MAPPING URBAN TREES WITH DEEP LEARNING AND STREET-LEVEL IMAGERY

by

Stefanie Lumnitz

BSc. Geography, Ludwig-Maximilian's University Munich, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Forestry)

The University of British Columbia
(Vancouver)

December 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

MAPPING URBAN TREES WITH DEEP LEARNING AND STREET-LEVEL IMAGERY

submitted by **Stefanie Lumnitz** in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE** in **Forestry**.

Examining Committee:

Dr. Verena Griess, Forest Resource Management
Supervisor

Dr. Nicholas Coops, Forest Resource Management
Supervisory Committee Member

Dr. Tahia Devisscher, Forest Resource Management
Supervisory Committee Member

Dr. Helge Rhodin, Computer Science
Additional Examiner

Abstract

Planning and managing urban trees and forests for livable cities remains an outstanding challenge worldwide owing to scarce information on their spatial distribution, structure and composition. Sources of tree inventory remain limited due to a lack of detailed and consistent inventory assessments. In practice, most municipalities still perform labor-intensive field surveys to collect and update tree inventories.

This thesis examines the potential of deep learning to automatically assess urban tree location and species distribution from street-level photographs. A robust and affordable method for detecting, locating, classifying and ultimately, creating detailed tree inventories in any urban region where sufficient street-level imagery is readily available was developed.

The developed method is novel in that a Mask Regional Convolutional Neural Network is used to detect and locate tree instances from street-level imagery, creating shape masks around unique fuzzy urban objects like trees. The novelty of this method is enhanced by using monocular depth estimation and triangulation to estimate precise tree location, relying only on photographs and images taken from the street. In combination with Google Street View, a technique for the rapid development of an extensive tree genera training dataset was presented based on the method of tree detection and location. This tree genera dataset was used to train a Convolutional Neural Network (CNN) for tree genera classification.

Experiments across four cities show that the novel method for tree detection and location can be transferable to different image sources and urban ecosystems. Over 70% of trees recorded in a ground-truth campaign (2019) were detected and could be located with a mean error in the absolute position ranging from 4m to

6m, comparable to GPS accuracy used for geolocation in classical manual urban tree inventory campaigns. The trained CNN classifies 41 fine-grained tree genera classes with 83% accuracy. The detection and classification models were then used to generate maps of urban tree genera distribution in the Metro Vancouver region.

Results of this research show that developed methods can be applied across different regions and cities and that deep learning and street-level imagery show promise to inform smart urban forest management, including bio-surveillance campaign planning.

Lay Summary

Urban trees play a vital role in making our cities more livable, sustainable and resilient to climate change. In order to manage and maximize benefits urban trees provide for cities and their inhabitants, city officials need to know the location of trees and how different species are distributed throughout urban environments. This thesis explored a novel approach to collect information on the species and distribution of urban trees from photographs taken from the streets. Using new technologies like deep learning, a tool was developed that detected over 70% of all trees growing on streets and classified 41 tree species for different cities in the Metro Vancouver region. In addition, it was examined if the developed tool can be transferred to other urban areas. In future, the developed methods, tools and data inform urban tree inventories to assist planning decisions and management schedules of city planners and urban forest practitioners.

Preface

This research was proposed, designed and carried out in its entirety by myself, with support of my Masters supervisory committee. I identified and designed this original research, with suggestions and guidance from the committee. I performed all phases of methodological development, field data collection, data analysis, interpretation of results and manuscript preparation as primary researcher. I sought advice and support of my committee and co-authors, who provided clarification and modifications. This research project was undertaken as part of the BioSAFE project and in partnership with the Canadian Food Inspection Agency, the Canadian Urban Environmental Health Research Consortium and the University of California, Riverside. A list of publications and presentations of thesis content as part of the BioSAFE project can be found in appendix A.

Chapters 2 and 3 are independent research chapters that have been structured and written as scientific articles. A list of publications for each research chapter is presented as follows:

Chapter 2:

- Lumnitz, S., Devisscher, T., Mayaud, J., Coops, N. and Griess, V. (in prep): Mapping urban trees with deep learning and street-level imagery.

Chapter 3:

- Lumnitz, S., Devisscher, T., Coops, N. and Griess, V. (in prep): Mapping urban tree diversity with deep learning and street-level imagery.

Table of Contents

List of Tables

List of Figures

Glossary

AGM Asian Gypsy Moth

AI Artificial Intelligence

ALB Asian Longhorned Beetle

API Application Programming Interface

APP Application

BC British Columbia

CFIA Canadian Food Inspection Agency

CNN Convolutional Neural Network

COCO Common Objects in Context

CONV LAYER Convolutional Layer

DED Dutch Elm Disease

DL Deep Learning

DNN Deep Neural Network

EAB Emerald Ash Borer

FASTER R-CNN Faster Region-based Convolutional Neural Network

FIS Forest Invasive Species

FPN Feature Pyramid Network

FC LAYER Fully-connected Layer

FOV Filed of View

FPN Feature Pyramind Network

FP1 Single-precision Floating Points

FP16 16-bit Floating Points

GPS Global Positioning System

GPU Graphics Processing Unit

GVI Green View Index

GSV Google Street View

IOU Intersection over Union

ISPRS International Society for Photogrammetry and Remote Sensing

ITCD Individual Tree Crown Delineation or Detection

KDE Kernel Density Estimates

LIDAR Light Detection and Ranging

MASK R-CNN Mask Region-based Convolutional Neural Network

ML Machine Learning

PPM Pine Processionary Moth

RESNET101 Residual Network with 101 Layers

RESNET50 Residual Network with 50 Layers

RGB Red Green Blue

RMSLE Root Mean Squared Log Error

ROI Region of Interest

SLAM Simultaneous localisation and mapping

SOD Sudden Oak Death

SFM Structure from Motion

NN Neural Network

RS Remote Sensing

VHR Very High Resolution

Acknowledgments

This research was funded as part of the bioSAFE project and Genomics Canada, BC and Quebec. The Canadian Food Inspection Agency provided insights and knowledge in current bio-surveillance methods and gaps. The Canadian Urban Environmental Health Research Consortium provided access to a High Performance Compute Cluster. Many thanks to the organizations that offered financial assistance including Google and the Python Software Foundation for a Google Summer of Code Fellowship, the Mitacs Globalink Fellowhsip, the fast.ai International Fellowships, and the Mary and David Macaree Fellowship.

I thank my supervisor Dr. Verena Griess for your mentorship and unwavering support in persuing this research and other projects that came along during this Masters. Thank you Dr. Nicholas Coops for sharing your expertise and always inviting me to join the IRSS lab. Dr. Tahia Devisscher, I am grateful for all the inspiring discussions, ideas and projects we shared. I thank the members of the Fresh lab for your company, encouragement and valuable feedback in all aspects of life at UBC and in Canada. Thank you Valentine Lafond and Kathleen Coopland for your help in the field and lab.

I thank the Center for Geospatial Sciences team at the University of Riverside and all members of the Python Spatial Analysis core-development team for your friendship and encouragement. Our conversations and support always brought new perspectives, ideas and hacks into my day-to-day life and work. Lastly, thank you Sam Anderson, Jerome Mayaud, Sophie Nitoslawski, Christina Draeger and Ralf Gommers for your friendship, consolidation, guidance and never tiring support. You made my work days lighter and I feel extremely fortunate to have benefited from meeting and spending time with you.

Dedication

This work is dedicated to my parents and brother - Martina, Ramon and Fritz Lumnitz - for their love and support, reaching over continents.

Chapter 1

Introduction

The great green city of the future is ecologically and economically resilient; it's made up of healthy, livable neighborhoods where the benefits of nature are available to all people. — Pascal Mittermaier, The Nature Conservancy's Global Managing Director for Cities (2019)

1.1 Healthy green cities of the future

By 2050, three out of four people on Earth will live in cities [139]. As urbanization continues in the epoch of the Anthropocene [30], cities embody the forefront of action against global change impacts, but also become vulnerable to their detrimental effects [35, 55]. It is increasingly recognized that urban trees play a critical role in mitigating negative effects of global change for people and the planet [38, 53]. Numerous studies have shown that urban trees are key in making cities more livable, resilient and help adapt for impacts of future climate change [64]. By providing shade, a natural way of air-cooling and absorbing CO_2 through growth, trees help mitigate climate change and save energy by reducing the need for air conditioning [137]. They clean the air and environment, by capturing particulates and urban pollutants through natural gas-exchange with the atmosphere [137]. Urban trees promote storm water runoff as they intercept rainfall and increase infiltration [60]. The presence of healthy trees in urban areas is known to have beneficial effects on human health and well being, promoting mental health, reducing stress, prevent-

ing obesity and accelerating recovery from illnesses [138, 140]. Urban forest have the potential to foster biodiversity, by providing shelter and food for animals and plants [7]. It has been shown that urban forests have direct social impacts such as increasing property values, positively impacting social cohesion and strengthening communities [37, 101, 145].

As evidence gathers about the diverse benefits healthy urban trees provide to resilience and livability in cities through various ecosystem services, the demand for effective urban forest management and planning grows [106, 140]. Expanding and maintaining a healthy urban forest is a recognized challenge to-date [38]. Poorly managed urban forests do not provide the same ecosystem services and can even lead to property damage, personal injury or other disservices [97, 114]. Threats to urban tree health and challenges associated with their mitigation and management are diverse [38]. Street trees for example often suffer from water stress caused by decreasing water availability to root systems from de-icing salts, barriers to root growth, poor soil quality or presence of toxic substances [12]. Above ground stressors include heat radiation from buildings and impervious surfaces, high winds channelized through urban canyons, cutting of tree crowns and growth inhibiting light patterns, especially for trees planted on the north side of buildings in the northern hemisphere [87]. Owing to urban trees proximity to centers of human activity and international trade routes, they are further threatened by damage through native and invasive pests and pathogens [110]. Many of these threats to urban tree health are expected to intensify with the effects of climate change, such as rising temperatures in cities [91].

Proactive management and decision making is required to protect, improve and extend urban forests with a direct influence on over 60% of the world population in the future [38, 53]. One of the biggest limitations for proactive urban tree management is the scarcity of up-to-date urban tree inventory data, used as a basis for planning and decision-making [42]. To date, the most common practice to retrieve such important information is the manual collection and measurement of single urban trees with hand held devices (s. section 1.1.4). At present, existing tree inventory data are mostly restricted to public street trees or other trees on public land and there is a lack of information on a large proportion of the urban forest, especially trees on residential property. Cost-efficient and widely applicable tools

are needed to provide high-resolution spatial information to enhance urban tree management and support decision making [144].

1.1.1 Smart urban forest management

The field of urban forestry is growing rapidly alongside novel, technically oriented urban sciences, like ecological engineering and smart city planning [11, 38]. "Smart urban forest management" describes the integration of urban forest management into emerging smart city planning concepts, applying novel technological developments like Artificial Intelligence (AI), open source mapping platforms or mobile Application (APP) driven citizen engagement to address the diverse challenges urban trees face [105]. In the context of proactive, resilient and smart urban forest management, the purpose of this work is to explore the suitability of novel Deep Learning (DL) architectures and openly available street-level imagery to develop a tool that meets the need for up-to-date urban tree inventory data (s. chapter 1).

The approach proposed in this thesis is based on recent advances in "instance segmentation" for fuzzy objects and monocular depth estimation to locate features detected on photographs in space (s. chapter 2). Ultimately, the aim is to produce a robust, cost-effective and rapid method for creating detailed tree location and diversity data in any urban region where sufficient street-level imagery is readily available. To demonstrate the value of the developed model for smart urban forest management, the created tree diversity data is used to inform urban biosurveillance management in the Metro Vancouver region (s. chapter 3).

1.1.2 Urban trees as vectors for tree pests and pathogens

One of the major challenges enhanced by climate change for urban tree management is the introduction and spread of native and invasive pests and pathogens in urban areas [32]. Canada's urban forests, for example, are increasingly threatened by Forest Invasive Species (FIS) such as the Emerald Ash Borer (EAB), Asian Longhorned Beetle (ALB), the Asian Gypsy Moth (AGM), Dutch Elm Disease (DED) or Sudden Oak Death (SOD) [109]. These FIS can cause irreversible damage to both natural and urban ecosystems and are associated with high management

costs after establishment. In 2003 ALB infestations in Toronto, for example, lead to the replacement of 28,700 urban trees, after the infestation was detected in campaigns lead by the the Canadian Food Inspection Agency (CFIA) responsible for invasive species management in Canada. Within this ALB management effort, the replacement of one tree was subsidized by CA$300 in private areas, CA$150 in public areas and CA$40 in urban woodlands, generating a total cost of $>$ CA$6 millio for eradication of ALB in Toronto alone (correspondence M. Marcotte, October 2019, CFIA). Furthermore, Canada wide FIS are estimated to cost CA$800 million annually in management efforts and further generate a threat to export markets estimated up to CA$2.2 billion annually.

Increasingly, new FIS are expected to enter Canada, with urban forests acting as key nodes in their dispersal pathways [110]. Yemshanov et al. [147] identified Metro Vancouver and the Greater Toronto area as the two major points of entry for invasive pests and pathogens through international trade and transportation networks. Urban trees located close to centers of human activity and international trade routes such as ports, commercial zones or tree nurseries within urban environments are under constant risk to be exposed to native and invasive pests and pathogens [110]. Trees that are unhealthy are under considerable threat of additional damage caused by tree pests and pathogens, as defense mechanisms in unhealthy trees are weakened [86]. Once an invasive pest or pathogen is established in urban areas, urban trees can act as vectors for spread of these harmful diseases into surrounding ecosystems [38].

Next to maintaining a healthy forest, urban tree managers face additional pressure to detect these pests and pathogens early, to contain and prevent the spread and establishment throughout the urban ecosystem into surrounding areas [110]. The earlier FIS populations are noticed at the initial stages of infestation and invasion, the higher the cost-efficiency and probability of management success [86]. Consequently, invasion managers are faced with the task of maximising bio-surveillance and early detection efforts to support rapid decision making, and minimize potential FIS management costs [88]. This is significant because 1) established FIS in urban spaces can have negative impacts on ecosystem services urban forests provide, and 2) urban trees can act as sentinels for early detection and rapid response before establishment, preventing FIS spreading into natural ecosystems [21].

1.1.3 The need for tree inventory data

Unfortunately, urban forest planning and management remains an outstanding challenge worldwide owing to relatively scarce information on the spatial distribution and accessibility of urban forests, as well as their health condition, composition, structure and function. Urban forest assessments are the basis for all decision-making in managing and mitigating threats to tree health [60]. Tree inventories are an important aspect of urban forest assessments, and usually involve the collection of field data on the location, genus, species, crown shape and volume, diameter, height and health condition of urban trees [68]. Urban tree inventories predominantly focus on information on individual urban trees, less so on groups of trees as for example found in urban parks [102]. At present, existing tree inventory data are mostly restricted to public street trees or other trees on public land and exclude large areas of urban forest, especially trees on residential properties. For example, in Vancouver, about 37 percent of urban forest is located on private land, and not included in the city inventory [62].

For the purpose of bio-surveillance management, extensive inventories are needed to manage and contain the spread and economic cost of harmful FIS through early detection [8, 109]. Information about potential host-tree distribution over the urban space can help to identify hotspots of establishment for FIS [116]. Similarly, tree-health can be used as an indicator to detect trees already harmed by FIS. Knowing where unhealthy trees are located can also be valuable to predict where FIS are most likely to spread, since natural defense mechanisms of unhealthy trees are already weakened [115]. Furthermore, detailed tree inventories can be used to quantify the monetary value of environmental and aesthetic benefits of single trees, including ecosystem services [126]. Weighting the cost of managing and protecting urban trees against the benefits or services they provide is often used as a basis for decision making or economic risk assessments [60]. Such information can be used as an economic baseline to inform decision makers which FIS mitigation strategies are economically valuable and what is at risk for different stakeholders in case urban trees are attacked [16, 88]. For the purpose of bio-surveillance it is therefore beneficial to provide up to date urban tree inventories, especially including information about location, tree health, tree genus and general tree structure. Furthermore, it is

important to collect data over both public and private trees, since pests and insects spread over public and private urban property. Additionally, detailed tree inventory information about tree location or genus provides a backbone for many other urban forest assessments, such as the prediction of urban tree health under changing climate conditions [47].

Methods used to collect data and the extent of these inventories is often governed by the direct application of the inventory and the municipality's budgetary constraints [102]. Depending on these factors cities need to decide whether they are collecting information for every single urban tree or for a subset of trees and if the inventory will be updated over time or information is only collected once, resulting in a fragmented patchwork of multiple data collections merged over time [68]. Keller and Konijnendijk [68] point out the need for inventory methods that can be scaled over multiple regions in order to develop national and international recommendations and standards for urban tree inventory data collection. More abundant and standardized urban tree inventory data, specifically including private trees, would open the possibility to plan and manage public green spaces more holistically, integrating both public and private green spaces into one urban forest. Aronson et al. [9] stress the need for this more holistic management approach in order to protect and manage urban trees and biodiversity sustainably in future. Hence, cost-efficient and widely applicable tools are needed to provide high-resolution spatial information to enhance early detection and support decision making [144].

1.1.4 Collecting urban tree data

Nielsen et al. [102] distinguish four main types of generating and updating previously identified information in urban tree inventories: satellite-supported methods, airplane-supported methods, on-the-ground scanning or digital photography, and field surveys. Satellite-supported methods have primarily been used for single-tree crown detection [67] or tree health assessments [117]. Data can be retrieved by multiple sensors ranging from Red Green Blue (RGB) colour space, over multi-spectral, hyper-spectral to panchromatic. Most satellite-based imagery used for urban tree inventories is of Very High Resolution (VHR) in order to detect the comparably small objects of trees [102]. Similarly, airplane-supported methods have

been used for tree detection, tree health assessments and more recently tree species classification [6]. Multi-spectral, hyper-spectral, and Light Detection and Ranging (LIDAR) sensors provide the advantage of more quickly generating relatively high resolution data over bigger areas than field surveys [102]. However, in contrast to satellite-based imagery, airplane-supported sensors have to explicitly be flown for the purpose of tree inventory data collection. On-the-ground scanning or digital photography can in general be used to retrieve more detail and volume of single tree inventory parameters than aerial based methods. [113] for example developed a semi-automated method to calculate tree crown volume and density from side view photographs. Nevertheless, classical field surveys with direct manual measurement and visual tree inspection are the most commonly used method to generate and update urban tree inventories [102].

All of these methods above are often limited in either geographical space, temporal coverage or the number of parameters covered by the urban tree inventory [144]. The need to perform labour-intensive field surveys or costly aerial campaigns often limits the detail and frequency of urban tree inventory updates [6]. Most data sources furthermore lack processing methods that can be generalized or automated over multiple cities. Often, expert knowledge is required to handle large file sizes or semi-automate classifications [144]. This leads to some municipalities not being able to collect tree inventory data at all, due to a constrained urban tree inventory budget [102]. [68] point out the need for methods scaling over various regions in order to develop national and international recommendations and standards for urban tree inventory data collection. Geospatial technologies and datasets that are more cost efficient or free could be used to support a larger number of municipalities and allow for urban tree inventory standardization [142]. A data source that has recently attracted a lot of attention by the urban forest research community, due to its low cost and global coverage, is street-level imagery in general and Google Street View (GSV) in particular [13].

