadway and bridge surface condition assessments. *Thesis and Dissertation , University of Wisconsin-Milwaukee*.

Amati, M., & Taylor, L. (2010). From green belts to green infrastructure. *Planning Practice & Research*, *25*(2), 143–155.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google Street View: Capturing the world at street level. *Computer*, *43*(6), 32–38.

Archipedia New England. (2019). Historic pavings and sidewalks in New England. http://www.archipedianewengland.org/1600-1699/historic-paving-and-sidewalks-in-new-england/

Ariffin, R. N. R., & Zahari, R. K. (2013). Perceptions of the urban walking environments. *Procedia-Social and Behavioral Sciences*, *105*, 589–597.

Arnold Jr, C. L., & Gibbons, C. J. (1996). Impervious surface coverage: The emergence of a key environmental indicator. *Journal of the American planning Association*, *62*(2), 243–258.

Asadi-Shekari, Z., Moeinaddini, M., & Shah, M. Z. (2015). Pedestrian safety index for evaluating street facilities in urban areas. *Safety Science*, *74*, 1–14.

Asadi-Shekari, Z., Moeinaddini, M., & Zaly Shah, M. (2013). Disabled pedestrian level of service method for evaluating and promoting inclusive walking facilities on urban streets. *Journal of Transportation Engineering*, *139*(2), 181–192.

Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., et al. (2021). Big self-supervised models advance medical image classification. *arXiv preprint arXiv:2101.05224*.

Baker, C. D., & Hou, Q. (2019). *Improving pedestrian infrastructure inventory in massachusetts using mobile lidar* (tech. rep.). Massachusetts Department of Transportation (Mass DOT).

Balado, J., Dıaz-Vilariño, L., Arias, P., & González-Jorge, H. (2018). Automatic classification of urban ground elements from mobile laser scanning data. *Automation in Construction*, *86*, 226–239.

Balali, V., Rad, A. A., & Golparvar-Fard, M. (2015). Detection, classification, and mapping of us traffic signs using google street view images for roadway inventory management. *Visualization in Engineering*, *3*(1), 15.

Bastani, F., He, S., Abbar, S., Alizadeh, M., Balakrishnan, H., Chawla, S., Madden, S., & DeWitt, D. (2018). Roadtracer: Automatic extraction of road networks from aerial images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4720–4728.

Basu, R., Sevtsuk, A., & Li, X. (2022). How do street attributes affect willingness-to-walk? city-wide pedestrian route choice analysis using big data from boston and san francisco. [Upcoming]. *Transportation Research A*.

Bell, C. D., Tague, C. L., & McMillan, S. K. (2019). Modeling runoff and nitrogen loads from a watershed at different levels of impervious surface coverage and connectivity to storm water control measures. *Water Resources Research*, *55*(4), 2690–2707.

Bise, R. D., Rodgers III, J. C., Maguigan, M. A., Beaulieu, B., Keith, W., Maguigan, C. L., & Meng, Q. (2018). Sidewalks as measures of infrastructure inequities. *Southeastern Geographer*, *58*(1), 39–57.

Bloodgood, M., & Vijay-Shanker, K. (2014). A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping. *arXiv preprint arXiv:1409.5165*.

Bloomberg, M. R., & Burden, A. (2006). New york city pedestrian level of service study phase i. *NYC DCP, Transportation Division*.

Boccalatte, A., Thebault, M., Ménézo, C., Ramousse, J., & Fossa, M. (2022). Evaluating the impact of urban morphology on rooftop solar radiation: A new city-scale approach based on geneva gis data. *Energy and Buildings*, *260*, 111919.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*.

Boeing, G. (2020). The right tools for the job: The case for spatial science tool-building. *Transactions in GIS*.

Boone-Heinonen, J., Guilkey, D. K., Evenson, K. R., & Gordon-Larsen, P. (2010). Residential self-selection bias in the estimation of built environment effects on physical activity between adolescence and young adulthood. *International Journal of Behavioral Nutrition and Physical Activity*, *7*(1), 70.

Boston PWD. (2014). Boston Sidewalk Inventory. https : / / data . boston . gov / dataset / sidewalk-inventory

Brandt, J. W., & Algazi, V. R. (1992). Continuous skeleton computation by voronoi diagram. *CVGIP: Image understanding*, *55*(3), 329–338.

Brenner, A. B., & Clarke, P. J. (2019). Difficulty and independence in shopping among older americans: More than just leaving the house. *Disability and rehabilitation*, *41*(2), 191–200.

Brezina, T., Graser, A., & Leth, U. (2017). Geometric methods for estimating representative sidewalk widths applied to vienna's streetscape surfaces database. *Journal of Geographical Systems*, *19*(2), 157–174.

Cain, K. L., Millstein, R. A., & Geremia, C. (2012). Microscale audit of pedestrian streetscapes (maps): Data collection & scoring manual. *University California San Diego. Available for download at: http://sallis. ucsd. edu/Documents/Measures_documents/MAPS% 20Manual_v1_010713. pdf (accessed 08.08. 13.)*

Cambra, P. J., Gonçalves, A., & Moura, F. (2019). The digital pedestrian network in complex urban contexts: A primer discussion on typological specifications. *Finisterra*, *54*(110), 155–170.

Cambridge GIS. (2018a). Cambridge Sidewalk.

Cambridge GIS. (2018b). Pavement markings.

Cambridge GIS. (2018c). Public Footpaths.

Cambridge GIS. (2018d). Roads.

Campbell, A., Both, A., & Sun, Q. C. (2019). Detecting and mapping traffic signs from google street view images using deep learning and gis. *Computers, Environment and Urban Systems*, *77*, 101350.

Cao, W., Liu, Q., & He, Z. (2020). Review of pavement defect detection methods. *IEEE Access*, *8*, 14531–14544.

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

Casanova, A., Pinheiro, P. O., Rostamzadeh, N., & Pal, C. J. (2020). Reinforced active learning for image segmentation. *arXiv preprint arXiv:2002.06583*.

Cervero, R. (1998). *The transit metropolis: A global inquiry*. Island press.

Charreire, H., Mackenbach, J. D., Ouasti, M., Lakerveld, J., Compernolle, S., Ben-Rebah, M., McKee, M., Brug, J., Rutter, H., & Oppert, J.-M. (2014). Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: A systematic review (the spotlight project). *Health & place*, *25*, 1–9.

Chen, L.-C., Yang, Y., Wang, J., Xu, W., & Yuille, A. L. (2016). Attention to scale: Scale-aware semantic image segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3640–3649.