1.2 Computer vision for urban tree inventories

Two recent trends have gained attention in smart city planning because they allow remote data collection, can be applied over large areas at low cost, and promote

uptake from a larger number of municipalities [132]. First, the growing availability of low-cost, detailed and increasingly crowd-sourced street-level imagery (photographs of street scenes taken from the ground) [13, 77]. Second, the success of DL and Convolutional Neural Network (CNN) out-competing other methods for extracting abstract features and objects in imagery [89].

1.2.1 Street-level imagery

Abundance of street-level imagery

In this thesis, street-level imagery is defined as photographs taken on the street, captured by RGB sensors mounted on different types of vehicles (i.e cars, bikes) or hand-held devices (i.e mobile phones, cameras). Easy access to affordable RGB sensors and cameras by companies and the public, in combination with the willingness to share imagery on the internet has lead to wide availability of street-level imagery data [125]. To-date, there is an increasing abundance of platforms and services providing street-level images. Imagery can, for example, either be accessed through relevant Application Programming Interface (API)'s from e.g. GSV and Bing Maps Streetside or it can be crowdsourced and collected through services like Mapillary and OpenStreetCam. Currently efforts to standardize street-level imagery are limited, resulting in data coming with varying quality and quantity depending on the imagery provider.

GSV is a geospatial platform that offers standardized and geocoded street-level imagery in different formats and resolutions at relatively low cost [52]. GSV provides extensive spatial coverage of North America and other countries in the world. Google [52] gives a detailed and up to date description which places are covered by GSV, when they will be recorded next and when they were previously recorded. GSV Street-level imagery is typically collected through a panoramic camera mounted on a car roof. Panoramic recordings are single snapshots in time covering a range of view of 360 degrees, spaced every 15 meters apart on public roads which means that one tree can be seen in multiple images [144]. GSV updates street-level images of public roads every 1-4 years. GSV data can be accessed online via an official API, that allows querying of the closest street-level image ac-

cording to a given geographic position or a geographic latitude and longitude data pair [144].

Street-level imagery for smart urban forest assessments

GSV imagery has already found its way into smart urban forestry assessments in recent literature. Berland and Lange [13] manually inspected single street-level imagery to generate tree inventory data through "virtual tree inventory surveys". They found that these "virtual surveys" of street trees conducted with GSV agreed with field data with over 93% of documented trees and discovered that it was possible to assess genus, species, location, diameter at breast height and tree health. Rousselet et al. [118] tested if GSV data could be used to identify trees under attack by the Pine Processionary Moth (PPM), through manual visual inspection of imagery by bio-surveillance professionals. A comparison of field data retrieved by a large-scale analysis based on a mesh of 16 km grid size with a GSV based approach recorded 96% of matching positive findings. Nevertheless, these studies and many others, are still not automated and limited by expensive manual labour [36].

Advances in Computer Vision (s. section 1.2.2), the field of study that assesses possibilities to automate tasks of the human vision system with a computer [83], and DL are enabling automatic and robust information extraction for street-level imagery in urban environments [57]. Computer vision algorithms developed for smart city research using street-level imagery have been applied to assess demographics [45], urban change [99], wealth [48], perceived urban safety [100], building types [66] and urban morphology [94]. In the field of smart urban forestry, street-level imagery in combination with computer vision has been applied in three key areas: 1) estimation of shade provision for urban trees [78, 80, 81], 2) quantification of perceived urban canopy cover [19, 36, 79, 124, 132], and 3) mapping the location of urban trees [15, 144].

Li et al. [81], for example, calculated a sky view factor from GSV, which indicates the level of enclosure of street canyons, in order to quantify shade provisioning from trees. Seiferling et al. [124] used GSV imagery in combination with Machine Learning (ML) techniques to quantify perceived urban canopy cover. Similarly, Li et al. [79] assessed the percentage of vegetation in streets, by quantifying

the amount of green pixels seen in a Street View Scene. These methodologies helped to shape the so called Green View Index (GVI) [36]. This index indicates how green streets are perceived by pedestrians and has been synthesized for multiple cities all over the world [36]. Wegner et al. [144] designed a workflow for automatic street tree detection and geolocation from GSV and Google Maps imagery, based on the Faster Region-based Convolutional Neural Network (FASTER R-CNN) framework.

Research gaps

Most workflows generating urban tree inventory data using GSV and computer vision techniques are currently limited by quantifying and classifying single pixels without distinguishing between separate trees, detecting tree position without the possibility to quantify tree characteristics or relying on secondary data sources for precise location predictions of trees [36]. This thesis proposes a workflow for the detection, classification, and geolocation of separate trees ready to use as part of tree inventories and introduces a method to streamline geolocation only relying on street-level imagery.

1.2.2 Theoretical background of deep learning

DL has proven to be a powerful tool for extracting abstract features and objects from raw imagery, and is increasingly adopted in ecology [27], environmental research and the Remote Sensing (RS) community [154]. In the following section relevant DL concepts for computer vision and their technical background will be presented. In order to define DL, the umbrella concepts of AI and ML will be introduced, under which DL can be placed as a sub-discipline (s. section 1.2.2). Next, CNNs and their technical concepts, challenges, and limitations of use will be discussed (s. section 1.2.2). Finally, the section will conclude by explaining the application of modern computer vision to solve problems in environmental monitoring and earth observation (s. section 1.2.2).

Artificial Intelligence and Machine Learning

AI is the study to automate intellectual tasks normally performed by humans [119]. Figure 1.1 provides a conceptual overview of how DL can be placed in the field of AI. ML is a subfield of AI and represents the new paradigm in the development of algorithms that computers are able to learn without being explicitly programmed [121]. In comparison to classical programming where specific rule sets are hard-coded to process data to best match an output, ML systems allow to automatically generate these rule sets though exposure to an input-output data pair [136]. The process of generating and refining these rule sets, by automatically comparing the systems current output to its expected output, is commonly described as training or learning [26].

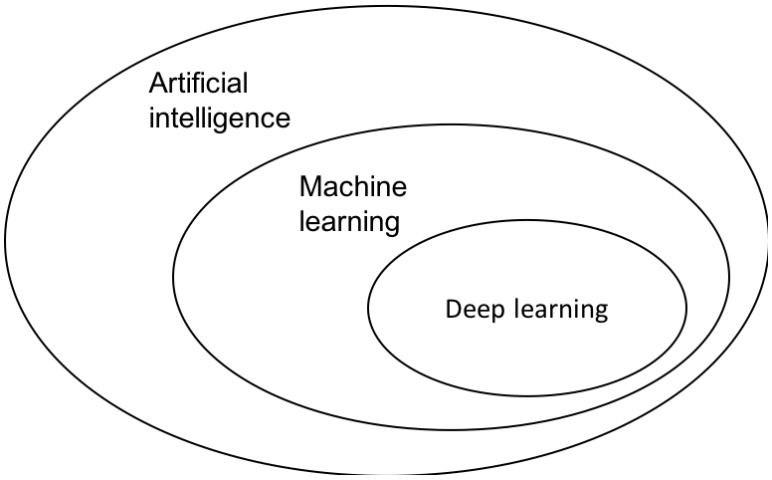


Figure 1.1: Deep learning, machine learning and artificial intelligence. Deep learning is a concept used in machine learning, which is in turn a subfield of artificial intelligence [26].

The field of ML can be further separated into three broad categories: Unsupervised Learning, Reinforcement Learning and Supervised Learning. Unsupervised Learning systems transform data without the previously described use of specific targets, answers or outputs in the training process [56]. The two best known Unsupervised Learning tasks are Clustering and Dimensionality Reduction. These tasks are used to compress, denoise or visualize data and are often a necessary step

to analyze a dataset before using it in a supervised-learning problem [26]. Reinforcement Learning conceptually places an agent at the core of a system who learns to take action in order to gain the maximum reward by receiving information about its environment [95]. Reinforcement learning has recently made a major break through, for example, with Google DeepMind's AlphaGo, a system using reinforcement learning to master and succeed in the game of Go [128].

Supervised Learning, however, is currently the most commonly used type of ML and the most dominant form of DL. At its core, Supervised Learning is the process of meaningfully transforming input data into new data representations, which is learned by exposing the system to known input-output data pairs. In other words, Supervised Learning systems automatically learn new data representations by mapping input data onto a set of known output targets, so called annotations or labels [26]. The trained system can then be used to generate predictions, its own annotations and labels, on new input data. This process is called inference [75].

Deep learning

As a sub-field of Supervised Learning, DL represents the notion of learning multiple, hierarchical layers of representations in between input data and output target [75]. Shallow learning in comparison only transforms input data into one or maximal two successive representation layers. Representations in DL gain in complexity with each successive deeper layer, whereby complex representations are build out of simpler representation [51]. DL is most commonly applied in a Neural Network (NN), also called a Deep Neural Network (DNN). These terms are often used interchangeably [26]. Figure 1.2 shows how a DNN conceptualizes the image of a person by layers of successively more and more complex representations.

DL in NN's could be thought of as a multistage information-distillation process [26]. A NN "learns" by creating a sequence of data transformations to map a given input to the target output (s. fig 1.2). Each step transforming the data is implemented in successive layers and parameterized by so called weights or parameters [26]. In the process of learning, these parameters are automatically updated, depending on a distance or loss score between the prediction and the target, calculated with the so called loss function (s. fig 1.3 (b)). The goal in DL is to minimize the

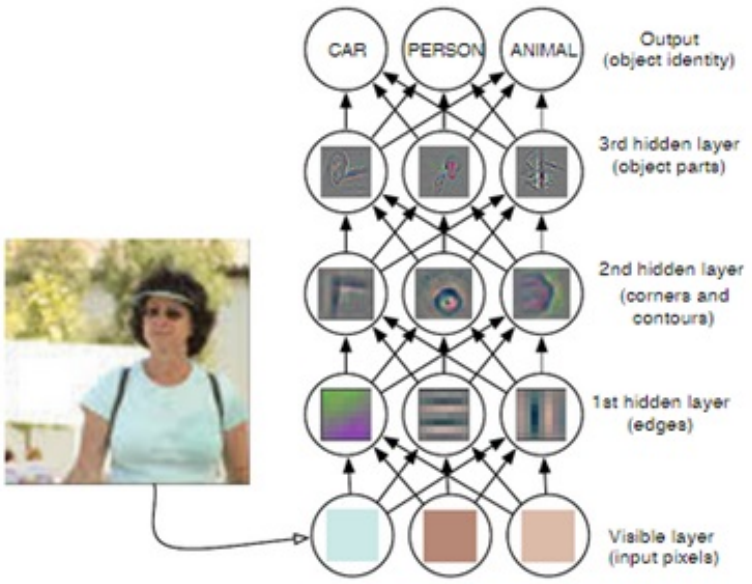


Figure 1.2: Concept of deep neural networks. The image of a person is presented in a concept of successive layers of representations. Representations get more complex the ”deeper” the layers of the NN. These deep layers build on top of shallower layers containing simpler representations, like colors, vertical or horizontal lines [51].

loss score, to achieve a close match between prediction and target. To optimize the loss score, parameters will be adjusted depending on an optimizer, a mechanism that implements the so called back-propagation algorithm, i.e. a specific variant of stochastic gradient descent [26]. In other words, training a DL model means that data is transformed into new data representations or features by exposing the model to a set of input variables and output targets, so called training data, to automatically update parameters and minimize the loss function. Typically, the ”deeper” the NN the larger the amount of training data needed in order to learn meaningful representations.

Convolutional Neural Networks

Even though DL has a long history, the field has only recently achieved a major break through in near-human-level to even superhuman performance in image classification [28]. The increasing depth of DNN frameworks, scale of computational power available for training and the amount of openly available training data has accelerated scientific discovery and development within the field and popularity of methods transferable to other areas of research since the early 2010s [26]. Most important though was the invention of a so called Convolution Operation used by CNN [51].

CNNs are the most common algorithmic architecture to implement DL for analyzing imagery (s. fig 1.3). They are explicitly designed to process large, multi-dimensional tensors such as volumetric data or images, with typical dimensions of nr. of pixels in width x nr. of pixels in height x 3 (red, green and blue) for true color images [51]. An ordinary NN, relying mainly on fully-connected layers, where a pixel would represent one neuron (i.e. unit within a DL model), would need to learn a vast amount of parameters even for relatively small images, i.e. an image of size $64x64x3$ would result in a parameter vector of length $12,288$ only for one shallow layer. In contrast to regular NN's, CNN's leverage the concept of parameter sharing [74]. They learn the parameters of convolutional filters to directly extract meaningful features from images, reducing the amount of parameters that need to be learned and therefore enhancing scalability of data input and processing speed [26]. The three basic building blocks of CNN's are the Fully-connected Layer (FC LAYER), the Convolutional Layer (CONV LAYER) and the pooling layer (s. fig 1.3).

FC LAYERs, often referred to as "densely connected layers" are the standard layers representing the original idea of a NN, where all input elements, so called "neurons", are connected to all output elements, another array of neurons (s. fig 1.3). Parameters in FC LAYERs are one dimensional arrays and can be very large, i.e. of length 12,288 for a small image. In CNN architectures, FC LAYERs are typically applied at the "top" of the network for classification, to transform the input into a last desired number of outputs, i.e. a list of labels or a one dimensional array.

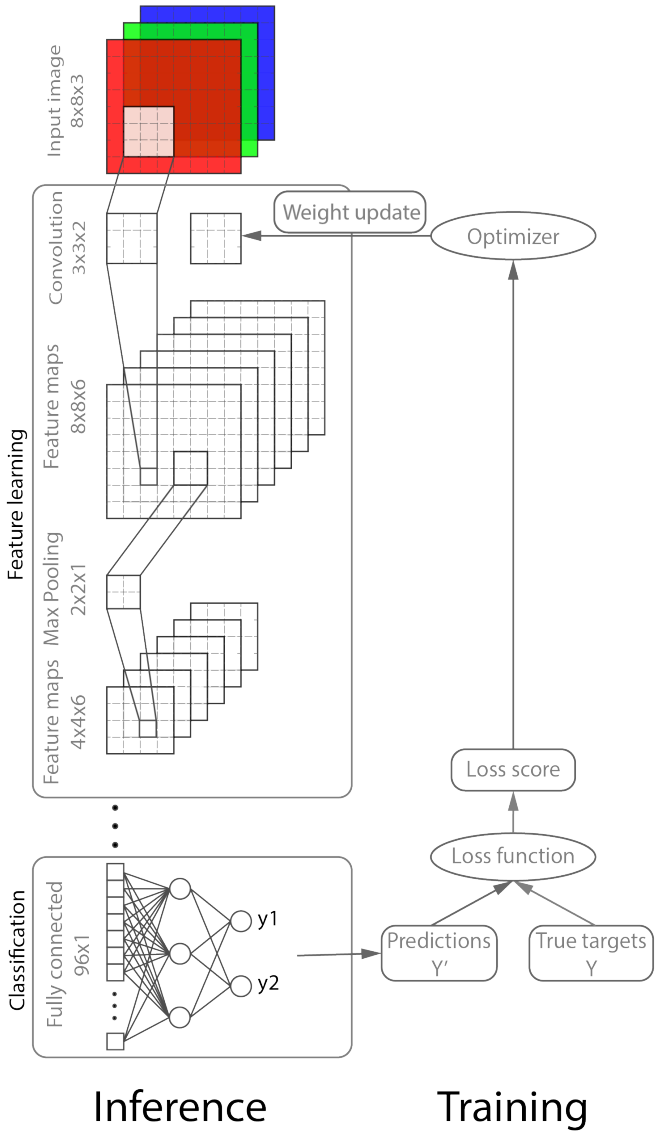


Figure 1.3: Convolutional neural network architecture, inference and training. A CNN consists of a sequence of CONV LAYERs and pooling layers for feature learning, typically followed by FC LAYERs for classification of learned features. Inference denotes the process of running an input image through all layers subsequently to generate a prediction Y'. Training a CNN requires to optimize weights stored in filters through calculating a loss score between predictions Y' and target Y.

CONV LAYERs are at the core of a CNN and use convolutional filters to implement the concept of parameter sharing. In contrast to FC LAYERs, parameters in CONV LAYERs are directly represented by convolutional filters. In figure 1.3, for example, a layer of two convolutional filters of size $3x3$, has a total of $3x3x2 = 8$ parameters that need to be learned for the same small input image with the dimensions of $64x64x3$. Typically, each filter is a tensor of relatively small size compared to the input image. In other words, the fundamental difference between FC LAYERs and CONV LAYERs is that FC LAYERs learn global patterns and CONV LAYERs learn local patterns in the input feature space, found through small filters. In addition to reducing the amount of parameters that need to be learned, CONV LAYERs allow the model to learn patterns that are translation invariant, i.e. if the model can recognize horizontal edges in the upper left corner of the image, it can recognize the same pattern everywhere, reducing the amount of training images that are needed. An FC LAYER would need to learn the same pattern at a different location again. Furthermore, CONV LAYERs allow models to learn spatial hierarchies of patterns, starting with less complex patterns like edges and increasingly learning more complex and abstract representations (s. fig 1.2). A convolution works by sliding the filter over the last 3D feature map, and transforming the extracted 3D patch (of shape $width_{filter}\ x\ height_{filter}\ x\ depth_{input}$) into a 1D vector (of shape ($depth_{output}$,)). This process quantifies the presence of the filter's pattern at different locations in the image and results in a new feature volume of size ($width_{filter}\ x\ height_{filter}\ x\ depth_{output}$) (s. fig 1.3).

Lastly, pooling layers are applied to reduce the spatial extent of feature maps [51]. These layers do not contain learnable parameters, but are used to reduce computational cost of the model by reducing the image resolution while preserving the depth and allowing for more filters to be applied. Max pooling is the most common pooling strategy, convolving the maximum pixel value while sweeping a filter over an image (s. fig 1.3). For a complete complete overview of pooling and CNNs see Goodfellow et al. [51].

The invention of CNN has already revolutionized research in the field of robotics and self driving cars [54] or medical imagery analysis [85]. The application of DL in urban and ecological assessments in general and on street-level imagery for updating urban forest inventories and monitoring urban trees in particular, is still

sparse (s. section 1.2.1). Owing to a fast growing open source community, the availability of pre-trained CNNs and extensive, openly available annotated datasets on street-level imagery, CNNs show great potential to be successfully applied to problems with small-scale data availability and fine-grained classification problems [51].

Classification, detection, segmentation and instance segmentation

The four main problems in computer vision to solve with CNNs are (s. fig 1.4): (1) thematic image classification (e.g. classifying an image as tree vs non-tree) [122], (2) multiple object detection (e.g. retrieving bounding box pixel-locations of all trees in the image) [51], (3) pixel-wise semantic segmentation (e.g. classifying every single-pixel into the class tree or the class non-tree) [18], and (4) pixel- and object-wise instance segmentation (e.g. retrieving every pixel that belongs to the class tree, differentiated per single tree instance) [26]. Zhang et al. [152] provide a detailed overview and a technical tutorial for RS data analysis using DL algorithms and Mountrakis et al. [98] highlight state-of-the-art examples for traditional and novel RS applications enhanced by DL.