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. (2020). Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*.

Chen, X., & Zhang, Y. (2017). Impacts of urban surface characteristics on spatiotemporal pattern of land surface temperature in kunming of china. *Sustainable Cities and Society*, *32*, 87–99.

Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., & Sun, J. (2018). Cascaded pyramid network for multi-person pose estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7103–7112.

Chin, G. K., Van Niel, K. P., Giles-Corti, B., & Knuiman, M. (2008). Accessibility and connectivity in physical activity studies: The impact of missing pedestrian data. *Preventive medicine*, *46*(1), 41–45.

Chippendale, T. (2020). Outdoor falls prevention strategy use and neighborhood walkability among naturally occurring retirement community residents. *Health Education & Behavior*, 1090198120980358.

Chippendale, T., & Boltz, M. (2015). The neighborhood environment: Perceived fall risk, resources, and strategies for fall prevention. *The Gerontologist*, *55*(4), 575–583.

Chithra, S., Nair, M. H., Amarnath, A., & Anjana, N. (2015). Impacts of impervious surfaces on the environment. *International Journal of Engineering Science Invention*, *4*(5), 27–31.

Christensen, K. M., Holt, J. M., & Wilson, J. F. (2010). Effects of perceived neighborhood characteristics and use of community facilities on physical activity of adults with and without disabilities. *Preventing chronic disease*, *7*, A105.

Clarke, P., Ailshire, J. A., Bader, M., Morenoff, J. D., & House, J. S. (2008). Mobility disability and the urban built environment. *American journal of epidemiology*, *168*(5), 506–513.

Clifton, K. J., Smith, A. D. L., & Rodriguez, D. (2007). The development and testing of an audit for the pedestrian environment. *Landscape and Urban Planning*, *80*(1), 95–110.

Colling, P., Roese-Koerner, L., Gottschalk, H., & Rottmann, M. (2020). Metabox+: A new region based active learning method for semantic segmentation using priority maps. *arXiv preprint arXiv:2010.01884*.

Cooper, R., Wolf, E., Fitzgerald, S., Dobson, A., Ammer, W., & Smith, D. (2003). Interaction of wheelchairs and segmental pavement surfaces. *Proceedings*.

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

Cottrill, C. D., Pereira, F. C., Zhao, F., Dias, I. F., Lim, H. B., Ben-Akiva, M. E., & Zegras, P. C. (2013). Future mobility survey: Experience in developing a smartphone-based travel survey in singapore. *Transportation Research Record*, *2354*(1), 59–67.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21–27.

Credit, K. N. (2018). *Economic development for the 21 st century: How proximity to transit and walkability influence business creation and performance*. Michigan State University.

Crews, D. E., & Zavotka, S. (2006). Aging, disability, and frailty: Implications for universal design. *Journal of physiological anthropology*, *25*(1), 113–118.

Cura, R., Perret, J., & Paparoditis, N. (2018). A state of the art of urban reconstruction: Street, street network, vegetation, urban feature. *arXiv preprint arXiv:1803.04332*.

Curtis, J. W., Curtis, A., Mapes, J., Szell, A. B., & Cinderich, A. (2013). Using google street view for systematic observation of the built environment: Analysis of spatio-temporal instability of imagery dates. *International journal of health geographics*, *12*(1), 1–10.

DC GIS. (2019a). Roads 2019.

DC GIS. (2019b). Sidewalks 2019.

DC GIS. (2020). Aerial photography (orthophoto sid) - 2019.

Deitz, S. (2021). Free movement: Enhancing open data to facilitate independent travel for persons with disabilities.

Deitz, S., Lobben, A., & Alferez, A. (2021). Squeaky wheels: Missing data, disability, and power in the smart city. *Big Data & Society*, *8*(2), 20539517211047735. https://doi.org/10.1177/20539517211047735

Department of Justice. (2010). 2010 ADA standards for accessible design. https://www.ada.gov/2010ADAstandards%5C%5Findex.htm

Dibble, J., Prelorendjos, A., Romice, O., Zanella, M., Strano, E., Pagel, M., & Porta, S. (2019). On the origin of spaces: Morphometric foundations of urban form evolution. *Environment and Planning B: Urban Analytics and City Science*, *46*(4), 707–730. https://doi.org/10.1177/2399808317725075

Diez Roux, A. V. (2003). Residential environments and cardiovascular risk. *Journal of Urban Health*, *80*(4), 569–589.

Ding, H., Jiang, X., Shuai, B., Liu, A. Q., & Wang, G. (2018). Context contrasted feature and gated multi-scale aggregation for scene segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2393–2402.

Doraiswamy, H., Tzirita Zacharatou, E., Miranda, F., Lage, M., Ailamaki, A., Silva, C. T., & Freire, J. (2018). Interactive visual exploration of spatio-temporal urban data sets

using urbane. *Proceedings of the 2018 International Conference on Management of Data*, 1693–1696. https://doi.org/10.1145/3183713.3193559

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, *10*(2), 112–122.

Druschel, S. J. (2020). *Hot shots for cold climates–evaluating treatment of the hardest icy spots* (tech. rep.).

Du, H., Cai, W., Xu, Y., Wang, Z., Wang, Y., & Cai, Y. (2017). Quantifying the cool island effects of urban green spaces using remote sensing data. *Urban Forestry & Urban Greening*, *27*, 24–31.

Eisenberg, Y., Heider, A., Gould, R., & Jones, R. (2020a). Are communities in the united states planning for pedestrians with disabilities? findings from a systematic evaluation of local government barrier removal plans. *Cities*, *102*, 102720. https://doi.org/https://doi.org/10.1016/j.cities.2020.102720

Eisenberg, Y., Heider, A., Gould, R., & Jones, R. (2020b). Are communities in the united states planning for pedestrians with disabilities? findings from a systematic evaluation of local government barrier removal plans. *Cities*, *102*, 102720.

Eisenberg, Y., Vanderbom, K. A., & Vasudevan, V. (2017). Does the built environment moderate the relationship between having a disability and lower levels of physical activity? a systematic review. *Preventive Medicine*, *95*, S75–S84. https://doi.org/https://doi.org/10.1016/j.ypmed.2016.07.019

Ellis, G., Hunter, R., Tully, M. A., Donnelly, M., Kelleher, L., & Kee, F. (2016). Connectivity and physical activity: Using footpath networks to measure the walkability of built environments. *Environment and Planning B: Planning and Design*, *43*(1), 130–151.