These four problems have been addressed with CNN algorithms in different areas of RS research in the past. Xing et al. [146] used thematic image classification (problem 1) to classify the land cover seen on geo-tagged photos. In combination with the photos' distance to pixels in the GlobeLand30-2010 land cover map, the photos' classification results were used to validate the map product's accuracy. By combining geo-tagged GSV imagery with DL, Kang et al. [66] were able to perform thematic image classification (problem 1) of individual buildings and map them in space. DNN were also used to extract the position of multiple smaller objects and features (problem 2) from aerial and satellite imagery, including vehicles [23], aircrafts [20], oil tanks [151] and sport fields [24]. Wegner et al. [144] designed a workflow for automatic detection and geolocation of street trees, by combining FASTER R-CNN bounding box detection (problem 2) scores from GSV and aerial imagery, with information retrieved from Google Maps in a probabilistic model. The authors then used street-level and aerial imagery to classify 18 different species among the detected trees (problem 1). Branson et al. [15] subse-

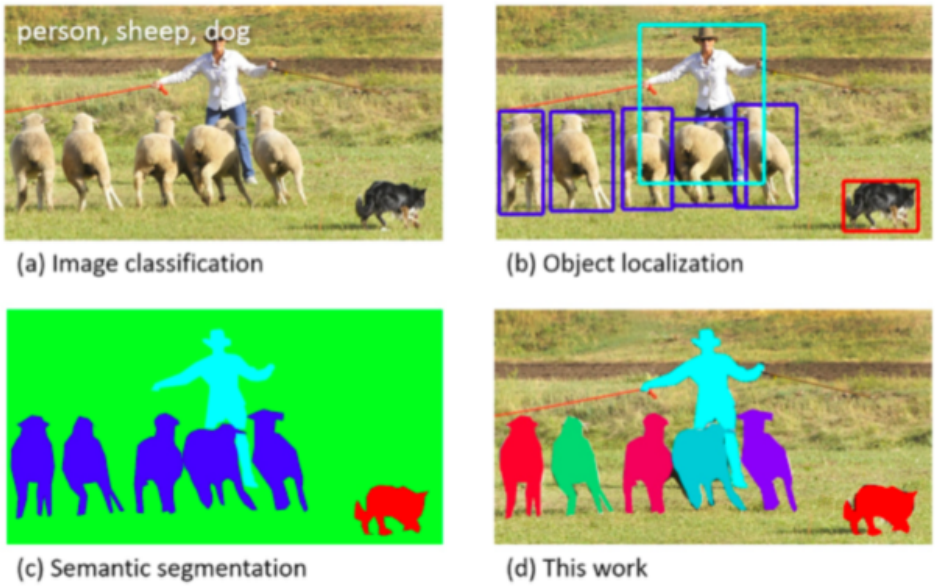


Figure 1.4: The four tasks of computer vision. This thesis is based on instance segmentation (d). [83]

quently built upon methods for object detection in [144] by including a Siamese CNN, to verify whether detected trees had changed visibly over time. Pixel-wise semantic segmentation (problem 3) with DNN has been dominating the International Society for Photogrammetry and Remote Sensing (ISPRS) semantic segmentation challenge, and is increasingly used for a variety of land cover classification projects [65]. Despite the proven suitability of DL for semantic segmentation, research applying CNNs to quantify urban greenery from street-level imagery is sparse. Cai et al. [19] recently tested tree canopy segmentation (problem 3) with different CNNs and estimated the GVI with a custom architecture based on a Residual Network with 50 Layers (RESNET50).

The fourth problem (instance segmentation, i.e. pixel-wise detection of separate objects of the same class) is challenging, and DNN frameworks have only recently shown great potential for this task [57]. Detecting and masking each distinct object in an image for 'Stuff classes' (fuzzy object classes without clearly delineated shapes, like the sky, trees or other vegetation) was brought to the at-

tention of the DL community via the 2016 Common Objects in Context (COCO) Stuff Challenge [18]. Models such as Mask Region-based Convolutional Neural Network (MASK R-CNN) have since had great success in performing instance classification on Stuff classes, which opens the door for assessing and mapping the rich data on fuzzy objects contained in side-view imagery. Combining street-level imagery and DL techniques to analyse urban features and objects is a promising avenue in urban research.

1.3 Research questions and research design

Limited awareness of cost-efficient street-level imagery and DNN potential to support environmental management and policy making constrains the utilization of these data and technologies. The direct and automated implementation of CNN architectures and street-level imagery as a promising tool for high-resolution data generation for decision support in smart urban forestry management is sparse.

The main objective of this thesis is to develop a cost-efficient and automated approach to generate fine-scale tree inventory data in urban areas. The research aims to assess the potential and advantages of an open source DL based approach for decision support in smart urban forest management in general and in FIS management in particular. Therefore, the thesis explores the potential to use readily available street-level imagery in combination with new and emerging open source tools. A case study demonstrates how an automated data generation approach could assist bio-surveillance procedures in Metro Vancouver, Canada.

1.3.1 Research questions

Specifically, the project addressed the following questions:

1. How can deep learning algorithms assist in improving tree detection in the urban landscape?

2. How can monocular depth estimation be combined with tree detection for urban tree geolocation from single street-level images?

3. How can urban trees be classified using emerging deep learning techniques and what are potential applications for smart urban forest management?

1.3.2 Research design

This research project explores the potential to extract valuable information about urban forests from street-level imagery using DL techniques. An approach to automatically generate fine-scale urban tree inventory data is investigated and the potential for implementation of the generated information in decision support for bio-surveillance efforts is outlined. Novel CNN architectures are implemented in three different modules that constitute an open source software package (s. fig 1.5) implemented in a workflow for urban tree location and genera mapping. One module (Detection) is for the task of tree instance segmentation on street-level imagery, a second module (Geolocation) for the task of location prediction with monocular depth estimation for detected trees (s. chapter 2) and the third module (Classification) is for the tree genus classification based on street-level imagery (s. chapter 3). Inference results generated through the three modules are combined to map urban tree genera distributions in the Metro Vancouver region. The final workflow is then evaluated for implementation in decision support for bio-surveillance (s. chapter 3).

A more detailed description of how street-level imagery and benchmark datasets (s. section B.1.1) were acquired and processed (s. section B.1.2) for training and evaluation purpose can be found in the appendix B. Details about the base architecture MASK R-CNN of the proposed workflow (s. section B.2.1), evaluation strategies to assess the suitability and performance of the chosen architecture (s. section B.3), and the training strategy used for tree detection (s. section B.4) can also be found in the appendix.

The three research questions outlined in section 1.3.1 were tested with the methodological workflow depicted in figure 1.6. The proposed experimental workflow proceeds in four stages. Stage one (s. fig 1.6, orange) is characterized by data acquisition and pre-processing. Developing a reproducible Python-based pipeline for automated GSV image acquisition allowed for a user-friendly download of open source imagery from any desired area for any desired view. Figure 1.7 shows a typical GSV scene (right), used in the analysis. Acquired imagery was further combined with tree location and genus information from the existing street tree inventories in Vancouver in order to build tree genera classification datasets. All datasets were

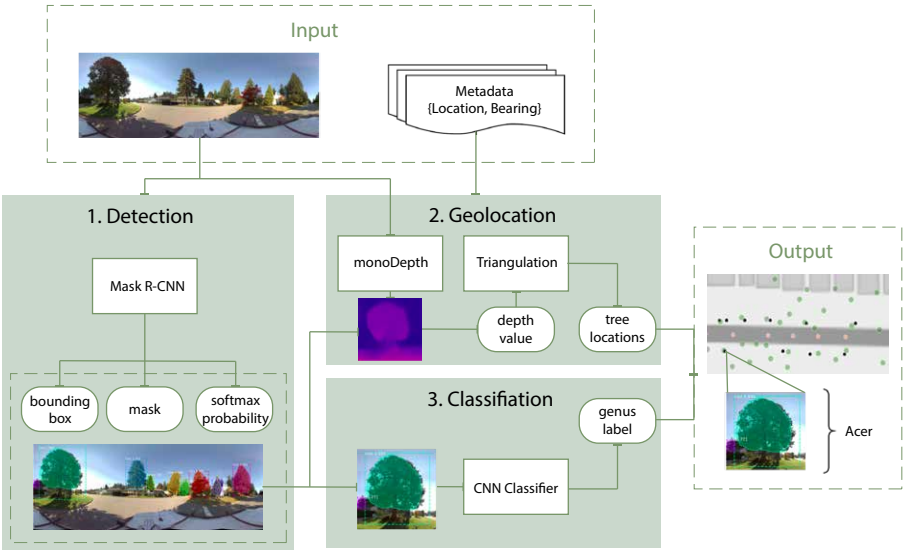


Figure 1.5: AiTree: Open source software for urban tree mapping. The full software will be published under *github/slumnitz/aiTree* and contains three modules. One module for the task of tree instance segmentation on street-level imagery (1), a second module for the task of location prediction with monocular depth estimation for detected trees (2) and the third one for tree genus classification on street-level imagery (3).

separated for the use as training data, development data and as test data (explained in stage three).

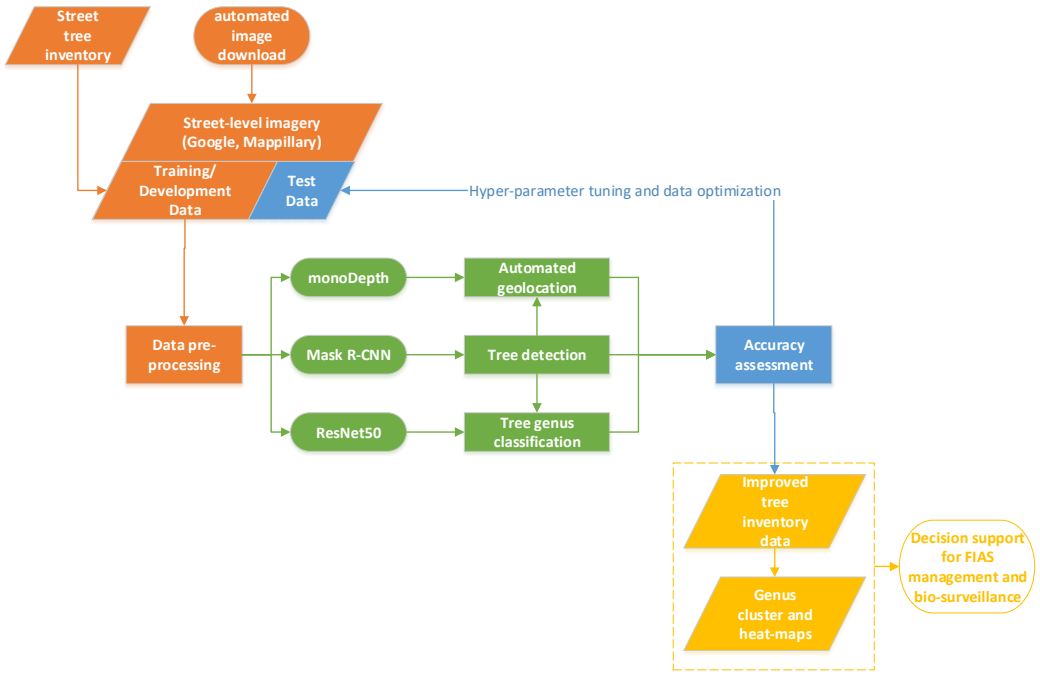


Figure 1.6: Methodology for the development of trained deep neural network models for automatic tree detection, geolocation and genus classification. Stage one (orange) is characterized by data acquisition. Stage two (green) includes the design and training of tree detection, geolocation and classification models with the MASK R-CNN, monodepth and RESNET50 architectures. In stage three (blue), the accuracy of the trained neural network models is assess through development and test datasets. In a last step, the software and workflow are tested for inference on Metro Vancouver imagery and decision support for bio-surveillance efforts (yellow).

Figure 1.7: Google Street View data. Current tree inventory data and GSV camera positions depicted on Google Maps imagery from 8th West Avenue, Vancouver (left). Street trees and trees on private property seen in GSV imagery (right). (Source: Google Maps, 2018; Google Street View; 2018)

Stage two (s. fig 1.6, green) included the design, training and deployment of tree detection and classification models built on top of three different CNN architectures (s. fig 1.5). The final scientific software was implemented in Python and will be made available as open source software on github. The software was primarily based on instance segmentation done with the MASK R-CNN architecture. MASK R-CNN is currently the most promising openly available framework for the core instance segmentation model of this tree inventory generation workflow. At present, MASK R-CNN is the best performing framework able to carry out both object classification and pixel-level instance segmentation [57]. MASK R-CNN allows for the automatic generation of bounding boxes, shape masks and classification scores with only one CNN architecture (s. fig 1.5, Detection). For more detail on

the MASK R-CNN architecture and training see appendix B.2.1. Furthermore, the design of the proposed workflow can be easily adapted replacing MASK R-CNN, depending on the accuracy of results or improvements in published state-of-the-art algorithms. The trained MASK R-CNN model was utilized to classify tree instances in GSV and Mapillary images from a human's perspective (s. section 2.2.4). Generated bounding boxes and tree shapes were then used for extracting tree images of interest that were fed into the subsequent geolocation (s. section 2.2.5) and classification modules (s. section 3.2.5). Stage two, in which detection, classification and geolocation models were primarily designed and trained, was continuously informed by stage three, in which model performance was assessed. Model training in stage two was then repeated until model performance was optimized.

In stage three (s. fig 1.6, blue), model performance was analyzed and optimized. Pre-processed data were split into training, development, and test datasets. During training, a split dataset was required to test for both bias and over-fitting. Model accuracy was assessed using both development and test datasets. For each model (tree detection, geolocation, and genus classification), different optimization measures (and their corresponding accuracies) were tested in order to find the best set of model hyper-parameters, training data, and optimization algorithms and achieve the highest possible model performance. The most accurate models were used for inference and an overall workflow accuracy was calculated. Achieved accuracy for all models were compared to state-of-the-art models or human-level performance and used as an indicator for performance of the proposed methodologies. Software and model output data were exported as geographic data stored in *.csv*, *.shp* or *.geo json* format.

In stage four (s. fig 1.6, yellow) the software was used to collect tree inventory information for the Metro Vancouver region. The generated tree inventory information was tested for implementation in bio-surveillance efforts in the Metro Vancouver region. The location and intensity of tree genera hotspots are reported in chapter 3.

1.4 Thesis structure

In chapter 2 the tree detection and geolocation methodology and workflow are described and model performance is assessed. In chapter 3 the tree genus classification model is presented and evaluated for the purpose of generating estimates of tree genera accumulation in the Metro Vancouver area. To conclude, key findings, implications, limitations and future work is discussed in chapter 4. The appendix contains additional information

Chapter 2

Mapping urban trees with deep learning and street-level imagery

2.1 Introduction

2.1.1 Urban tree assessment

Urban forests are gaining global attention as evidence is gathered about the diverse benefits they provide to human health and well-being through various ecosystem services [106, 140] (s. section 1.1). Planning and managing urban forests and trees on the basis of urban tree inventories is increasingly coming to the fore in the context of global urbanization trends, rapid climate change and increasingly connected trade [110]. Nevertheless, urban forest planning and management remains an outstanding challenge worldwide owing to relatively scarce information on the spatial distribution and accessibility of urban forests and trees, as well as their health condition, composition, structure and function [69]. In practice, most municipalities still perform labor-intensive field surveys to collect and update inventories of public trees. Despite the importance of urban trees, national and municipal sources of tree inventory lack in detail, consistency and quantity due to the cost associated with mapping and monitoring trees through time and over large areas [102] (s. section 1.1.3).

The study aims to produce an automatic, affordable and novel method for tree detection and geolocation that can be used in any urban region where sufficient street-level imagery is readily available. I introduce state-of-the-art instance segmentation (object detection and pixel masking) with DL frameworks to extract and mask fuzzy features like trees in images (s. section 2.2.4 and appendix B.2.1). In addition, the novelty of this method is enhanced by using monocular depth estimation and triangulation to estimate precise tree locations without the need to rely on secondary datasets (s. section 2.2.5). Ultimately, I aim to fulfill the need for inventory methods that can be automated and generalized over multiple cities in order to develop national and international recommendations and standards for urban tree inventory data collection [102] (s. section 1.1.3).

2.1.2 Remote sensing for individual tree mapping

Recent research has focused on RS data and techniques allowing for the remote and automated recognition and characterization of individual trees [67]. Individual tree mapping from remotely-sensed data, termed as Individual Tree Crown Delineation or Detection (ITCD), has gained popularity since the mid-1980s as an alternative to ground-truth measurements [153]. However, mapping and monitoring of individual trees in heterogeneous urban areas using remotely-sensed data and current ITCD methods remains challenging [10]. The small size of individual tree crowns in urban areas binds the use of most satellite imagery sources to analyzing clusters of urban trees or requires a process for spectral unmixing [129]. VHR satellite or aerial imagery ($<$80 centimeter) can help provide the level of detail required for individual urban tree assessments, but are often impacted by urban shadows [33, 82]. Similarly, the use of high-resolution LIDAR data in individual tree assessments is often impacted by vertical urban structures such as power lines and lamp posts [153]. Datasets such as LIDAR or VHR aerial imagery are usually collected at one-point in time and can be expensive to acquire [6, 82]. Novel, readily-accessible methods and data sources to build standardized tree inventories on a large spatial scale allowing for cheap, seamless and recurrent data collection and rapid processing are still needed [67].

2.1.3 Trends in automatic tree inventory assessments

Two recent trends have gained attention to filling the gap in assessing urban trees over large areas at low cost, and promote uptake from a larger number of municipalities in recent literature [132]: first, the success of CNN out-competing other methods for extracting abstract features and objects in imagery [89] (s. section 1.2.2), and second, the growing availability of low-cost, detailed and increasingly crowd-sourced street-level imagery (photographs of street scenes taken from the ground) [13, 77] (s. section 1.2.1).

Street-level imagery is, for example, used to quantify 'perceived urban canopy cover' by estimating the percentage of detected tree canopy cover pixels relative to the total number of pixels in an image [124]. Similarly, Li et al. [80] assessed the percentage of vegetation in streets by quantifying the amount of green pixels seen in a street view scene. Both of these methodologies calculated a GVI, a metric that quantifies the proportional amount of green pixels in each image. The index serves as a proxy for how urban vegetation is perceived by pedestrians, and has since been applied to a variety of cities all over the world [36]. Wegner et al. [144] designed a workflow for automatic detection and geolocation of street trees, by combining FASTER R-CNN tree detection results from GSV and aerial imagery, with information retrieved from Google Maps in a probabilistic model. The authors then used street-level and aerial imagery to classify 18 different species among the detected trees. Branson et al. [15] subsequently built upon methods for object detection in [144] by including a Siamese CNN, to verify whether detected trees had changed visibly over time. Both approaches, however, still rely on a data fusion approach with VHR aerial imagery to locate urban trees.

Localizing objects that have previously been detected in photographs acquired with smart phones or cameras from the street is a unique challenge for RS practitioners. Satellite or aerial imagery pixels and LIDAR point clouds inherently store either relative or absolute geographic location information and make the need to additionally compute three dimensional geographic pixel coordinates redundant. The translation process for features from street-level imagery to a geographical location is usually achieved using one of two principal approaches: (1) objects are either matched to locations using overhead or 3-dimensional data (e.g. passive aerial

imagery or active LIDAR) [76], or (2) the location is directly retrieved from street-view imagery by reconstructing 3-dimensional space or feature data (e.g. camera-to-object depth) [5]. The latter approach can be achieved through multi-view stereo methods (using multiple images to reconstruct the objects) [25], binocular methods (using two images) [59] or monocular methods (using only one image) [50]. Because monocular depth estimation can be made using a single image, it does not require a large amount of images taken from multiple perspectives or additional knowledge about the analysed scene, and allows for the analysis of features from various data sources taken at different points in time [134]. Monocular depth estimation has benefited from recent development of novel DL approaches, particularly in the field of self-driving cars [92]. The potential to retrieve location information of detected objects from a single image through the use of monocular depth estimation enhances the potential to use street-level imagery collected with different sensors over time to match detected objects to specific locations [144].

2.1.4 Chapter objectives

In this chapter, I propose a novel, low-cost method for urban tree detection and geolocation using readily available geo-tagged street-level images. I investigate the generalizability and transferability of the tree detection model by applying it to different geographical locations in three cities in the Metro Vancouver region (Canada) and the city of Pasadena (US). I further test the robustness of the approach on images provided by two different street-level imagery sources, namely GSV, which provides proprietary data, and Mapillary, which offers crowd-sourced data. I validate the model by comparing its output with on-the-ground tree location measurements in the Metro Vancouver region. Ultimately, the aim is to produce a robust, cheap and rapid method for creating detailed tree inventories in any urban region where sufficient street-level imagery is readily available.

2.2 Data and methods

I started by detecting trees and generating tree instance masks in all available panorama imagery prior to mapping detected trees in space using two DL architectures (s. fig 2.1). DL has proven to be a powerful tool for extracting abstract

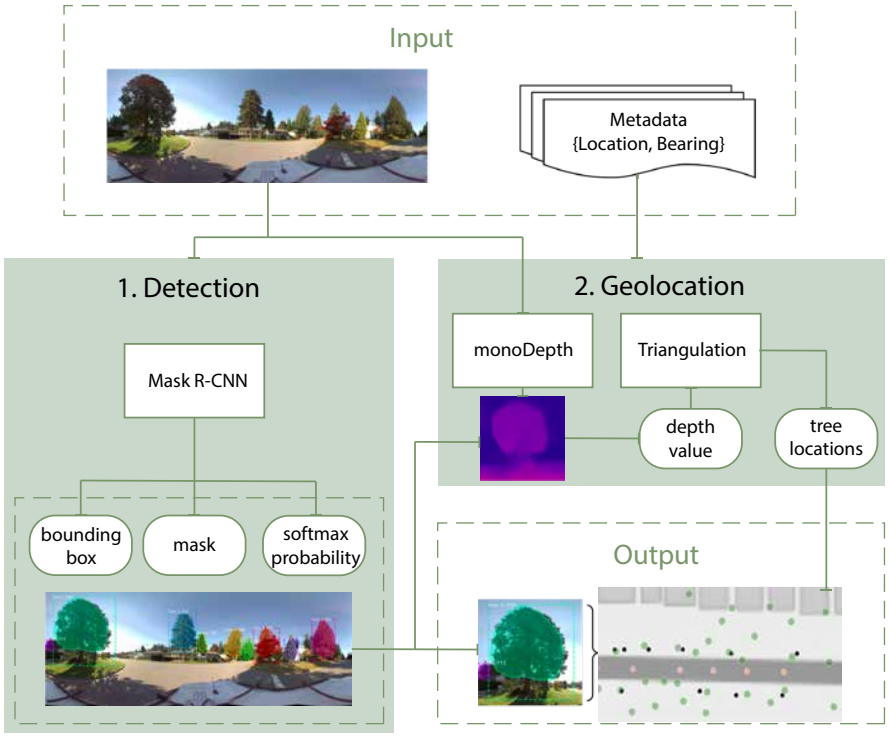


Figure 2.1: Urban tree mapping workflow: I first generate a bounding box, a mask and a probability score for all urban trees for the input imagery using the trained MASK R-CNN algorithm. I then compute a dense depth mask with monoDepth for the same input imagery and extract a depth value for every previously generated tree mask. I use the depth value, and imagery metadata in the triangulation pipeline to generate tree locations as geographic coordinates. The output is a map of urban tree positions connected to generated tree masks.

features and objects from raw imagery, and is increasingly adopted in the RS community [154]. The most common algorithmic architecture to implement DL for analyzing imagery are CNNs (s. section 1.2.2). Zhang et al. [152] provide a detailed overview and a technical tutorial for RS data analysis using DL algorithms and Mountrakis et al. [98] highlight state-of-the-art examples for traditional and novel RS applications enhanced by DL. CNNs were, for example, used to extract the position of multiple smaller objects and features from aerial and satellite im-

agery, including vehicles [23], air crafts [20], oil tanks [151] and sport fields [24]. Xing et al. [146] used CNNs to classify the land cover seen on geo-tagged photos. In combination with the photos' distance to pixels in the GlobeLand30-2010 land cover map, the photos' classification results were used to validate the map product's accuracy. The combination of street-level imagery and DL is also showing promising applications in the urban context. Kang et al. [66] classified the use of buildings and mapped them in space by combining geo-tagged GSV imagery with DL. The approach assesses the potential to map urban trees in four areas of interest and consists of the following steps (s. fig 2.1):

1. Street-level imagery retrieval for areas of interest (s. section 2.2.3)
2. Tree instance segmentation with MASK R-CNN architecture (s. section 2.2.4)
3. Geolocation of trees detected in panoramas through monocular depth estimation and panorama metadata (s. section 2.2.5)
4. Geolocation correction of trees present in multiple panoramas through triangulation (s. section 2.2.5)

2.2.1 Study site

For the majority of model training and assessment, I chose imagery and ground-truth measurement plots distributed over the Metro Vancouver area (49° N, 123° W), specifically in the municipalities of Vancouver, Surrey and Coquitlam (s. fig 2.2). Metro Vancouver is located on Canada's south-west coast, being one of the warmest Canadian cities in winter and experiencing relatively high rainfall rates throughout the year. Mild climatic conditions are favouring growth and survival for tree species from harsher and milder climatic conditions [131]. Metro Vancouver's urban forest includes many exotic tree species imported from different climatic zones in North America as well as over 60% of Canada's native tree species resulting in one of the most diverse urban forests in Canada [130]. Metro Vancouver strives to be one of the world's greenest cities by 2020, resulting in spatially varying types of proactive urban forest management [62] (s. section 3.2.1). To demonstrate the potential transferability of the model to other urban ecosystems, I

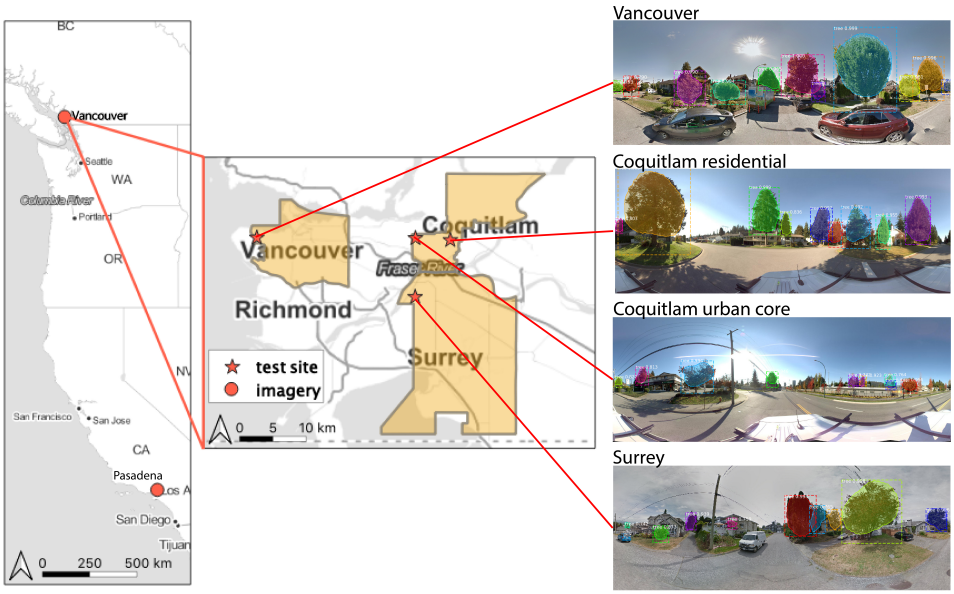


Figure 2.2: Location of street-level imagery datasets and ground-truth measurements. Street-level imagery is available for Metro Vancouver and Pasadena. Geolocation test sites are located in the cities of Vancouver, Coquitlam and Surrey, Canada. (Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL. Source: Mapillary 2019, Google Street View 2018)

evaluate the trained tree instance segmentation model on imagery of the Pasadena Urban Tree dataset (34° N, 118° W), located 2000 kilometer further south in the west coast of the United States (s. fig 2.2).

2.2.2 Ground-truth measurements

To evaluate the model's geolocation performance, I conducted a field campaign in March 2019 to collect ground-truth location measurements of all public and private trees in four areas of interest: Vancouver, Surrey and two in Coquitlam (urban and suburban areas) (s. fig 2.2). I define a tree as any vegetation with a clearly distinguishable stem and crown, that has the potential to grow over 5 meters in height with full maturity [103]. I recorded the Global Positioning System (GPS) positions of each visible tree from the sidewalk using a Trimble Geo 7X handheld

device in an unobstructed position to maximize GPS signal. I used the rangefinder on the handheld device to measure the offset between the measurement position and the tree, from which a precise tree location could be determined during post-processing. The measured GPS locations were corrected using Base Station Data for Vancouver (BCVC: 491632.73535 N, 123521.58021 W) and Surrey (BCSF: 491131.49655 N, 1225136.24849 W). This provided us with an overall accuracy of under 0.5 meters for 294 recorded trees in Vancouver, 152 trees in Surrey, and 336 trees in Coquitlam.

2.2.3 Street-level imagery

It is common practice in training a CNN to divide a dataset into training, development and testing datasets [122]. This practice ensures that model parameters are adjusted to support the best possible generalization of the model and its applicability to various datasets and tree objects. Given the range of imagery providers, cameras used and quality of street-level imagery, I decided to test the transferability of the model between two different data providers: proprietary GSV data (Vancouver, Surrey, Pasadena) and crowdsourced Mapillary data (Coquitlam).

Building training and development datasets

In total, I compiled two datasets for two separate training steps described in section 2.2.4. I split each of the compiled datasets with a 80:20 ratio into training and development datasets. First, I extracted 36,500 images from the openly available COCO Stuff semantic segmentation dataset that contains semantic labels (classified pixels) for trees. The COCO Stuff dataset is, to date, the most expansive collection of images with semantic segmentation labels ($\sim$164,000) for "amorphous Stuff" classes (e.g. sky, roads, brick walls, trees, etc.) [18]. In contrast, most datasets focus on clearly delineated "thing" classes (e.g. people, cars, traffic lights, etc) [14]. Second, I acquired GSV images from Vancouver and Surrey in March 2018 (s. fig 2.2, tab 2.1). I used the "Labelbox" web tool to create single tree instance labels for combined 60 images for Vancouver and Surrey by manually masking all visible trees, resulting in approximately 1200 tree masks (i.e. instance labels), 453 for Vancouver and 711 for Surrey [73].

Table 2.1: Street-level imagery and mask annotations for the fine-tuning procedure. Compiled datasets consist on 30 panorama images each.

	dataset	provider	tree masks	green infrastructure
Vancouver	train/dev	Google	453	street and private trees
Surrey	train/dev	Google	711	private trees
Pasadena	test	Google	365	street and private trees
Coquitlam	test	Mapillary	471	street and private trees

Building test datasets

To assess the models generalization performance and transferability to imagery of a different urban ecosystem, I evaluated the model's performance on an independent test dataset consisting of imagery from the city of Pasadena (s. tab 2.1). The dataset, which covers all of Pasadena, was created in March 2016 and is available from Branson et al. [15]. In addition, I used imagery acquired from Mapillary for the city of Coquitlam as a second independent test dataset to demonstrate the robustness of the model applied to different street-level imagery providers (s. fig 2.2 and tab 2.1). For Coquitlam, I downloaded panorama images in February 2019 [133]. I randomly sampled 30 test images from both datasets using the NumPy random number generator (assuming a univariate Gaussian distribution) in order to test imagery from different types of city structure. I masked and annotated approximately 360 individual tree masks for Pasadena and 470 for Coquitlam using the same method for labeling Vancouver and Surrey imagery described above. All panorama imagery contained metadata about the camera location of capture, and a 360° bearing reference to true north.

2.2.4 Tree instance segmentation

Tree instance segmentation model

I trained the MASK R-CNN architecture for tree instance segmentation, the task of pixel-wise detection and delineation of separate objects of interest in an image. Instance segmentation for "Stuff classes" (fuzzy object classes without clearly delineated shapes, like the sky, trees or other vegetation) is technically challenging

and has only recently gained attention in the DL community, resulting in a relative scarcity of architectures performing well for this task [18]. I chose MASK R-CNN due to its generality, flexibility and the best performing architecture in the recent COCO 2017 Instance Segmentation Challenge [14]. MASK R-CNN is a state-of-the-art architecture that is implemented in the model in a modular way so that it can be replaced, by surpassing instance segmentation algorithms in future. The implementation of MASK R-CNN adopts the original framework He et al. [57] implemented in Python 3, Keras and TensorFlow and can be accessed through Abdulla [4]. Generated outputs include: (1) bounding boxes in pixel coordinates around each detected tree object, (2) a probability score of the class label assigned to a detected object (binary: tree or non-tree), and (3) a pixel mask through the assignment of single pixels to individually detected object [57]. For more detailed information about the architecture of MASK R-CNN c.f. He et al. [57] and supplementary material in appendix B. After finalizing training, the MASK R-CNN model was applied to detect and mask new tree objects in images which were not exposed during the previous training step, which I refer to as the process of inference (s. section 1.2.2) [51].

Training strategy

I used all three, transfer learning, a layered-training approach and fine-tuning to train the MASK R-CNN architecture [26]. Through transfer learning, a model pre-trained for one task (e.g. semantic segmentation of "thing" classes in the COCO dataset) is re-trained for another task (e.g. tree instance segmentation on street-level imagery) through the transferal of weights [111][14]. I transferred weights (i.e. feature representations) of a MASK R-CNN model pre-trained on the COCO dataset to the first (deep) layers of the fresh model, and initialized the training process with these COCO weights. In this way, the model first learned to distinguish tree structure in general and was later able to separate single trees more effectively [19].

In the subsequent training iteration, defined as the layered-training approach, I used images containing tree objects in the COCO Stuff dataset. For this, I trained the last 5+ top layers of MASK R-CNN with 50 epochs on COCO Stuff using a learning rate of 1e-4. An epoch refers to all images in the training dataset being run through

the entire model and the internal model parameters being updated at least once [51].

Finally, I fine-tuned the model by training with the labeled data from Vancouver and Surrey. I therefore trained the model heads (the most shallow or last layers of the model) with 30 epochs, followed by another iteration training +5 layers of the Residual Network with 101 Layers (RESNET101). After a sparse grid search I found 1e-4/10 to be the most successful learning rate to use for the fine-tuning step. I set all other hyper-parameters following recommendations in existing research that employs MASK R-CNN [4, 57].

To avoid over-training the model on relatively few training samples from Surrey and Vancouver (1000 tree instances), I used heavy data augmentation at train-time during the fine-tuning procedure. In brief, I split the panorama images in half at the start of each training epoch, downscaled the halves to 1024x1024 pixels in size. Each time an image was loaded into memory: (1) I flipped the images left to right 50% of the time; (2) I either re-scaled the image in the x and/or y direction by a variable factor between 0.8 and 1.2, rotated the image with a random angle between -4 and +4 degrees, or sheared the image with a random angle between -2 and +2 degrees; (3) and performed contrast normalization using a random target factor between 0.9 and 1.1 of the initial contrast of the image. The full model was trained on a NVIDIA GTX1080 Ti Graphics Processing Unit (GPU), and limited by 11GB of memory to stochastic gradient descent, or a mini-batch size of 1.

Evaluation strategy

I evaluated the method on two development datasets (Vancouver and Surrey, using GSV imagery) and two test datasets (Coquitlam using Mapillary, and Pasadena using GSV imagery). To evaluate model performance, I chose three commonly used evaluation metrics for instance segmentation frameworks [14]: (1) mean average precision (*mAP*), (2) average precision over Intersection over Union (IOU) using a threshold of 0.5 (AP_{50}), and (3) average precision over IOU with threshold 0.75 (AP_{75}) (s. appendix B.3.2). To quantify the known, negative influence of small tree mask sizes on instance segmentation performance in detail, I iteratively excluded all smaller masks under a mask size threshold and compared recalculated

evaluation metrics [70]. Next, I manually inspect failure cases according to three different tree mask sizes to identify the most frequent tree detection error. For detailed tree instance segmentation error assessment, I defined small masks to be under 3000 pixels in size (approximately 0.3% of total image pixels per small mask), representing detections of very distant trees ($>$ 70 meters) relative to the camera position at image capture. I defined medium masks to be between 3000 and 30,000 pixels in size, roughly all private trees, found in front yards, and big masks to be over 30,000 pixels in size (approximately 3% of total image pixels per big mask), repetitive to large front yard trees, or street trees.

Additionally, I calculated precision and recall to compare detection results with annotated images [34]:

$$precision = \frac{tree_{pred}}{tree_{pred} + other_{pred}} \tag{2.1}$$

$$recall = \frac{tree_{pred}}{tree_{annotated}} \tag{2.2}$$

All final evaluation metrics and precision-recall curves to compare model performance for different datasets were calculated excluding very small masks under a size of 3000 pixels (approximately 0.3% of total image pixels per small mask) [18].

2.2.5 Geolocation of trees

Depth Estimation

To geolocate individual trees, I first created a dense depth estimate layer for each panorama using monocular depth estimation [50]. I adapted Godard et al. [50]'s monocular depth estimation architecture, MonoDepth, to develop dense depth masks because of its applicability to images with varying lens types (e.g. panoramic or narrow view) typically found in street-level imagery datasets [132]. MonoDepth is available off-the-shelf as a fully trained unsupervised DL model with a depth error margin of less than 20%. Godard et al. [50] provide multiple trained weights for non-commercial usage, which allows researchers to use the model for infer-

ence purposes without having to perform laborious training stages. For the analysis, I followed Godard et al. [50]'s recommendations to adopt the best performing weights when MonoDepth was pre-trained on the Cityscapes dataset and fine-tuned using the KITTI vision benchmark dataset [29, 46].

The depth estimation model typically computes disparities (D) between objects in each panorama image which then need to be translated by a factor (F) into meters. During the translation process, values need to be calibrated for the specific lens (C) used and corrected for differences in image sizes between the original Cityscapes and KITTI datasets (W_0) and the input image (W_1). Depth can then be interpreted as per pixel depth estimate ($depth$) between the object in the image and the camera position at the time of image capture:

$$depth = \frac{W_0 * F}{(W_1 * D) * C} \tag{2.3}$$

Based on Godard et al. [50], I set F to 0.54 meters, W_0 to 721 pixels, W_1 to 6656 pixels and C to 1.5 to account for the use of a camera lens that captures panorama images. The calibration parameter C was determined by minimizing the average error in the monocular depth estimates compared to distances of measured ground-truth trees to the recorded camera position. Once dense depth layers were computed, I extracted a single depth value in meters for each previously detected tree using depth pixel values at the center of mass calculated for each instance mask. Finally, I translated each observation of a detected tree into geographic coordinates by combining the depth value, the bearing of each tree instance in respect to the position the panorama was captured, and the panoramas coordinates. Both of the later are recorded in the imagery's metadata.