Emery, J., Crump, C., & Bors, P. (2003). Reliability and validity of two instruments designed to assess the walking and bicycling suitability of sidewalks and roads. *American Journal of Health Promotion*, *18*(1), 38–46.

EPA. (2021). Sources of greenhouse gas emissions.

Erath, A. L., van Eggermond, M. A., Ordóñez Medina, S. A., & Axhausen, K. W. (2015). Modelling for walkability: Understanding pedestrians' preferences in singapore. *14th International Conference on Travel Behavior Research (IATBR 2015)*.

Ess, A., Mueller, T., Grabner, H., & Van Gool, L. (2009). Segmentation-based urban traffic scene understanding. *BMVC*, *1*, 2.

Estoque, R. C., Murayama, Y., & Myint, S. W. (2017). Effects of landscape composition and pattern on land surface temperature: An urban heat island study in the megacities of southeast asia. *Science of the Total Environment*, *577*, 349–359.

Etten, A. V. (2020). City-scale road extraction from satellite imagery v2: Road speeds and travel times. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1786–1795.

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design*, *14*(1), 65–84.

Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and measuring urban design qualities related to walkability. *Journal of Physical Activity and Health*, *3*(s1), S223–S240.

Feliciotti, A. (2018). *Resilience and urban design* (Doctoral dissertation). University of Strathclyde.

Ferreira, A., & Giraldi, G. (2017). Convolutional neural network approaches to granite tiles classification. *Expert Systems with Applications*, *84*, 1–11.

Ferreira, & da Penha Sanches, S. (2007). Proposal of a sidewalk accessibility index. *Journal of Urban and Environmental Engineering*, *1*(1), 1–9.

Fix, E. (1985). *Discriminatory analysis: Nonparametric discrimination, consistency properties* (Vol. 1). USAF school of Aviation Medicine.

Fleischmann, M. (2019). Momepy: Urban morphology measuring toolkit. *Journal of Open Source Software*, *4*(43), 1807.

Fleischmann, M., Romice, O., & Porta, S. (2021). Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science*, *48*(8), 2133–2150.

Forsyth, A., Hearst, M., Oake, J. M., & Schmitz, K. H. (2008). Design and destinations: Factors influencing walking and total physical activity. *Urban Studies*, *45*(9), 1973–1996.

Frackelton, A., Grossman, A., Palinginis, E., Castrillon, F., Elango, V., & Guensler, R. (2013). Measuring walkability: Development of an automated sidewalk quality assessment tool. *Suburban Sustainability*, *1*(1), 4.

Frank, L. D., & Engelke, P. O. (2001). The built environment and human activity patterns: Exploring the impacts of urban form on public health. *Journal of planning literature*, *16*(2), 202–218.

Froehlich, J. E., Brock, A. M., Caspi, A., Guerreiro, J., Hara, K., Kirkham, R., Schöning, J., & Tannert, B. (2019). Grand challenges in accessible maps. *Interactions*, *26*, 78–81. https://doi.org/10.1145/3301657

Froehlich, J. E., Saugstad, M., Saha, M., & Johnson, M. (2022). Towards mapping and assessing sidewalk accessibility across sociocultural and geographic contexts. *arXiv preprint arXiv:2207.13626*.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3146–3154.

Gahegan, M. (2018). Our gis is too small. *The Canadian Geographer/Le Géographe Canadien*, *62*(1), 15–26.

Gal, Y., Islam, R., & Ghahramani, Z. (2017). Deep bayesian active learning with image data. *International Conference on Machine Learning*, 1183–1192.

Gehl, J. (2011). *Life between buildings: Using public space*. Island Press.

Gehl, J. (2013). *Cities for people*. Island press.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361.

Gil, J., Beirão, J. N., Montenegro, N., & Duarte, J. P. (2012). On the discovery of urban typologies: Data mining the many dimensions of urban form. *Urban morphology*, *16*(1), 27.

Gillies, S., Ward, B., Petersen, A., et al. (2013). Rasterio: Geospatial raster i/o for python programmers.

Gissin, D., & Shalev-Shwartz, S. (2019). Discriminative active learning. *arXiv preprint arXiv:1907.06347*.

Glaeser, E. (2010). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin Press.

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

Golestaneh, S. A., & Kitani, K. M. (2020). Importance of self-consistency in active learning for semantic segmentation. *arXiv preprint arXiv:2008.01860*.

GoogleDevelopers. (n.d.). Machine learning glossary.

Grasser, G., Van Dyck, D., Titze, S., & Stronegger, W. (2013). Objectively measured walkability and active transport and weight-related outcomes in adults: A systematic review. *International journal of public health*, *58*(4), 615–625.

Greenwald, M. J., & Boarnet, M. G. (2001). Built environment as determinant of walking behavior: Analyzing nonwork pedestrian travel in portland, oregon. *Transportation research record*, *1780*(1), 33–41.

Griew, P., Hillsdon, M., Foster, C., Coombes, E., Jones, A., & Wilkinson, P. (2013). Developing and testing a street audit tool using google street view to measure environmental supportiveness for physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *10*(1), 1–7.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo, Z. D., Azar, M. G., et al. (2020). Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*.

Guan, J., Yang, X., Ding, L., Cheng, X., Lee, V. C., & Jin, C. (2021). Automated pixel-level pavement distress detection based on stereo vision and deep learning. *Automation in Construction*, *129*, 103788.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *International Conference on Machine Learning*, 1321–1330.

Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. *Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, 47–57.

Haans, A., & De Kort, Y. A. (2012). Light distribution in dynamic street lighting: Two experimental studies on its effects on perceived safety, prospect, concealment, and escape. *Journal of Environmental Psychology*, *32*(4), 342–352.

Hara, K., Azenkot, S., Campbell, M., Bennett, C. L., Le, V., Pannella, S., Moore, R., Minckler, K., Ng, R. H., & Froehlich, J. E. (2013). Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street

view. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 1–8. https://doi.org/10.1145/2513383.2513448

Hara, K., Azenkot, S., Campbell, M., Bennett, C. L., Le, V., Pannella, S., Moore, R., Minckler, K., Ng, R. H., & Froehlich, J. E. (2015). Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis. *ACM Transactions on Accessible Computing*, *6*, 1–23. https://doi.org/10.1145/2717513

Hara, K., Le, V., & Froehlich, J. (2013). Combining crowdsourcing and google street view to identify street-level accessibility problems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 631–640. https://doi.org/10.1145/2470654.2470744

Hara, K., Le, V., & Froehlich, J. (2012). A feasibility study of crowdsourcing and google street view to determine sidewalk accessibility. *Extended Abstracts of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*, 273–274. https://doi.org/10.1145/2384916.2384989

Hara, K., Sun, J., Moore, R., Jacobs, D., & Froehlich, J. (2014). Tohme: Detecting curb ramps in google street view using crowdsourcing, computer vision, and machine learning. *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 189–204.