Triangulation

In a subsequent processing step, triangulation is used to reduce duplicate observations of individual tree predictions and correct their position estimation where multiple observations of the same tree are recorded (s. fig 2.3 (a)). I assumed that there are no false-positive tree instances (i.e. each tree detected as an object is a real tree). Through triangulation, I created nodes of intersections for all crossing edges, drawn between preliminary tree prediction positions and the correspond-

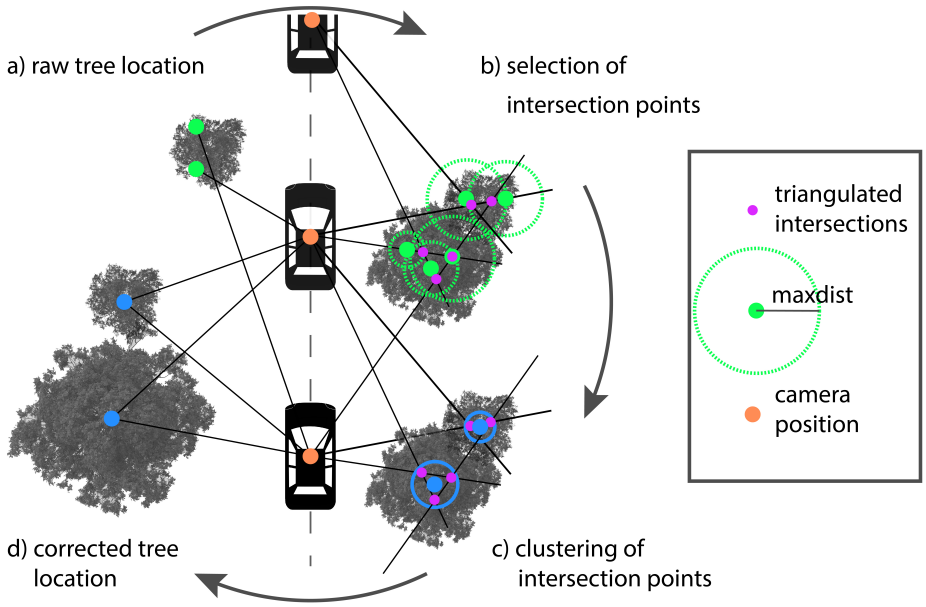


Figure 2.3: Correction of predicted tree locations through triangulation. I draw bearings between raw tree locations from monocular depth estimation and camera positions (a). The triangulated intersections of these bearings are selected if they are located within a maximum distance (*maxdist*) from raw tree locations, for a minimum of two raw predictions (b). The intersection which are the closest to the raw tree positions are chosen and clustered (c) to create the final corrected tree position (d).

ing camera location. I selected candidate intersections within a maximum distance (*maxdist*, s. equation (2.4)) to the preliminary tree location estimate for at least two preliminary tree locations (s. fig 2.3 (b)).

$$maxdist = c_0 + c_1 * depth \tag{2.4}$$

c_0 is a constant offset in meters, and c_1 describes the maximum relative error in the depth estimate, calculated as 65%. I chose a value of 3 meters for c_0 to account for the average inaccuracy in the GPS positioning of panorama locations.

Given that each edge between the preliminary tree position and the correspond-

ing panorama location has potentially multiple candidate intersections, I selected the closest intersection to each preliminary tree position. Finally, I used hierarchical clustering to assign the average position of all selected candidate points in the cluster as an output tree coordinate (s. fig 2.3 (c), (d)). I analyzed the distances of all ground-truth measurements with respect to each other, and found that over 99% of all points were separated by at least 3 meters. To avoid multiple detections of the same tree represented by multiple candidate intersection points, I chose a 3 meter threshold for the clustering as the minimum distance between observations.

2.2.6 Model evaluation

I used measured public and private tree positions to evaluate the tree location model performance for areas of interest in Vancouver, Surrey, an urban and a suburban area in Coquitlam. I evaluated the absolute positioning error of tree predictions as the euclidean distance between the ground-truth measurement and the tree location predictions [148]. I used a greedy algorithm to assign closest matching trees first, and then took matched trees out of the running process until no ground-truth measurements were left to match [144]. A match is kept as a true positive match if the distance between ground-truth measurement and tree prediction does not exceed 15 meters. A 15 meter threshold was chosen to include as many private trees as possible in the error assessment, typically found in a 15-30 meter range from the camera position at time of image capture. Trees located more than 15 meters from the measured private ground-truth data represent distant tree detections behind houses without comparable measured ground-truth data and were excluded in the error asssessment by the given threshold. I defined a measure of absolute tree positioning accuracy as the mean of absolute positioning errors $mean_{epos}$ [148]:

$$mean_{epos} = \frac{1}{n_{TP}} \sum_{i=1}^{n_{TP}} \sqrt{(x_i - x_{predi})^2 + (y_i - y_{predi})^2} \tag{2.5}$$

I evaluate absolute tree positioning accuracy and the ratio of matches to non-matches for public, private and all trees respectively.

2.3 Experiments

2.3.1 Instance segmentation

Effects of layered training

Performing layered training with COCO Stuff images improves the detection of tree objects from ∼80% without COCO Stuff training to a slight overcounting of trees with 103% detected trees compared to the labeled tree masks in the combined Vancouver and Surrey datasets (s. tab 2.2). The overall model accuracy with AP_{50} and mAP improves slightly while values for AP_{75} decline. The small decrease in mask accuracy for AP_{75} may be a direct result of the layered model including detections of trees which are more difficult to detect compared to the lower number of detections when layered training is not used.

Table 2.2: Evaluation metrics for training with and without COCO Stuff on combined Vancouver and Surrey dataset

	AP_{50}	AP_{75}	mAP	mask predicted	mask annotations
COCO only	0.608	0.252	0.281	74	90
COCO Stuff	0.661	0.199	0.290	93	90

Transferability between different ecosystems and data sources

I find that MASK R-CNN developed on Vancouver and Surrey training datasets was successfully applied to detecting trees across all four datasets (s. fig 2.4 and tab 2.3). AP_{50} values ranging from 0.620 to 0.682 and values of other evaluation metrics are consistent with current tree or plant semantic segmentation performances found in other studies, with the difference that I not only evaluate pixel-based classification results, but also distinguish between different tree objects [18] (s. tab 2.3). AP_{75} and AP_{50} values are lowest for Surrey (0.157, 0.262) and highest for Pasadena (0.262, 0.316) and Coquitlam (0.261, 0.342) datasets. Overall, at a recall threshold of 0.6, precision is above 0.8 for all four datasets (s. fig 2.4). With a recall of 0.35 or higher, precision-recall curves for both development datasets are slightly higher than for the testing datasets, which is to be expected since assessed

features in the development datasets directly influence the training process (s. fig 2.4) [75].

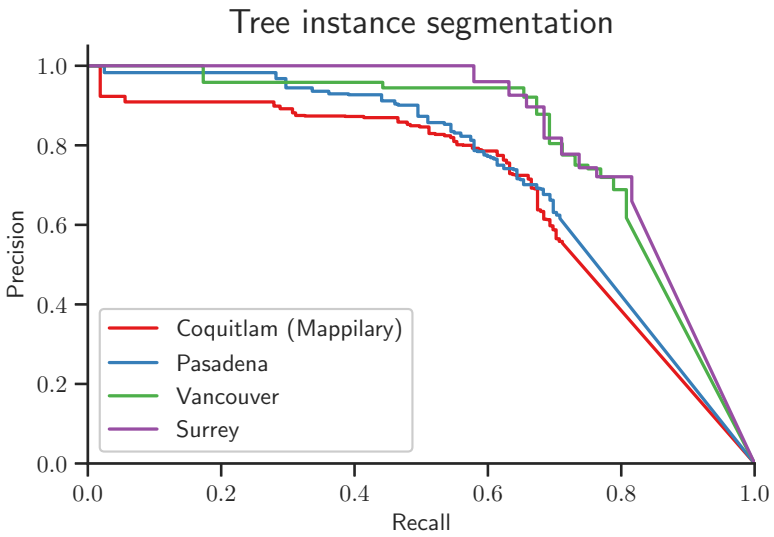


Figure 2.4: Precision-recall curves for development (Surrey, Vancouver) and test (Coquitlam, Pasadena) datasets.

Table 2.3: Evaluation metrics for Vancouver, Surrey and Pasadena. Excluding masks under 3000 pixels in size.

	AP_{50}	AP_{75}	mAP	mask predicted	mask annotations
Vancouver	0.682	0.232	0.312	53	52
Surrey	0.634	0.157	0.262	40	38
Pasadena	0.628	0.262	0.316	194	202
Coquitlam	0.620	0.261	0.342	238	215

MASK R-CNN performance for the Pasadena and Mappillary test datasets is very similar (mAP) to slightly better (AP_{75}, AP_{50}) than that of the Vancouver and Surrey dataset (s. tab 2.3). Slight differences in precision-recall curves and variations in evaluation metrics may be attributed to overall varying tree shapes and sizes found in each dataset. Features learned throughout the trained MASK R-CNN model appear to be sufficient to detect a variety of urban trees in different urban greenspace management settings, i.e. they are not limited to tree species and forms

observed in Vancouver and Surrey (i.e. detection of palm trees in Pasadena). The tree detection model is therefore robust to a variety of ecosystems and urban green space design without the need for extensive retraining.

Furthermore, performing inference on Coquitlam (Mapillary) test imagery without retraining results in the highest *mAP* value of the four datasets. Model performance appears therefore robust to the different data source and sensor used for street-level photography, i.e. Mapillary, (s. fig 2.4 and tab 2.3). Precision-recall curves for both testing datasets appear to be very similar. This indicates that the presented model has the ability to generalize well for both a city with a very different ecosystem (Pasadena) and imagery from different data sources or sensors (Mapillary in the case of Coquitlam). The consistency of model performance with an $AP_{50} > 0.6$ regardless of data and sensor source implies that panoramas acquired from both GSV and Mapillary are suitable for use in the urban tree mapping model.

Instance Segmentation performance as a function of mask size

Next, I assessed the influence of tree mask sizes on MASK R-CNN instance segmentation performance. Plotting *AP* metrics values against mask size thresholds shows that larger masks get predicted more accurately (*AP* values over 0.6) (s. fig 2.5). This is expected, since the ratio between the outline of a tree and the tree mass contained by the outline (i.e. outline-mass-ratio) decreases with object size, resulting in a bigger weight of the fuzzy tree outline with decreasing mask size in the calculation of model evaluation metrics [70]. Predicting precise fuzzy tree outlines (tree to sky interface) is often harder than predicting the tree mass. Outlines therefore often differ more from the ground-truth outline than the actual mass of a tree, resulting in declining evaluation metrics numbers with smaller tree size. Notably, the accuracy for large masks are similar for both the Vancouver (GSV) and Coquitlam (Mapillary) datasets, while accuracy for Surrey (GSV) and Pasadena (GSV) datasets are approximately 0.2 lower. This indicates that similarity in tree object structure, to a degree, may have a bigger influence on image segmentation performance than similarity in image quality when instance segmentation is performed on multiple datasets. Values for AP_{50} increase once masks under 3000 pixels (which represent very small or distant trees) are removed. There is a corresponding significant de-

crease in accuracy for masks smaller than 3000 pixels once all masks over 3000 pixels in size are discarded.

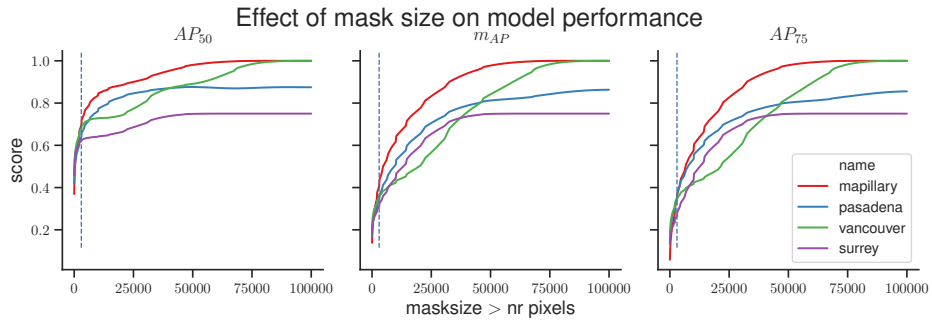


Figure 2.5: Effect of mask size on model performance. I iteratively excluded smaller masks, measured by the number in pixels per mask in calculating AP_{50}, *mAP* and AP_{75}. The dotted vertical line indicates the cut-off size of >3000 pixels for final calculation of the evaluation metrics. The model can predict big tree masks of Coquitlam (Mappilarry) in red and Vancouver in green datasets the most accurate.

Error analysis for instance segmentation

I manually inspected 296 failure cases (including small masks) to identify the most frequent tree detection error (s. fig 2.6 and 2.7). The majority of errors arise from densely planted public and private trees, resulting in two trees being detected as one combined tree, occlusion of trees, or otherwise overlapping trees (s. fig 2.6 (a), (b), (e), (h)). This source of error confirms that distinguishing between visually overlapping amorphous objects is a difficult task [18]. Detecting trees using multiple street-level perspectives potentially offsets this error source, as occluded or overlapping trees can either be seen in the foreground or are otherwise distinguishable from another perspective and image in the full model [72]. Detecting hedges as false positives, small trees, trees in shadows of buildings and trees with leaf-off condition (s. fig 2.7, (c), (d), (f), (g), (i)) is a direct result of having very few training examples of these special cases in the datasets. As expected, most of these errors were detected for small mask sizes (<3000 pixel) (s. fig 2.6).

Trees seen far in the distance were often disregarded in the manual labeling

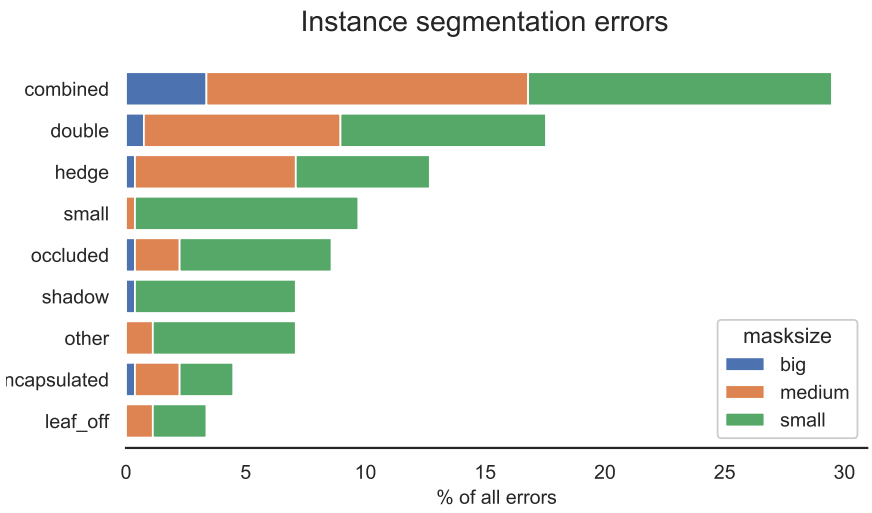


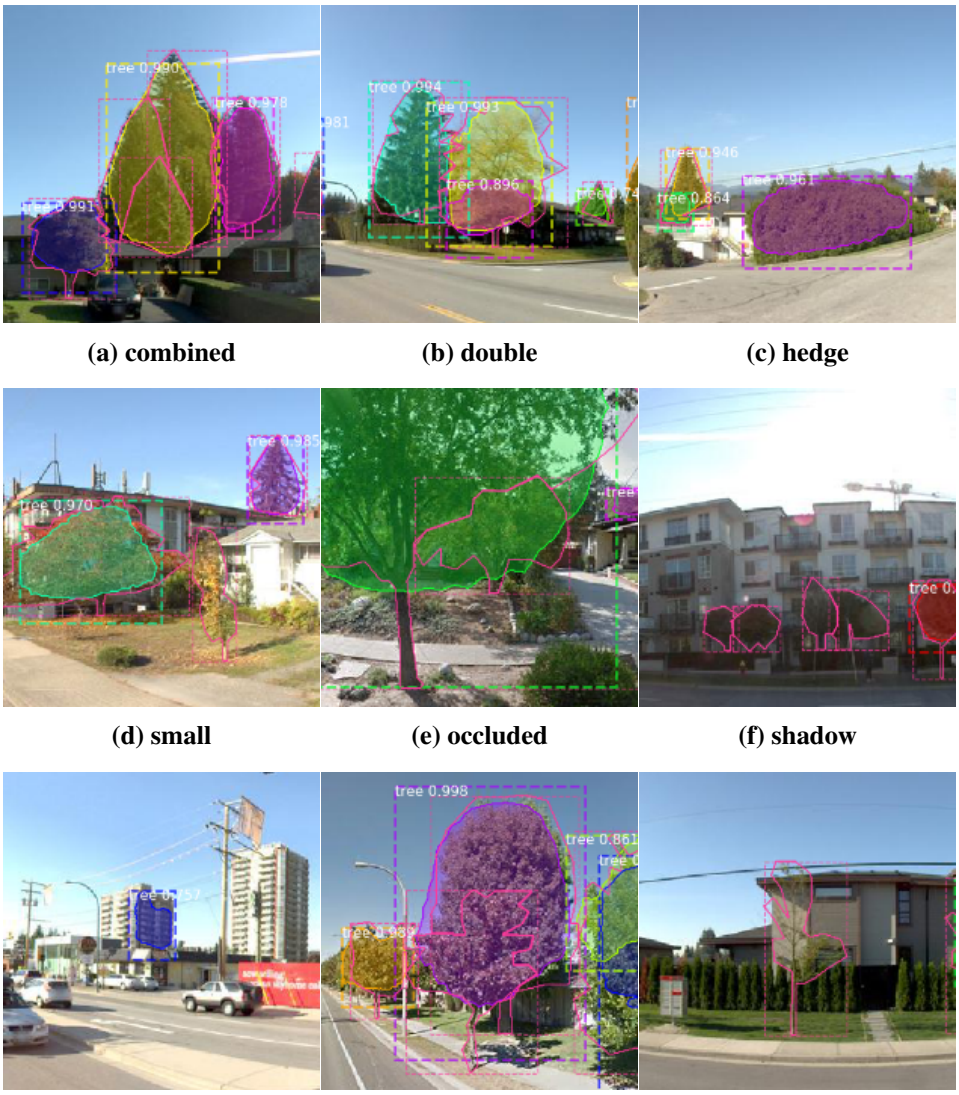
Figure 2.6: The most common inference errors of tree instance segmentation with Mask R-CNN (in percent).

process due to their small mask size and relative distance for the direct camera location at image capture time. I note that 200 of these human labeling errors were recorded, describing instances where the model correctly identified a tree but no corresponding tree label was created. I disregarded human labeling errors for small masks, as masks under a threshold of > 3000 pixels in size were not included in the final evaluation (s. section 2.3.1). These smaller masks, which represent trees found in backyards or distant trees, could be included with help of additional data augmentation methods mentioned in Kisantal et al. [70] in future analyses.

2.3.2 Localization

Comparison to ground-truth

I matched 70% of all ground-truth tree measurements with tree predictions after excluding all matches over 15 meters in distance as false positive matches (s. fig 2.8). Non-matched ground-truth measurements often result from a tree missing in the tree detection process, through either occlusion by larger trees in the front of an image (s. fig 2.8 (a)), or by the absence of a tree in either the ground-truth



(a) combined **(b) double** **(c) hedge**

(d) small **(e) occluded** **(f) shadow**

(g) other **(h) encapsulated** **(i) leaf-off**

Figure 2.7: Examples of masking errors in instance segmentation with Mask R-CNN. Most common errors include: separate trees detected as one (a), one tree split into multiple detections (b), hedges or shrubs detected as trees (c), small trees that were not detected (d), undetected trees behind detected trees (e), undetected trees in shadows of buildings (e), non-tree objects detected as trees and masking errors (g), undetected small trees in front of large trees (h) and undetected trees with leaf-off condition and non-sky background (i)

measurements or the street-level imagery, due to a two or more year time difference in the ground-truth and imagery datasets (s. fig 2.8 (b)). Localizing trees using monocular depth estimation can potentially help to prevent loss of information since every single tree detection can be localized and is not dependent on many photographs from different views [72]. I note that triangulation, in comparison to raw tree location predictions, successfully reduced the mean absolute position error for all areas by approximately 2 meters, from 9 to 7 meters, and the total count of tree predictions by 45-55%.

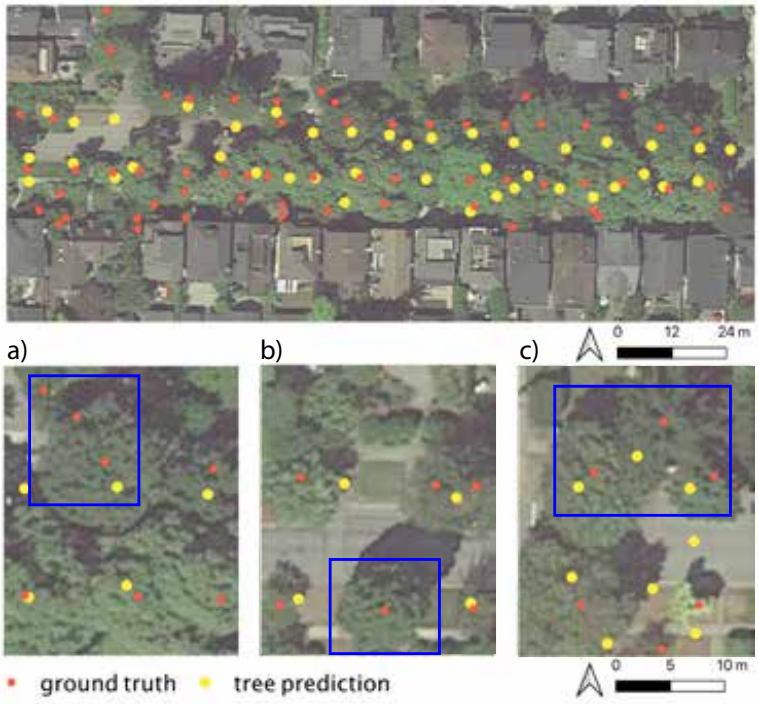


Figure 2.8: Location prediction results of trees. Predicted tree positions (yellow), ground-truth measurements (red) and common detection errors (blue). 70% of all measured trees were detected, 30% are missing through occlusion by large trees (a). The time difference between ground-truth measurements and when a street-level image was taken (>=2 years) results in the absence of either a tree prediction or a ground-truth measurement (b). Geolocation accuracy decreases slightly with increasing distance of trees to the car position of image capture (c).

Minimum distances between tree location prediction and ground-truth measurements for all areas are 0.26 meters or higher (s. tab 2.4). Tree location prediction in Vancouver is, with a mean of 5.28 and a median position accuracy of 4.36 meters approximately 2 meters more accurate than all other areas, followed by Coquitlam's urban and suburban areas with mean and median slightly above 6 meters. With a mean of 7.06 meters and a medium of 6.87 meters, geolocation performance in Surrey is lower than in all other areas. Overall lower position accuracy in the Surrey could be attributed to the area of interest being located on a slope $>15\%$ negatively influencing triangulation, compared to other areas with no or a relatively low slope ($<5\%$). Another source of error may be the overall more spread building and green space structure of the Surrey area. This structural characteristic is leading to trees being located further away from the camera position of image capture with a slight decline in detection and location accuracy with resulting smaller tree masks, discussed in section 2.3.2 (s. fig 2.8 (c)).

Table 2.4: Absolute geolocation accuracy (in meters)

	match	min	mean	median	std
Vancouver	235	0.26	5.28	4.36	3.59
street trees	143	0.26	4.31	3.92	2.76
private trees	97	1.56	8.55	8.38	3.77
Surrey	94	0.42	7.06	6.87	3.36
Coquitlam (urban)	64	0.46	6.58	6.26	3.22
Coquitlam (suburban)	159	0.55	6.83	6.07	3.73

Location accuracy for private and street trees

After triangulation, 143 street trees (93% of the ground-truth measurement) were successfully located in the Vancouver area after triangulation. 11 trees (7%) of 154 street trees remained unmatched (<15 meter ground-truth to prediction). The majority (9 trees) of ground-truth street trees that were not matched were either newly or re-planted small trees in between the date of capture for street-level images and the collection of ground-truth data (i.e. 2 years) (s. fig 2.8, (c)). The two remaining unmatched trees were not detected in the instance segmentation step, due to large vehicles obstructing the trees in respective street-level images. Distances from

ground-truth to tree predictions for street trees range from 0.26 to 13.14 m meters with a mean of 4.31 meters, a median of 3.92 meters and a standard deviation of 2.76 meters (s. tab 2.4). The mean of street trees (red) can be detected almost 1 meter more accurately than the mean of all trees (blue) and 4 meters more accurately than the mean of private trees (green) in the Vancouver area (s. fig 2.9). The overall more accurate predictions in Vancouver are possibly a result of the presence of uniformly and separately planted street trees (s. fig 2.9, red, and fig 2.8). Owing to street trees proximity to the camera position, tree masks are bigger and predicted more accurate which has a potential influence on the location prediction of street trees (s. section 2.3.1).

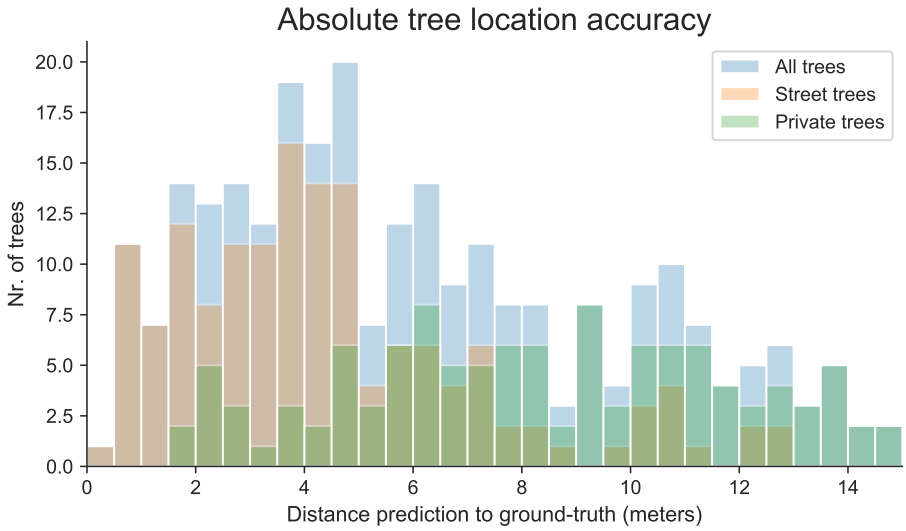


Figure 2.9: Absolute geolocation accuracy for street trees (red), private trees (green), and all trees (blue) in the Vancouver area.

Absolute location accuracy for private trees with the presence of street trees is 3 meters less accurate compared to the previously discussed overall accuracies of all of Vancouver, Surrey and Coquitlam areas (s. tab 2.4). Values for Vancouver's private trees are a minimum of 1.56 meters, a mean of 8.55 meters, a median of 8.38 meters and a standard deviation of 3.77 meters. 70% (97 trees) of all recorded private trees were matched, 30% (43 trees) were not matched. Both, non-matches and

low position accuracy of private trees may be influenced by the more varying spatial pattern of planted private trees. Surrey and Coquitlam areas with similar spatial tree heterogeneity, not influenced by street trees, also recorded approximately 30% of non-matched ground-truth trees. As previously discussed, the combination of two trees detected as a single tree is the most common tree detection error for all mask sizes (s. section 2.3.1). This error is expected to occur more often for densely planted private trees with overlapping canopy, than uniformly planted street trees [144].

It is also possible that the presence of street trees influences the model's location accuracy. Surrey and Coquitlam's (suburban) private trees (no presence of street trees) show lower positional errors than Vancouver's private trees. Street trees often overlap with private trees in street-level photographs due to their proximity to the camera position. Street trees therefore influence both, monocular depth estimation of private trees and the bearing information of the tree detection bounding box from the camera position, as the center of mass shifts towards the larger part of the mask, the street tree. These detection errors negatively influence the localization process and may result in lower positional accuracy for private trees in areas with street tree presence.

Location accuracy with distance of tree from position of image capture

Distances of street tree measurements to the camera positions range between 6-14 meters, distances of private tree measurements to the camera position are typically $>$15 meters away, resulting in a bi-modal distribution of all distances between ground measurements and car positions for all areas (s. fig 2.10). Another reason for more accurate geolocation in Vancouver may be attributed to Vancouver trees being positioned closer to the camera at the time of image capture than in Surrey and Coquitlam image datasets (s. fig 2.10). The range of absolute position error from the tree location prediction to the tree ground-truth measurement is slightly lower for trees closer to the camera position than the error range for trees further away from the camera (s. fig 2.8, (b)), indicated by the shape of the Kernel Density Estimates (KDE) in figure 2.10. However, only a low correlation with R-square of 0.17 can be recorded between the distance of the predicted tree to the ground-truth

measurement and the distance of the ground-truth measurement to the car position. The mean positional error increases with distance to the camera by approximately 0.23 times the distance between ground-truth tree location and camera, indicated by the slope of figure 2.10. This aligns in magnitude with a Root Mean Squared Log Error (RMSLE) of approximately 0.2 reported by Godard et al. [50] for the increase in error for monocular depth estimation with distance from the camera position. Random noise of 6.3 meters is introduced, likely through different street slopes, described tree detection errors and resulting triangulation errors. I also detect a systematic error, an intercept of 2.7 meters with a potential cause through systematic car position GPS inaccuracies in urban landscapes [41]. Another cause for this systematic error could be the initial tree location prediction using the center of mass for each tree crown, retrieving a depth measurement for the outside of the crown diameter instead of the usually measured stem position.

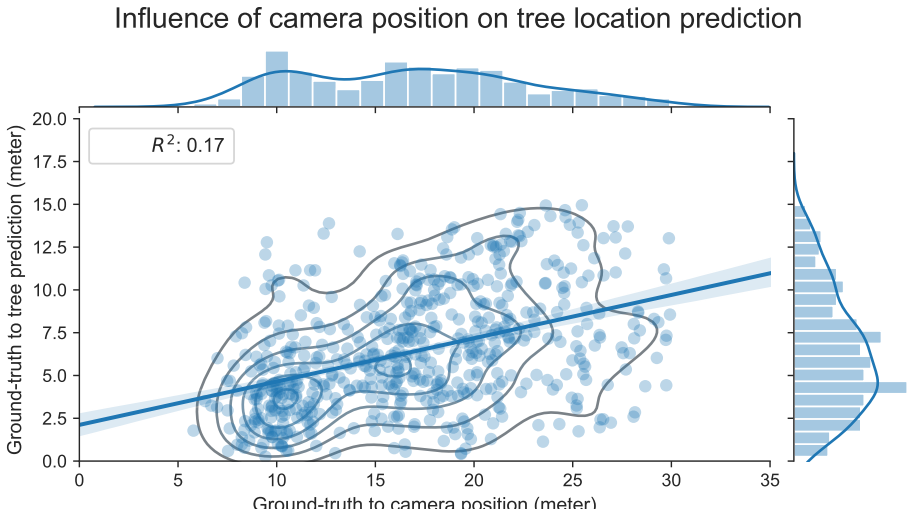


Figure 2.10: Influence of camera position at time of image capture on tree location prediction. Comparison of the distance of ground-truth measurements to the camera location at time of image capture vs. ground-truth measurements to predicted tree locations for all data points (Vancouver, Surrey, Coquitlam))

2.4 Conclusion

To support decision-making and research that can improve the management of urban forests, cities need more cost-efficient and widely applicable tools that can provide high-resolution spatial information on single urban trees for the entire urban and peri-urban landscape (s. section 1.1.3) [69]. I presented a promising low-cost framework for mapping individual urban trees over large areas that shows potential to be adopted in different cities around the world. This novel model relies solely on street-level imagery as a data input and does not require any additional, potentially expensive VHR aerial or satellite imagery for the geolocation of trees. Furthermore, it is developed and tested to be transferable over different image sources and geographical regions as evidenced by the experimental results.

The approach can be applied to a diversity of urban trees and forests, both public and private, and could form the basis for urban assessments that require single tree detection. I found that MASK R-CNN can be successfully trained to identify fuzzy objects like trees to a high precision with a minimal amount of training images (48 images) and a layered training approach integrating open source imagery datasets (COCO Stuff). The experimental results of this study demonstrate that a layered training approach resulted in a 23% higher tree detection rate compared to only using transfer learning. AP_{50} values over 0.62 are consistent with state-of-the-art results in other studies segmenting fuzzy objects [18]. The instance segmentation model, in combination with the layered training approach, has shown potential to learning a broad range of tree shapes, species and sizes without the need for extensive training on particular tree features. For instance, palm trees in the Pasadena test set were detected without palm trees being included in the Vancouver and Surrey training set. The combination of DL and street-level imagery appears promising towards the detection of trees in different urban ecosystems. Further, the model is not limited to the use of the same sensor or dataset. Both Mapillary and GSV panoramas showed suitable for urban tree mapping.

I accurately geolocated trees using one or two street-level images, a monocular depth estimation algorithm, and triangulation that requires no additional or context information. The geolocation of street trees with a mean accuracy of around 4 meters was approximately 2 meters more accurate than the mean accuracy of 6

meters for private trees seen from the street. Most trees clustered at 10 meters distance from the camera position for the tested Google and Mapillary imagery. The proximity of trees resulted in a generally larger tree mask in the tree detection step. Inversely, the further away a tree was away from the camera, the more errors in tree detection occured. This suggests that for future application the distance from the camera position at time of image capture to trees of interest should be a consideration when choosing or generating a dataset for urban tree mapping for future application. Detection errors influenced our tree geolocation module and I recommend collecting images in a range of 7-14 meters away from trees of interest for best positioning results.

Street-level imagery in combination with DL brings a new perspective to assessing urban forests. Accurate masking and geolocation of trees can provide the basis for a variety of quantitative urban forest assessments (s. section 4). For example, this assessment could be used in the future to quantify ecosystem services of urban trees at a large scale using tree structure attributes as proxies to estimate multi-functionality. Future research directions aimed at optimizing street-level imagery capture could include: assessing the influence of spacing between camera positions of image capture on triangulation and positional accuracy of tree predictions; evaluating the influence of slope and vertical terrain variability on geolocation performance; and improving geolocation performance for areas with high terrain variability.

Chapter 3

A machine learning tool for mapping urban tree diversity

3.1 Introduction

3.1.1 The importance of urban tree diversity

A diverse and healthy urban forest enhances the ability of cities to adapt to climate change impacts, such as droughts or floods, improves wildlife habitat, contributes to the protection of native ecosystems and, importantly, increases resilience of cities to pest and disease outbreaks [7, 104]. Mass tree mortality, i.e. of *Fraxinus* or *Ulmus*, due to disease and invasive pests has been known to occur; famous examples in Canada and the United States being DED and EAB [90]. These outbreaks can be hugely detrimental to the health of urban ecosystems, including the health of people living in cities, and can come at a great cost to municipalities (s. section 1.1.2) [135].

Given the many benefits of a diverse urban forest (s. section 1.1), preserving and improving urban forest health and diversity are key goals of many urban forest strategies across North America and Europe [7, 62, 108]. Tree diversity metrics used in the context of urban forest management often relate to richness (the count of different tree genera), evenness (the proportion of a given tree genus with the

total urban forest), composition (the identity of present tree genera), and distribution (the spatial abundance of tree genera) [104]. These metrics are used to assess the state of diversity in urban forests, and they require detailed and accurate urban tree inventories as a baseline. In the face of resource constraints and lack of capacity, municipalities are increasingly looking for new ways to carry out urban forest inventories and continuously assess the state of their urban forest resource (s. section 1.1.4), especially as climate change is predicted to increase urban forest vulnerability to pests and disease (s. section 1.1.3) [96].

3.1.2 Bio-surveillance in the Metro Vancouver region

Urban decision makers in Canada will require detailed urban tree inventory data to predict FIS spread patterns, to minimise impacts on valuable urban forests and to assess the efficacy and efficiency of bio-surveillance programs [43]. Emerging national and international policy statements and strategies will greatly impact the management of urban forests for prevention, detection and rapid response for FIS infestations [39]. Key goals of these policies and strategies are to improve surveillance activities in geographic areas under risk of pest and pathogen introduction, such as residential areas close to ports, tree nurseries, or industrial zones, to evaluate the effectiveness of international policies and pest contamination procedures in regards to introduction prevention.

The implementation of the above strategies will require a sound understanding of baseline conditions such as tree species and genera richness, evenness, composition, and distribution. The availability of cost-efficient, fine-scale urban tree inventory data, therefore, has the potential to direct successful bio-surveillance efforts and identify areas of high economic and invasion risk for many known and unknown FIS [116]. Detailed tree inventory data can, for example, provide valuable information about the location of native and planted host trees susceptible to attack from specific FIS. Knowledge of the spatial distribution of urban trees and their genera composition allows to determine the most effective bio-surveillance activity in varying urban forest landscapes (s. section 1.1.3) [71].

Two common bio-surveillance activities for early detection of FIS in the Metro Vancouver region involve: 1) the distribution of pheromone traps around areas of

high introduction risk, such as harbours or commercial zones where FIS first come in contact with trees, and 2) manual visual inspection of potential host trees present at intersection points of a 1km by 1km triangular grid placed over the Metro Vancouver area (correspondence with Kimoto T., April 2019, CFIA). Both of these activities directly rely on up-to-date, spatially comprehensive data on the distribution and host tree or genera composition of the urban forest. Current tree inventories of the 21 municipalities in the Metro Vancouver region are mostly restricted to public street trees or other trees on public land and exclude large areas of urban forest, especially trees on residential properties. For example, in Vancouver, about 37% of urban forest is located on private land [62]. As outlined in section 1.2 automatic processing of spatially extensive GSV imagery has the potential to provide detailed information on the abundance and distribution of host trees or tree genera in urban areas. With the aim of improving urban bio-surveillance activities in the Metro Vancouver region, I propose a method to automatically classify tree species at the genus level, leveraging GSV imagery and CNNs.