Harris, F., Yang, H.-Y., & Sanford, J. (2015). Physical environmental barriers to community mobility in older and younger wheelchair users. *Topics in Geriatric Rehabilitation*, *31*, 42–51. https://doi.org/10.1097/TGR.0000000000000043

Haunert, J.-H., & Sester, M. (2008). Area collapse and road centerlines based on straight skeletons. *GeoInformatica*, *12*(2), 169–191.

He, J., Deng, Z., & Qiao, Y. (2019). Dynamic multi-scale filters for semantic segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3562–3572.

He, J., Deng, Z., Zhou, L., Wang, Y., & Qiao, Y. (2019). Adaptive pyramid context network for semantic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7519–7528.

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9729–9738.

He, L., Chao, Y., Suzuki, K., & Wu, K. (2009). Fast connected-component labeling. *Pattern recognition*, *42*(9), 1977–1987.

Horváth, E., Pozna, C., & Unger, M. (2022). Real-time lidar-based urban road and sidewalk detection for autonomous vehicles. *Sensors*, *22*(1), 194.

Hosseini, M., Araujo, I. B., Yazdanpanah, H., Tokuda, E. K., Miranda, F., Silva, C. T., & Cesar Jr, R. M. (2021). Sidewalk measurements from satellite images: Preliminary findings. *Spatial Data Science Symposium 2021 Short Paper Proceedings*. https://doi.org/https://doi.org/10.25436/E2QG6F

Hosseini, M., Miranda, F., Lin, J., & Silva, C. T. (2022). Citysurfaces: City-scale semantic segmentation of sidewalk materials. *Sustainable Cities and Society*, 103630.

Hosseini, M., Saugstad, M., Miranda, F., Sevtsuk, A., Silva, C. T., & Froehlich, J. E. (2022). Towards global-scale crowd+ ai techniques to map and assess sidewalks for people with disabilities. *arXiv preprint arXiv:2206.13677*.

Hosseini, M., Sevtsuk, A., Miranda, F., Cesar Jr, R., & Silva, C. (n.d.). Mapping the walk: A scalable computer vision approach for generating sidewalk network datasets from aerial imagery. *Available at SSRN 4086624*.

Hou, Q., & Ai, C. (2020). A network-level sidewalk inventory method using mobile lidar and deep learning. *Transportation Research Part C: Emerging Technologies*, *119*, 102772.

Houle, K. M. (2008). Winter performance assessment of permeable pavements: A comparative study of porous asphalt, pervious concrete, and conventional asphalt in a northern climate.

Huang, Q., Xia, C., Wu, C., Li, S., Wang, Y., Song, Y., & Kuo, C.-C. J. (2017). Semantic segmentation with reverse attention. *arXiv preprint arXiv:1707.06426*.

Huang, S.-J., Jin, R., & Zhou, Z.-H. (2010). Active learning by querying informative and representative examples. *Advances in neural information processing systems*, *23*, 892–900.

Iglovikov, V., Mushinskiy, S., & Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint arXiv:1706.06169*.

Jacobs, J. (1961). *The death and life of american cities*. Random House, New York.

Jain, S., & Gruteser, M. (2018). Recognizing textures with mobile cameras for pedestrian safety applications. *IEEE Transactions on Mobile Computing*, *18*(8), 1911–1923.

Jaskiewicz, F. (2000). Pedestrian level of service based on trip quality. *Transportation Research Circular, TRB*.

Jenkins, M. D., Carr, T. A., Iglesias, M. I., Buggy, T., & Morison, G. (2018). A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks. *2018 26th European Signal Processing Conference (EUSIPCO)*, 2120–2124.

Jordahl, K. (2014). Geopandas: Python tools for geographic data.

Joshi, P., Leitão, J. P., Maurer, M., & Bach, P. M. (2021). Not all suds are created equal: Impact of different approaches on combined sewer overflows. *Water Research*, *191*, 116780.

Kamel, I., & Faloutsos, C. (1993). *Hilbert r-tree: An improved r-tree using fractals* (tech. rep.).

Kang, B., Lee, S., & Zou, S. (2021). Developing sidewalk inventory data using street view images. *Sensors*, *21*(9), 3300.

Kang, H., & Han, J. (2020). Safexr: Alerting walking persons to obstacles in mobile xr environments. *The Visual Computer*, *36*(10), 2065–2077.

Karimi, H. A., & Kasemsuppakorn, P. (2013). Pedestrian network map generation approaches and recommendation. *International Journal of Geographical Information Science*, *27*(5), 947–962.

Kasarla, T., Nagendar, G., Hegde, G. M., Balasubramanian, V., & Jawahar, C. (2019). Region-based active learning for efficient labeling in semantic segmentation. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1109–1117.

Kasemsuppakorn, P., & Karimi, H. A. (2008). Data requirements and a spatial database for personalized wheelchair navigation. *Proceedings of the 2nd International Convention on Rehabilitation Engineering Assistive Technology*, 31–34.

Kasemsuppakorn, P., & Karimi, H. A. (2013). Pedestrian network extraction from fused aerial imagery (orthoimages) and laser imagery (lidar). *Photogrammetric Engineering & Remote Sensing*, *79*(4), 369–379.

Katzmarzyk, P. T., Denstel, K. D., Beals, K., Carlson, J., Crouter, S. E., McKenzie, T. L., Pate, R. R., Sisson, S. B., Staiano, A. E., Stanish, H., Ward, D. S., Whitt-Glover, M., & Wright, C. (2018). Results from the united states 2018 report card on physical activity for children and youth. *Journal of Physical Activity and Health*, *15*(S2), S422–S424.

Kelly, C. M., Wilson, J. S., Baker, E. A., Miller, D. K., & Schootman, M. (2013). Using google street view to audit the built environment: Inter-rater reliability results. *Annals of Behavioral Medicine*, *45*(suppl_1), S108–S112.

Kharazi, B. A., & Behzadan, A. H. (2021). Flood depth mapping in street photos with image processing and deep neural networks. *Computers, Environment and Urban Systems*, *88*, 101628.

Kim, J. H., Lee, S., Hipp, J. R., & Ki, D. (2021). Decoding urban landscapes: Google street view and measurement sensitivity. *Computers, Environment and Urban Systems*, *88*, 101626.