3.1.3 Training data for tree genus classification

The main challenge of using new DL technology in urban and environmental research is the lack of availability of large scale, public training datasets. Most publicly available datasets still focus on classical areas of machine learning like face recognition [112], self-driving cars or medical imagery. There are few very specialized datasets available for tree species or plant recognition [141]. These datasets represent best available imagery or specific use cases and often do not represent an operational dataset needed for learning correct feature representations by the CNN.

3.1.4 Chapter objectives

The study presented in this chapter has two key goals. The first goal is to propose a multi-stage strategy for building large imagery datasets for research in urban areas, with only a limited amount of manual annotation needed. Therefore, I propose a method that integrates readily available geospatial information (i.e environmental in-situ datasets, such as street-tree inventories) and geo-tagged street-view imagery, leveraging the tree detection method presented in chapter 2. The second goal is to

create a DL model to classify urban tree genera from street-level imagery. I explore state-of-the-art procedures in transfer-learning for fine-grained classification problems. Ultimately, I create a new dataset of tree genera hotspots for the Metro Vancouver area in British Columbia (BC), Canada, to inform urban bio-surveillance management and planning.

3.2 Data and methods

3.2.1 Case study site

The Metro Vancouver region spans 2,700 square kilometers, including the three cities (vancouver, Surrey, and Coquitlam) introduced in chapter 2, section 2.2.1. Metro Vancouver is a federation of 21 municipalities, with 26 urban centers ranging in size and character. Given proximity to one of the major trade nodes to Asia, urban forests in the Metro Vancouver region are particularly vulnerable to invasive tree pests and insects, such as ALB, AGM, DED or SOD arriving through international trade [147] (s. section 1.1.2). Detailed maps of tree genera can for example help urban bio-surveillance managers to identify areas of high invasion risk and target areas for visual inspections according to host tree abundance and accumulation (s. section 1.1.3). In this study, I define areas of high invasion risk as areas that show an accumulation of detected trees of the same genus (as investigated by KDE), located in relative proximity to points of entry for pests and pathogens, such as industrial areas or ports, or connected to areas where the assessed tree genus can be observed. *Pseudotsuga*, *Thuja*, *Acer*, *Ulmus*, *Quercus* and *Fraxinus* are native and planted tree genera, that are abundant within the Metro Vancouver area and can act as host to either ALB, AGM, DED or SOD, pests and pathogens that are expected to arrive in the region in future [69, 120, 127, 147].

Metro Vancouver imagery dataset

A GSV imagery dataset for assessing tree species distribution in parts of the Metro Vancouver area (excluding Abbotsford, Mission, Maple Ridge, Langley Township and Pitt Meadows) was acquired from GSV in 2017. The dataset consists of a total of over 2 million images of size 512x512 pixels, predominantly collected from

April to September, 2017. It contains images for over 690,000 car positions spread over Metro Vancouver, with four images per car position representing a Filed of View (FOV) of 0°, 90°, 180° and 270° from true north, respectively.

Tree genus classification training dataset

To build a tree genera classification model to classify the retrieved images of Metro Vancouver, I developed a workflow to build a labeled dataset containing approximately 40,000 curated training images for 120 different tree genera compiled into 41 genera classes and one "other genera" class. For a full list of tree genera classes see appendix D. The method for download and retrieval of imagery for building training, development and testing datasets for tree genus classification is described in section 3.2.4.

3.2.2 Full mapping workflow

Computer vision tasks are frequently approached as end-to-end DL problems. In end-to-end learning, learning is highly automated, meaning that all stages of learning are performed as a holistic learning process (i.e. detection and classification of trees as one model and learning process). The main drawback of end-to-end DL models is that they usually require enormous amounts of data to train (millions of images) [49]. I promoted the strategy of sub-problem solving which requires less training imagery compared to training end-to-end DL models for the task of tree genera classification from street-level imagery [49] (s. fig 3.1). In an automated manner, I first detected trees in street scenes, as described in chapter 2, and then classified cropped images displaying the tree of interest in the center of the imagery. Adopting this strategy I received a tree count per street-level car location as well as a tree genera label for each detected tree. Lastly, I created maps of kernel-density estimates, using the open source package "seaborne" to visualize hot spots of tree genera in the Metro Vancouver area [143].

3.2.3 Tree detection

In this study, MASK R-CNN was used to detect and outline trees in imagery by generating bounding boxes and binary segmentation masks for each tree instance (s.

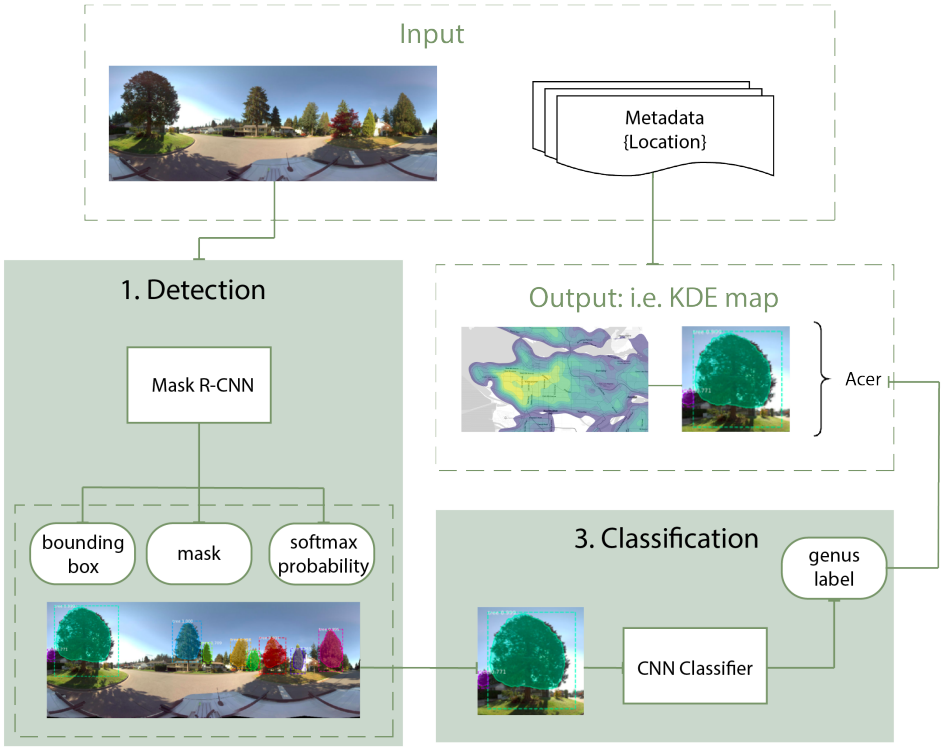


Figure 3.1: Tree genus classification workflow. GSV images are cropped to display single trees and further augmented to build training, development and testing datasets.

section 2.2.4). These tree detections were then used for both tree location, as presented in chapter 2, and the following genus classification model. For the tree genus classification workflow, MASK R-CNN's binary segmentation masks were used for two purposes. First, the tree detection module, developed in chapter 2, was used to build a tree genera dataset outlined in section 3.2.4. Second, MASK R-CNN was applied to the full Metro Vancouver dataset and generated bounding boxes were used to extract images of single trees used for single-label tree genus classification. The MASK R-CNN architecture, its training and evaluation process, as well as the chosen training data, are described in detail in appendix B.

3.2.4 Multi-stage strategy for building tree genera dataset

I proposed a multi-stage strategy to rapidly collect and sample tree genera images from street-level imagery providers for training a tree genus classifier. The challenge of this task was to match known occurrences of tree genera recorded in Vancouver's official street-tree inventory to pictures of the recorded trees retrieved from GSV, to create a labeled dataset for training. The main stages of the proposed strategy to create this labeled street-level imagery dataset involved: 1) image acquisition, 2) cropping images to the tree of interest, and 3) manual removal of erroneously labeled images.

Step 1: Imagery acquisition

I leveraged location and genus information of existing street tree inventory data for the city of Vancouver to semi-automatically create a training and testing tree genera imagery dataset. Vancouver street tree inventory data contains geographic coordinates for manually recorded individual trees, connected to a single species and genus label. Metadata of street-level imagery from different providers (i.e. crowd-sourced Mapillary data or proprietary GSV data) generally contains the camera position at the time of image capture in geographical coordinates and bearing of the image center in relation to magnetic north in degrees. Given the spatial relation between the tree location and species name recorded in the street tree inventory and the camera position at the time of image capture I calculated the necessary imagery bearing to display the selected, manually recorded tree location in the middle of the image (s. fig 3.2 (a)). I then downloaded all available imagery with the calculated bearing information as input for the image center, a FOV of 90° and a pixel resolution of 512x512 pixels from the GSV platform in June 2018 [144] (s. fig 3.2 (b)). The choice of a relatively wide FOV of 90° accounted for known errors of GPS accuracy of 2-12 meters in urban environments for both the car position at time of image capture and the tree location recorded in the urban tree inventory [149]. Owing to inaccuracies in GPS measurements, and associated error in calculated bearings, a wide FOV ensured that each downloaded image displayed the tree of interest recorded in the tree inventory, even if the tree was off-set from the center of the image.

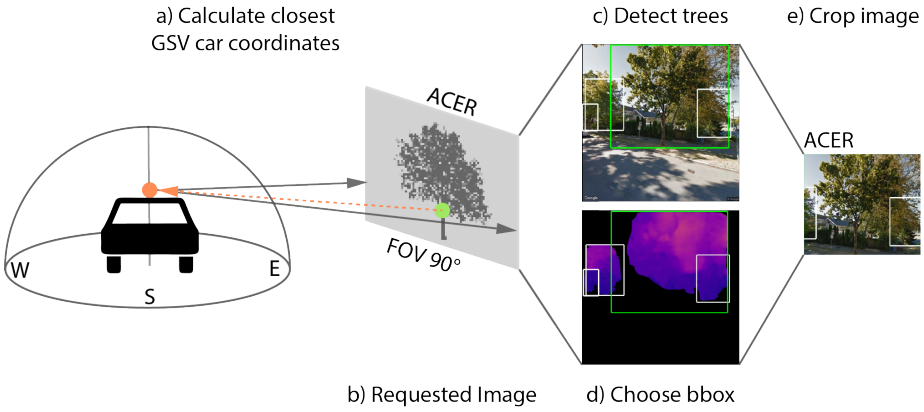


Figure 3.2: Training data generation for tree genus classification. Using existing street tree inventory, the closest GSV car coordinates to each recorded tree are calculated (a) and a corresponding image is requested (b). All trees in the requested image are detected with a trained MASK R-CNN model (c) and the closest and largest bounding boxes or tree detections are chosen to represent the street tree (d). Images are cropped to display the selected single tree in the center for training the tree genus classification model (e).

Step 2: Cropping images to tree of interest

I post-processed images in order to create an image displaying a single tree, connected to the correct label from Vancouver's street tree inventory. I first applied the trained MASK R-CNN model to detect all trees present in the image as described in chapter 2 and appendix B (s. fig 3.2 (c)). I then ran a monocular depth estimation model, to create a dense depth layer for each image as described in section 2.2.5. I computed a measure of distance between each tree detected in the image and the camera position at the time of image capture in meters by extracting the depth value of the pixel located at the center of mass for each calculated tree mask. I then selected one tree per downloaded GSV image as the labeled street-tree under the assumption that the particular street-tree must be the tree with the smallest depth value, or the closest tree to the camera position at the time of image capture (s. fig 3.2 (d)). Each GSV image was cropped according to the bounding box of teh selected tree, previously computed by MASK R-CNN (s. fig 3.2 (e)).

Step 3: Manual removal of erroneous images

Lastly, I manually inspected the created training dataset of cropped images, one genus class at a time, to identify all cropped images of trees with an incorrect label. Images that displayed a tree genus that did not correspond to the matched label were discarded. Furthermore, I discarded all images of size 50 KB or smaller as visual analysis revealed that their resolution was not fit for training the classifier. Following the recommendations of [15] for tree species classification, all genera with an image sample set over 125 images per class were used as separate classes training the genus classifier. All genera with an image sample set under 125 images per class were combined under the class label "other". Resulting in a tree genus dataset of 41 tree genera and one "other" class (s. fig 3.3).

3.2.5 Tree genus classification

To classify the detected trees into one of the 42 fine-grained genera classes, I trained an image classification model using the novel fast.ai DL framework built on PyTorch. fast.ai is an open source software package, designed for researchers and DL practitioners to quickly build and iteratively train DL models with state-of-the-art guidance on best practices for training. I used a transfer-learning approach with a modified RESNET50 architecture and a softmax classifier [58].

Balancing the training dataset

First, I split the genera training dataset into training, development and test datasets with a 80:10:10 ratio. I sought to prevent the classifier from over-fitting on tree genera which dominate the genera dataset, due to their abundance in the Vancouver, i.e. *Acer* or *Prunus* with over 5,000 images per class. In order to balance classes in the training dataset, I under-sampled respective classes, meaning, I selected a maximum of 4,000 images per class and removed the rest from training. I then over-sampled all other classes [17]. First, I added all downloaded, uncropped images of the respective tree, and second, multiplied images using a NumPy random number generator (assuming a univariate Gaussian distribution) until all classes were equalized to a count of 4,000 images per class.

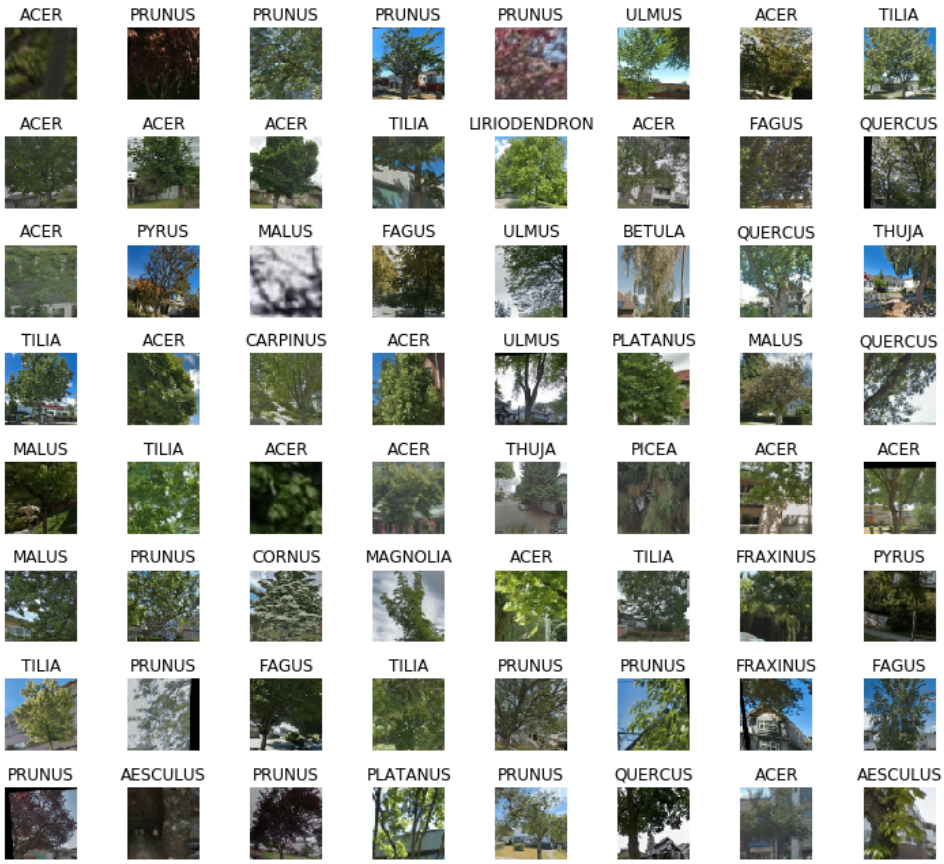


Figure 3.3: Examples of tree genera dataset and data augmentation. GSV images are cropped to display single trees and further augmented to build training, development and testing datasets. The final training size of each image is 256*x*256 pixels.

Mixup and data augmentation

Mixup is a novel data augmentation technique known to improve generalization error of models and avoid the memorization of corrupted labels [150]. As the name suggests, mixup constructs a training image through mixing two random examples from the training set and their labels through linear interpolation (60% image one, 40% image two) [150]. In order to avoid over-fitting on the oversampled dataset, I implemented mixup. Additionally, several data augmentation techniques were

applied to imagery as mixed training images were fed into the model. Data augmentation was applied randomly and included a horizontal flip with a probability of 50%, a rotation of up to 15°, a zoom up to 150%, lighting and contrast change of magnitude up to 0.4 and a symmetric warp of magnitude up to 0.2 (s. fig 3.3).

Mixed precision training and progressive resizing

Mixed precision training performs operations within the model using smaller sized data types when possible – so called half-precision or 16-bit Floating Points (FP16) and Single-precision Floating Points (FP1) – which improves training time and decreases the use of memory [93, 107]. I implemented mixed precision training to speed up the computational training process [63]. Furthermore, I used progressive resizing of images, starting from 64x64, over 128x128, to 256x256 pixels for training respective models [112]. As training with smaller sized images was less memory intensive the training process was accelerated through learning to distinguish tree genera on coarse resolution images first, in comparison to training on larger resolution images from the beginning. I used trained weights from each model with a smaller image size to initiate the training process of the model with the next bigger image size. I trained the RESNET50 implementation on an NVIDIA GPU (Tesla P100-PCIE-12GB) with 32 CPU cores and 32 GB of memory.

Evaluation

I used *mAP* as an evaluation metric, where precision for each genus class (p_{class}) was defined as the number of correctly classified trees (true positives, TP) divided by the number of correctly classified trees (true positives, TP), plus the number of incorrectly classified trees (false positives, FP) falling into the specific genus class. *mAP* was then calculated as the weighted mean of all class precisions $(p_1, p_2, ..., p_n)$, with the corresponding number of training images per class as weights $(w_1, w_2, ..., w_n)$::

$$p_{class} = \frac{TP_{class}}{TP_{class} + FP_{class}} \tag{3.1}$$

$$mAP = \frac{\sum_{i=1}^{n} w_i p_i}{\sum_{i=1}^{n} w_i} \tag{3.2}$$

In addition to *mAP*, I used top-3 accuracy to assess model performance. Top-3 accuracy was defined as the percentage of images of one class whose ground truth genus label is within the three highest ranked predicted labels. Last, I compared classification of separate genera classes in a detailed confusion matrix.

3.3 Experiments

3.3.1 Classification performance

The model achieved an overall classification accuracy of *mAP* of $> 82\%$ and top-3 accuracy of $> 95\%$ for 41 different tree genera and one "other" class in both training and development datasets (s. tab 3.1). I closely examined classification *mAP* for different genera classes (s. fig 3.4): the model classified 6 genera over 90%, 13 genera with over 80% and 27 genera over 70% precision; 12 genera were classified with a precision under 70%. *Laburnum*, *Abies*, *Ilex*, *Prunus* and *Betula* genera display the highest classification precision. *Amelanchier*, *Ginko*, *Cercis*, *Juglans* and *Tsuga* could not be classified successfully.