Kim, T., Hwang, I., Lee, H., Kim, H., Choi, W.-S., & Zhang, B.-T. (2020). Message passing adaptive resonance theory for online active semi-supervised learning. *arXiv preprint arXiv:2012.01227*.

Kontokosta, C. E. (2018). Urban informatics in the science and practice of planning. *Journal of Planning Education and Research*, 0739456X18793716.

Kopf, J., Chen, B., Szeliski, R., & Cohen, M. (2010). Street slide: Browsing street level imagery. *ACM Transactions on Graphics*, *29*(4), 96:1–96:8.

Krizek, K. J. (2003). Residential relocation and changes in urban travel: Does neighborhood-scale urban form matter? *Journal of the American Planning Association*, *69*(3), 265–281.

Kuo, W., Häne, C., Yuh, E., Mukherjee, P., & Malik, J. (2018). Cost-sensitive active learning for intracranial hemorrhage detection. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 715–723.

Lander, C., Wiehr, F., Herbig, N., Krüger, A., & Löchtefeld, M. (2017). Inferring landmarks for pedestrian navigation from mobile eye-tracking data and google street view. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2721–2729.

Landis, B. W., Vattikuti, V. R., Ottenberg, R. M., McLeod, D. S., & Guttenplan, M. (2001). Modeling the roadside walking environment: Pedestrian level of service. *Transportation research record*, *1773*(1), 82–88.

Lautso, K., & Murole, P. (1974). A study of pedestrian traffic in helsinki: Methods and results. *Traffic Engineering & Control*, *15*(9).

Law, S., Seresinhe, C. I., Shen, Y., & Gutierrez-Roig, M. (2018). Street-frontage-net: Urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, 1–27.

Lay, M., Metcalf, J., & Sharp, K. (2020). *Paving our ways: A history of the world's roads and pavements*. CRC Press.

Lee, B. J., Jang, T. Y., Wang, W., & Namgung, M. (2009). Design criteria for an urban sidewalk landscape considering emotional perception. *Journal of urban planning and development*, *135*(4), 133–140.

Lee, S., & Talen, E. (2014). Measuring walkability: A note on auditing methods. *Journal of Urban Design*, *19*(3), 368–388.

Lee, S. (2018). *The impact of the neighborhood environment on falls among older adults* (Doctoral dissertation).

Leinberger, C. B., & Lynch, P. (2014). Foot traffic ahead: Ranking walkable urbanism in america's largest metros. *Transportation Research Board*.

Lewandowicz, E., & Flisek, P. (2020). A method for generating the centerline of an elongated polygon on the example of a watercourse. *ISPRS International Journal of Geo-Information*, *9*(5), 304.

Li, A., Saha, M., Gupta, A., & Froehlich, J. E. (2018). Interactively modeling and visualizing neighborhood accessibility at scale: An initial study of washington dc. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 444–446. https://doi.org/10.1145/3234695.3241000

Li, H., Xiong, P., An, J., & Wang, L. (2018). Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*.

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

Li, X., Zhang, C., & Li, W. (2017). Building block level urban land-use information retrieval based on google street view images. *GIScience & Remote Sensing*, *54*(6), 819–835.

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

Li, X., Zhou, W., & Ouyang, Z. (2013). Relationship between land surface temperature and spatial pattern of greenspace: What are the effects of spatial resolution? *Landscape and Urban Planning*, *114*, 1–8.

Li, Z., Guan, R., Yu, Q., Chiang, Y.-Y., & Knoblock, C. A. (2021). Synthetic map generation to provide unlimited training data for historical map text detection. *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 17–26.

Liu, S., Higgs, C., Arundel, J., Boeing, G., Cerdera, N., Moctezuma, D., Cerin, E., Adlakha, D., Lowe, M., & Giles-Corti, B. (2021). A generalized framework for measuring pedestrian accessibility around the world using open data. *Geographical Analysis*.

Liu, W., Rabinovich, A., & Berg, A. C. (2015). Parsenet: Looking wider to see better. *arXiv preprint arXiv:1506.04579*.

Louch, H., Voros, K., & David, E. (2020). *Availability and use of pedestrian infrastructure data to support active transportation planning*.

Loutzenheiser, Felix. (2010). Boston Region's Pedestrian Transportation Plan. https://www.mapc.org/wp-content/uploads/2017/11/PedPlanFullReport.pdf

Ma, K., Hoai, M., & Samaras, D. (2017). Large-scale continual road inspection: Visual infrastructure assessment in the wild. *Proceedings of British Machine Vision Conference*.

Mackowiak, R., Lenz, P., Ghori, O., Diego, F., Lange, O., & Rother, C. (2018). Cereals-cost-effective region-based active learning for semantic segmentation. *arXiv preprint arXiv:1810.09726*.

Maghelal, P. K., & Capp, C. J. (2011). Walkability: A review of existing pedestrian indices. *Journal of the Urban & Regional Information Systems Association*, *23*(2).

Marshall, W. E., & Garrick, N. W. (2010). Effect of street network design on walking and biking. *Transportation Research Record*, *2198*(1), 103–115.

MassGIS. (2018). MassGIS Data: 2018 Aerial Imagery.

McCormack, G. R., McLaren, L., Salvo, G., & Blackstaffe, A. (2017). Changes in objectively-determined walkability and physical activity in adults: A quasi-longitudinal residential relocation study. *International journal of environmental research and public health*, *14*(5), 551.

McGarigal, K. (1995). *Fragstats: Spatial pattern analysis program for quantifying landscape structure* (Vol. 351). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

McKibbin, M. (2011). The influence of the built environment on mode choice–evidence from the journey to work in sydney. *34th Australasian Transport Research Forum (ATRF) Proceedings held on 28 - 30 September 2011 in Adelaide, Australia.*

Metropolitan Planning Council. (2021). Where the sidewalk ends: The state of municipal ADA transition planning for the public right-of-way in the Chicago region.

Millington, C., Thompson, C. W., Rowe, D., Aspinall, P., Fitzsimons, C., Nelson, N., & Mutrie, N. (2009). Development of the scottish walkability assessment tool (swat). *Health Place*, *15*(2), 474–481.

Miranda, F., Doraiswamy, H., Lage, M., Wilson, L., Hsieh, M., & Silva, C. T. (2019). Shadow accrual maps: Efficient accumulation of city-scale shadows over time. *IEEE Transactions on Visualization and Computer Graphics*, *25*(3), 1559–1574.

Miranda, F., Doraiswamy, H., Lage, M., Wilson, L., Hsieh, M., & Silva, C. T. (2020). Shadow Accrual Maps. https://github.com/VIDA-NYU/shadow-accrual-maps/

Miranda, F., Hosseini, M., Lage, M., Doraiswamy, H., Dove, G., & Silva, C. T. (2020). Urban mosaic: Visual exploration of streetscapes using large-scale image data. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15.

Mitchell, C. (2006a). Pedestrian mobility and safety: A key to independence for older people. *Topics in Geriatric Rehabilitation*, *22*, 45–52.

Mitchell, C. (2006b). Pedestrian mobility and safety: A key to independence for older people. *Topics in Geriatric Rehabilitation*, *22*(1), 45–52.

Monroy Licht, I. M. (2015). *Arc routing problems for road network maintenance* (Doctoral dissertation). École Polytechnique de Montréal.

Montoya-Zegarra, J. A., Wegner, J. D., Ladick, L., & Schindler, K. (2014). Mind the gap: Modeling local and global context in (road) networks. *German Conference on Pattern Recognition*, 212–223.

Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D., Teitler, J. O., Sheehan, D. M., Jack, D. W., & Rundle, A. G. (2016). Use of google street view to assess environmental contributions to pedestrian injury. *American journal of public health*, *106*(3), 462–469.

Moran, M. E. (2022). Where the crosswalk ends: Mapping crosswalk coverage via satellite imagery in san francisco. *Environment and Planning B: Urban Analytics and City Science*, 23998083221081530.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

Muench, S. T., Anderson, J., & Bevan, T. (2010). Greenroads: A sustainability rating system for roadways. *International Journal of Pavement Research & Technology*, *3*(5).

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 779–785.

Nations, U. (2020). *The new urban agenda*. United Nations Human Settlements Programme (UN-Habitat).

Neuhold, G., Ollmann, T., Rota Bulo, S., & Kontschieder, P. (2017). The mapillary vistas dataset for semantic understanding of street scenes. *Proceedings of the IEEE International Conference on Computer Vision*, 4990–4999.

Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. *European conference on computer vision*, 483–499.

Nickelson, J., Wang, A. R., Mitchell, Q. P., Hendricks, K., & Paschal, A. (2013). Inventory of the physical environment domains and subdomains measured by neighborhood audit tools: A systematic literature review. *Journal of Environmental Psychology*, *36*, 179–189.

Ning, H., Ye, X., Chen, Z., Liu, T., & Cao, T. (2022). Sidewalk extraction using aerial and street view images. *Environment and Planning B: Urban Analytics and City Science*, *49*(1), 7–22.

Nolte, M., Kister, N., & Maurer, M. (2018). Assessment of deep convolutional neural networks for road surface classification. *arXiv:1804.08872*.

Nwakaire, C. M., Onn, C. C., Yap, S. P., Yuen, C. W., & Onodagu, P. D. (2020). Urban heat island studies with emphasis on urban pavements; a review. *Sustainable Cities and Society*, 102476.

NYC DoITT. (2018). New york city planimetrics data.

NYC DOT. (2020). Street design manual.

NYC GIS. (2018). NYS Statewide Digital Orthoimagery Program.

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, *108*(455), 1–24.

Olaru, D., & Curtis, C. (2015). Designing tod precincts: Accessibility and travel patterns. *European Journal of Transport and Infrastructure Research*, *15*(1), 6–26.

Pathak, D., Krahenbuhl, P., & Darrell, T. (2015). Constrained convolutional neural networks for weakly supervised segmentation. *Proceedings of the IEEE international conference on computer vision*, 1796–1804.

Phillips, C. B., Engelberg, J. K., Geremia, C. M., Zhu, W., Kurka, J. M., Cain, K. L., Sallis, J. F., Conway, T. L., & Adams, M. A. (2017). Online versus in-person comparison of microscale audit of pedestrian streetscapes (maps) assessments: Reliability of alternate methods. *International journal of health geographics*, *16*(1), 27.

Pikora, T. J., Bull, F. C., Jamrozik, K., Knuiman, M., Giles-Corti, B., & Donovan, R. J. (2002). Developing a reliable audit instrument to measure the physical environment for physical activity. *American journal of preventive medicine*, *23*(3), 187–194.

Placematters and WalkDenver. (2014). Walkscope.

Pratt, R. H., Evans IV, J. E., Levinson, H. S., Turner, S. M., Jeng, C. Y., & Nabors, D. (2012). *Traveler response to transportation system changes. chapter 16-pedestrian and bicycle facilities* (tech. rep.).

Proulx, F. R., Zhang, Y., & Grembek, O. (2015). Database for active transportation infrastructure and volume. *Transportation research record*, *2527*(1), 99–106.

Qin, H., Curtin, K. M., & Rice, M. T. (2018). Pedestrian network repair with spatial optimization models and geocrowdsourced data. *GeoJournal*, *83*(2), 347–364.

Quackenbush, L. J. (2004). A review of techniques for extracting linear features from imagery. *Photogrammetric Engineering & Remote Sensing*, *70*(12), 1383–1392.

Ran, X., Xue, L., Zhang, Y., Liu, Z., Sang, X., & He, J. (2019). Rock classification from field image patches analyzed using a deep convolutional neural network. *Mathematics*, *7*(8), 755.

Retting, R. A., Ferguson, S. A., & McCartt, A. T. (2003). A review of evidence-based traffic engineering measures designed to reduce pedestrian–motor vehicle crashes. *American Journal of Public Health*, *93*(9), 1456–1463.

Rhoads, D., Solé-Ribalta, A., González, M. C., & Borge-Holthoefer, J. (2020). Planning for sustainable open streets in pandemic cities. *arXiv preprint arXiv:2009.12548*.

Rogers, S. H., Gardner, K. H., & Carlson, C. H. (2013). Social capital and walkability as social aspects of sustainability. *Sustainability*, *5*(8), 3473–3483.

Rosenfeld, A., & Pfaltz, J. L. (1966). Sequential operations in digital picture processing. *Journal of the ACM (JACM)*, *13*(4), 471–494.

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

Rzotkiewicz, A., Pearson, A. L., Dougherty, B. V., Shortridge, A., & Wilson, N. (2018). Systematic review of the use of google street view in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health & place*, *52*, 240–246.

Sachs, D. (2016). A Complete Map of Denver's Walking Network Is Now Within Reach.

Saha, M., Patil, S., Cho, E., Cheng, E. Y.-Y., Horng, C., Chauhan, D., Kangas, R., McGovern, R., Li, A., Heer, J., & Froehlich, J. E. (2022). Visualizing urban accessibility: Investigating multi-stakeholder perspectives through a map-based design probe study. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3491102.3517460

Saha, M., Saugstad, M., Maddali, H. T., Zeng, A., Holland, R., Bower, S., Dash, A., Chen, S., Li, A., Hara, K., & Froehlich, J. (2019). Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

Saha, P. K., Borgefors, G., & di Baja, G. S. (2016). A survey on skeletonization algorithms and their applications. *Pattern recognition letters*, *76*, 3–12.

Sallis, J. F., Floyd, M. F., Rodrıguez, D. A., & Saelens, B. E. (2012). Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*, *125*(5), 729–737.

Sampson, R. J. (2012). *Great american city: Chicago and the enduring neighborhood effect*. University of Chicago Press.

Santamouris, M. (2013). Using cool pavements as a mitigation strategy to fight urban heat island—a review of the actual developments. *Renewable and Sustainable Energy Reviews*, *26*, 224–240.

Santamouris, M., Synnefa, A., & Karlessi, T. (2011). Using advanced cool materials in the urban built environment to mitigate heat islands and improve thermal comfort conditions. *Solar Energy*, *85*(12), 3085–3102.

Schaeffer, K. H., & Sclar, E. (1980). *Access for all: Transportation and urban growth*. Columbia University Press.

Scheffer, T., Decomain, C., & Wrobel, S. (2001). Active hidden markov models for information extraction. *International Symposium on Intelligent Data Analysis*, 309–318.

Sener, O., & Savarese, S. (2017). Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*.

Settles, B. (2009). Active learning literature survey.

Sevtsuk, A. (2020). *Street commerce: Creating vibrant urban sidewalks*. University of Pennsylvania Press.

Sevtsuk, A., Basu, R., Li, X., & Kalvo, R. (2021). A big data approach to understanding pedestrian route choice preferences: Evidence from san francisco. *Travel behaviour and society*, *25*, 41–51.

Shatu, F., & Yigitcanlar, T. (2018). Development and validity of a virtual street walkability audit tool for pedestrian route choice analysis—swatch. *Journal of transport geography*, *70*, 148–160.

Shuster, W. D., Bonta, J., Thurston, H., Warnemuende, E., & Smith, D. (2005). Impacts of impervious surface on watershed hydrology: A review. *Urban Water Journal*, *2*(4), 263–275.

Slater, S. J., Nicholson, L., Chriqui, J., Barker, D. C., Chaloupka, F. J., & Johnston, L. D. (2013). Walkable communities and adolescent weight. *American journal of preventive medicine*, *44*(2), 164–168.

Speck, J. (2013). *Walkable city: How downtown can save america, one step at a time*. Macmillan.

Sun, C., Su, J., Ren, W., & Guan, Y. (2019). Wide-view sidewalk dataset based pedestrian safety application. *IEEE Access*, *7*, 151399–151408.

Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5693–5703.

Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W., & Wang, J. (2019). High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv:1904.04514*.

Takebayashi, H., & Moriyama, M. (2012). Study on surface heat budget of various pavements for urban heat island mitigation (G. Polacco, Ed.) [Publisher: Hindawi Publishing Corporation]. *Advances in Materials Science and Engineering*, *2012*.

Talbot, L. A., Musiol, R. J., Witham, E. K., & Metter, E. J. (2005). Falls in young, middle-aged and older community dwelling adults: Perceived cause, environmental factors and injury. *BMC public health*, *5*(1), 1–9.

Tang, J., & Long, Y. (2019). Measuring visual quality of street space and its temporal variation: Methodology and its application in the hutong area in beijing. *Landscape and Urban Planning*, *191*, 103436.

Tao, A., Sapra, K., & Catanzaro, B. (2020). Hierarchical multi-scale attention for semantic segmentation. *arXiv preprint arXiv:2005.10821*.

TCAT. (2016). OpenSidewalks, Openly mapping for the pedestrian experience, Taskar Center for Accessible Technology (TCAT), University of Washington.

Theodosiou, Z., Partaourides, H., Tolga, A., Panayi, S., & Lanitis, A. (2020). A first-person database for detecting barriers for pedestrians.

Thomas, N. D., Gardiner, J. D., Crompton, R. H., & Lawson, R. (2020a). Keep your head down: Maintaining gait stability in challenging conditions. *Human movement science*, *73*, 102676.

Thomas, N. D., Gardiner, J. D., Crompton, R. H., & Lawson, R. (2020b). Physical and perceptual measures of walking surface complexity strongly predict gait and gaze behaviour. *Human movement science*, *71*, 102615.

Thomas M. Menino, T. J. T. (2013). Boston Complete Streets. https://tooledesign.com/project/boston-complete-streets-manual

Thornton, C. M., Conway, T. L., Cain, K. L., Gavand, K. A., Saelens, B. E., Frank, L. D., Geremia, C. M., Glanz, K., King, A. C., & Sallis, J. F. (2016). Disparities in pedestrian streetscape environments by income and race/ethnicity. *SSM-population health*, *2*, 206–216.

Tillson, G. W. (1900). *Street pavements and paving materials: A manual of city pavements: The methods and materials of their construction*. John Wiley & Sons.

Todic, F. (2016). Centerline: Calculate the polygon's centerline.

Treccani, D., Dıaz-Vilariño, L., & Adami, A. (2021). Sidewalk detection and pavement characterisation in historic urban environments from point clouds: Preliminary results. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, 243–249.

Tucker, C. J., Grant, D. M., & Dykstra, J. D. (2004). Nasa's global orthorectified landsat data set. *Photogrammetric Engineering & Remote Sensing*, *70*(3), 313–322.

Twardzik, E., Duchowny, K., Gallagher, A., Alexander, N., Strasburg, D., Colabianchi, N., & Clarke, P. (2019). What features of the built environment matter most for

mobility? using wearable sensors to capture real-time outdoor environment demand on gait performance. *Gait & posture*, *68*, 437–442.

United States Department of Justice, C. R. D. (1990). Americans with disabilities act (ada) of 1990, pub. l. no. 101-336, 104 stat. 328.

US Geological Survey. (2018). USGS EROS Archive - Aerial Photography - High Resolution Orthoimagery (HRO).

Van Cauwenberg, J., Van Holle, V., Simons, D., Deridder, R., Clarys, P., Goubert, L., Nasar, J., Salmon, J., De Bourdeaudhuij, I., & Deforche, B. (2012). Environmental factors influencing older adults' walking for transportation: A study using walk-along interviews. *International journal of behavioral nutrition and physical activity*, *9*(1), 1–11.

Van Dam, T. J., Harvey, J., Muench, S. T., Smith, K. D., Snyder, M. B., Al-Qadi, I. L., Ozer, H., Meijer, J., Ram, P., Roesler, J. R., et al. (2015). *Towards sustainable pavement systems: A reference document* (tech. rep.). United States. Federal Highway Administration.

Walker, J., & Johnson, C. (2016). Peak car ownership: The market opportunity of electric automated mobility services.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al. (2020). Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*.

Wang, K., Zhang, D., Li, Y., Zhang, R., & Lin, L. (2016). Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(12), 2591–2600.

Wang, R., Liu, Y., Lu, Y., Zhang, J., Liu, P., Yao, Y., & Grekousis, G. (2019). Perceptions of built environment and health outcomes for older chinese in beijing: A big data approach with street view images and deep learning technique. *Computers, Environment and Urban Systems*, *78*, 101386.

Wasserman, S., Faust, K. et al. (1994). Social network analysis: Methods and applications.

Wei, Y., Zhang, K., & Ji, S. (2019). Road network extraction from satellite images using cnn based segmentation and tracing. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 3923–3926.

Wel, Casper van der. (2019). PyGEOS.

Weld, G., Jang, E., Li, A., Zeng, A., Heimerl, K., & Froehlich, J. E. (2019). Deep learning for automatically detecting sidewalk accessibility problems using streetscape imagery. *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 196–209. https://doi.org/10.1145/3308561.3353798

Williams, J. E., Evans, M., Kirtland, K. A., Cavnar, M. M., Sharpe, P. A., Neet, M. J., & Cook, A. (2005). Development and use of a tool for assessing sidewalk maintenance as an environmental support of physical activity. *Health Promotion Practice*, *6*(1), 81–88.

Wilson, N., Pearson, A. L., Thomson, G., & Edwards, R. (2018). Actual and potential use of google street view for studying tobacco issues: A brief review. *Tobacco control*, *27*(3), 339–340.

Wu, H., Sun, B., Li, Z., & Yu, J. (2018). Characterizing thermal behaviors of various pavement materials and their thermal impacts on ambient environment. *Journal of cleaner production*, *172*, 1358–1367.

Xie, S., Feng, Z., Chen, Y., Sun, S., Ma, C., & Song, M. (2020). Deal: Difficulty-aware active learning for semantic segmentation. *Proceedings of the Asian Conference on Computer Vision*.

Xue, J., Zhang, H., Nishino, K., & Dana, K. (2020). Differential viewpoints for ground terrain material recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yang, J., Jin, S., Xiao, X., Jin, C., Xia, J. C., Li, X., & Wang, S. (2019). Local climate zone ventilation and urban land surface temperatures: Towards a performance-based and wind-sensitive planning proposal in megacities. *Sustainable Cities and Society*, *47*, 101487.

Yang, L., Zhang, Y., Chen, J., Zhang, S., & Chen, D. Z. (2017). Suggestive annotation: A deep active learning framework for biomedical image segmentation. *International conference on medical image computing and computer-assisted intervention*, 399–407.

Yang, X., Tang, L., Ren, C., Chen, Y., Xie, Z., & Li, Q. (2020). Pedestrian network generation based on crowdsourced tracking data. *International Journal of Geographical Information Science*, *34*(5), 1051–1074.

Ye, Y., Richards, D., Lu, Y., Song, X., Zhuang, Y., Zeng, W., & Zhong, T. (2019). Measuring daily accessed street greenery: A human-scale approach for informing better urban planning practices. *Landscape and Urban Planning*, *191*, 103434.

Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015).

big data'for pedestrian volume: Exploring the use of google street view images for pedestrian counts. *Applied Geography*, *63*, 337–345.

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

Yu, F., Wang, D., Shelhamer, E., & Darrell, T. (2018). Deep layer aggregation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2403–2412.

Yuan, Y., Chen, X., & Wang, J. (2019). Object-contextual representations for semantic segmentation. *arXiv preprint arXiv:1909.11065*.

Zapata-Diomedi, B., Boulangé, C., Giles-Corti, B., Phelan, K., Washington, S., Veerman, J. L., & Gunn, L. D. (2019). Physical activity-related health and economic benefits of building walkable neighbourhoods: A modelled comparison between brownfield and greenfield developments. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 1–12.

Zhang, F., Zhang, D., Liu, Y., & Lin, H. (2018). Representing place locales using scene elements. *Computers, Environment and Urban Systems*, *71*, 153–164.

Zhang, H., & Zhang, Y. (2019). Pedestrian network analysis using a network consisting of formal pedestrian facilities: Sidewalks and crosswalks. *Transportation research record*, *2673*(7), 294–307.

Zhang, L., Yang, F., Zhang, Y. D., & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. *2016 IEEE International Conference on Image Processing (ICIP)*, 3708–3712.

Zhang, Y., Odeh, I. O., & Han, C. (2009). Bi-temporal characterization of land surface temperature in relation to impervious surface area, ndvi and ndbi, using a sub-pixel image analysis. *International Journal of Applied Earth Observation and Geoinformation*, *11*(4), 256–264.

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

Zhao, S., Wang, Y., Yang, Z., & Cai, D. (2019). Region mutual information loss for semantic segmentation. *arXiv preprint arXiv:1910.12037*.

Zhao, Y. (1997). *Vehicle location and navigation systems*.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 633–641.

Zhou, G., Chen, W., Kelmelis, J. A., & Zhang, D. (2005). A comprehensive study on urban true orthorectification. *IEEE Transactions on Geoscience and Remote sensing*, *43*(9), 2138–2147.

Zhou, H., Liu, L., Lan, M., Zhu, W., Song, G., Jing, F., Zhong, Y., Su, Z., & Gu, X. (2021). Using google street view imagery to capture micro built environment characteristics in drug places, compared with street robbery. *Computers, Environment and Urban Systems*, *88*, 101631.

Zhu, S., & Mai, X. (2019). A review of using reflective pavement materials as mitigation tactics to counter the effects of urban heat island. *Advanced Composites and Hybrid Materials*, *2*(3), 381–388.

Zhu, Y., Sapra, K., Reda, F. A., Shih, K. J., Newsam, S., Tao, A., & Catanzaro, B. (2019). Improving semantic segmentation via video propagation and label relaxation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8856–8865.