Table 3.1: Classification accuracy for development and test sets

	top-3 accuracy	*mAP*
development set	95.0%	82.4%
test set	95.5%	82.9%

The confusion matrix for the test dataset revealed that most tree genera were successfully distinguished from one another (s. fig 3.5). Few high values diverging from the center line were observed, indicating overall high prediction accuracy, precision and recall for all genera classes. *Cercis*, *Amelancier*, *Tsuga* and *Ginkgo* were the most confused genera classes (s. fig 3.5).

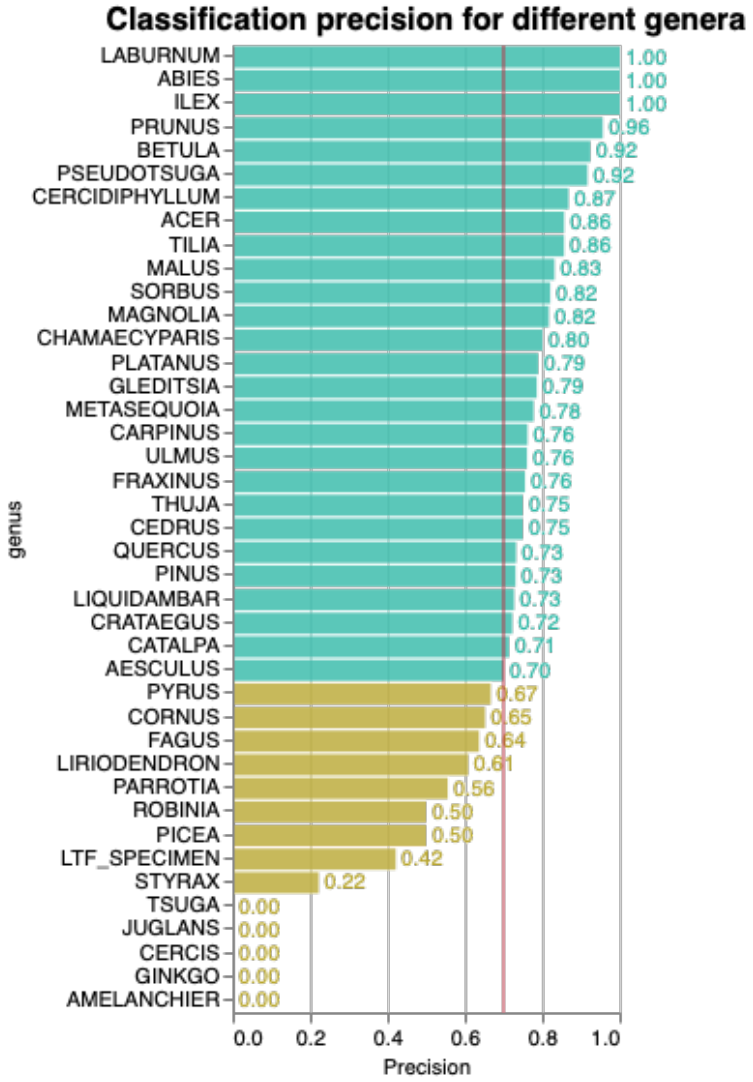


Figure 3.4: Precision for different genus classes in the test dataset. Up to 27 genera can be classified with a precision over 0.7 (Green).

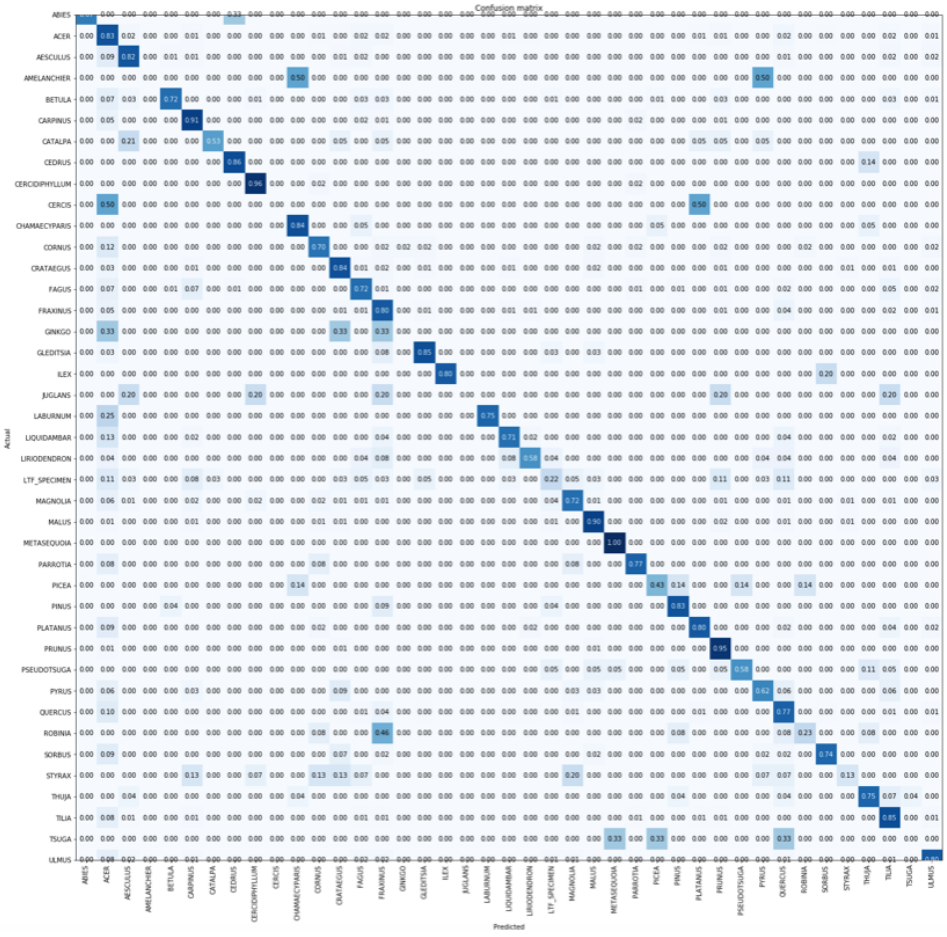


Figure 3.5: Confusion matrix for tree genera classification. Darker shades indicate higher classification accuracy, precision and recall. The darker blue the colour of the confusion matrix, the higher the displayed percentage value. The dark blue diagonal center line indicates that in most cases, the model generated a matching prediction to the actual tree genera displayed on the analyzed image.

3.3.2 Hotspot maps of Metro Vancouver

I detected a total of $>$ 4 million trees in street-level imagery dataset of the Metro Vancouver area. Generated image sizes varied from a minimum of 2 KB to a maximum of 450 KB with a mean of 32 KB (s. fig 3.6). An image of pixel size 256*x*256 used for the last iteration of training the genus model, corresponded approximately to 65,000 pixels or 64 KB in size. Visual analysis of generated images revealed that images under 20 KB or 30,000 pixels in size typically represented trees far away from the camera position of image capture (s. fig 3.6). Images over a threshold of $>$ 20 KB were used to generate the following hotspot maps (s. fig 3.7).

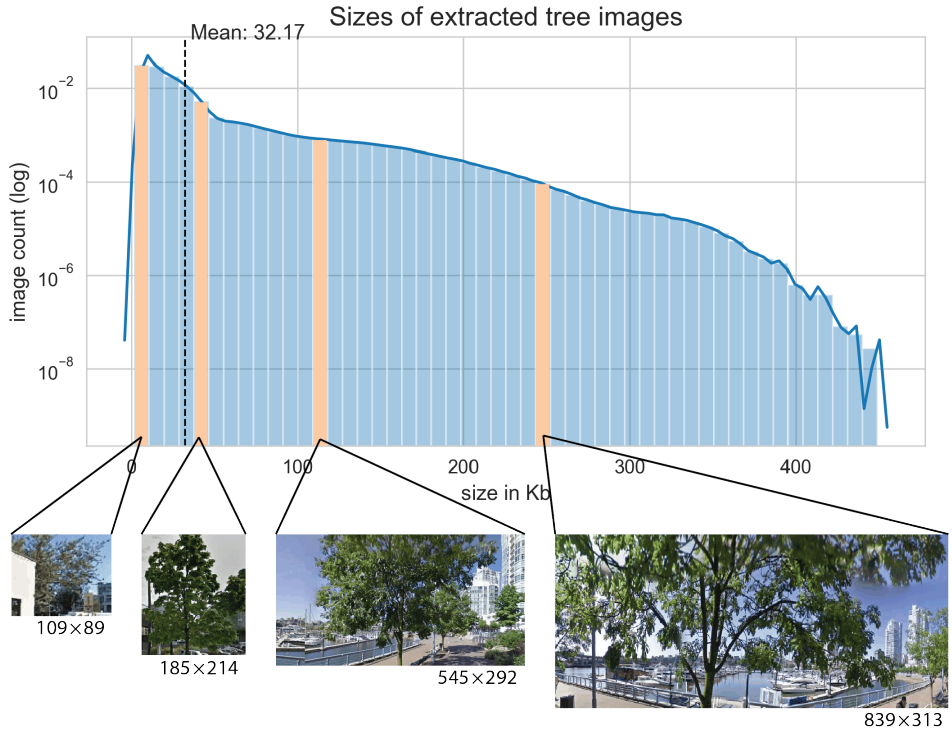


Figure 3.6: Distribution of sizes of generated tree cutouts with examples.

Applying the tree genera classifier to cropped images retrieved through tree detection with MASK R-CNN, I built maps of specific tree genera hotspots through KDE for Metro Vancouver, to aid bio-surveillance planning and management. Displayed in figure 3.7 are two coniferous (*Pseudotsuga*, *Thuja*), and four deciduous

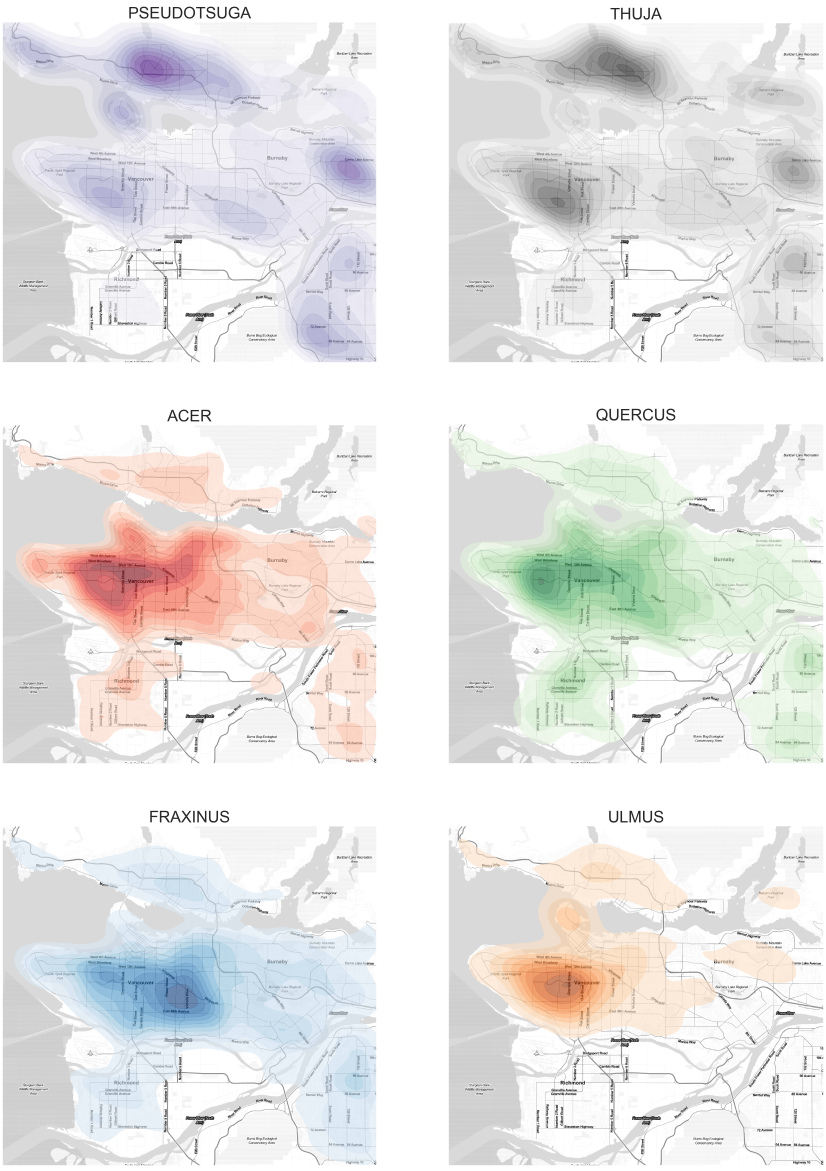


Figure 3.7: Tree genera distributions in Metro Vancouver.
Kernel density estimates are shown for two coniferous, *Pseudotsuga* and *Thuja*, and four deciduous, *Acer*, *Quercus*, *Fraxinus* and *Ulmus*, tree genera.

(*Acer*, *Ulmus*, *Quercus*, *Fraxinus*) tree genera. All six genera are currently under threat by invasive pests and pathogens and of high interest in bio-surveillance campaigns (s. tab 3.2). Appendix C provides a visual example, comparing the generated, underlying dataset used for KDE maps to existing street tree inventory data.

Table 3.2: Selected occurrences of tree genera in Metro Vancouver. Count of generated detections and example threats of two coniferous, *Pseudotsuga*, *Thuja*, and four deciduous, *Acer*, *Ulmus*, *Quercus*, *Fraxinus* tree genera

genus	Count	native	Threatened by
ACER	493,000	x	Asian Long-horned Beetle (ALB)
THUJA	175,000	x	Asian Gypsy Moth (AGM)
FRAXINUS	110,000		Emerald Ash Borer (EAB)
QUERCUS	95,000	x	Sudden Oak Death (SOD)
PSEUDOTSUGA	90,000	x	Sudden Oak Death (SOD)
ULMUS	47,000		Dutch Elm Disease (DED)

Generated data and KDE maps helped to answer diverse questions about the genera composition of Metro Vancouver's urban forest ranging from where most trees were detected (Vancouver West), to highest percentage of assessed coniferous trees (North Vancouver), or highest percentage of *Fraxinus* (new settlements in East Vancouver). KDE maps for both coniferous genera, which are native to Vancouver, show that *Pseudotsuga* and *Thuja* were found throughout Metro Vancouver, but were especially abundant in less densely populated areas, close to provincial parks or nature reserves, i.e. Stanley Park, North Vancouver or Coquitlam. *Ulmus* was mainly detected in the city of Vancouver, West Vancouver and East Vancouver, but was not observed in Surrey or Richmond. In comparison, *Acer* hotspots were very interconnected and spread wide over the region. This suggests that in the event of an infestation of *Acer*, negative impacts of pests and pathogens could be far reaching for Metro Vancouver, as pests and pathogens spread more easily and quickly when host trees are interconnected.

The highest count of tree observations on imagery among the displayed tree genera of interest were *Acer*, followed by *Thuja*. Trees of the genus *Ulmus* were observed the least in this example with a total of approximately 50,000 occurrences

in street-level images (s. tab 3.2). A full list of genera and corresponding occurrences in street-level imagery can be found in appendix D.

3.4 Discussion

3.4.1 Classification model performance

The model presented in this chapter is the first method currently available for tree genera classification from street-level imagery that tests applicability to a larger area like Metro Vancouver. *Cercis*, *Amelancier*, *Tsuga* and *Ginkgo* were the most confused and lowest performing genera classes. These four classes were strongly underrepresented in the training, developing and testing datasets with $< 1\%$ of the total dataset for each class in the testing dataset (s. fig 3.8). In contrast *Laburnum*, *Abies* and *Ilex* that are amongst the classes with the best model performance are also among the image class datasets with $< 1\%$ of the total dataset (s. fig. 3.8). Differences in model performance for these 6 classes could result from: 1) tree genera structures that are generally very difficult to classify through either high heterogeneity within the genus class or similarities to other genera, 2) the low amount of remaining test images after selecting 100 training images is not sufficient to accurately represent the real world distribution of potential GSV images of the respective class. It appears that 100 training examples per class are not enough to train a model to successfully classify *Cercis*, *Amelancier*, *Tsuga* and *Ginkgo*. Both of the above named reasons could be solved in future through increasing training and testing data for the respective classes.

On the other hand, the class with the highest number of training examples (*Acer*) was not the class with the highest classification precision (s. fig 3.4 and fig 3.8). This suggests that certain genera could be harder to classify either owing to similarities to different genera classes or to high heterogeneity within the respective genera class [123]. In the latter case it might be beneficial to separate these genera out to classify different tree species found within the genus. Assessing which of the above factors (lack of training imagery, insufficient testing data, heterogeneity within the genus class, similarity in between genera) is the ultimate cause of lower model performance for certain classes is currently still a challeng-

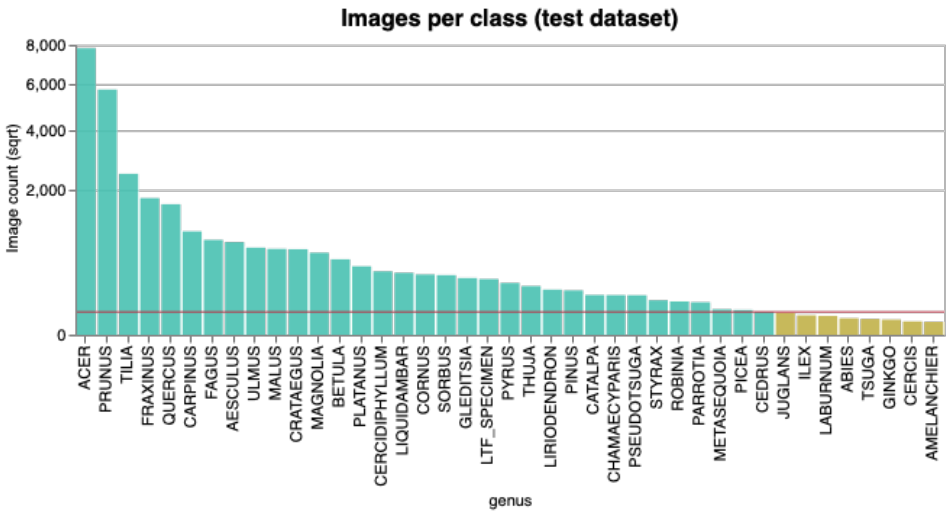


Figure 3.8: Images per class in the tree genus classification test dataset. All classes have a minimum of 125 images in the training set, all remaining images are in the development and test datasets. Up to 8 genera classes (yellow) have under 50 images (red) in the test dataset.

ing and time consuming task through the lack of tools to interpret DNNs [22]. This gap in available tools and standardized workflows opens up possibilities for future research in assessing the influence of tree genera class structure and training data availability on classification performance.

3.4.2 Transferability to other areas

Retraining the presented model with labelled imagery from other cities and including a greater number of tree genera would make it possible to use the model to assess urban forest diversity more extensively. However, as beneficial climatic conditions have made Vancouver's urban forest one of Canada's most diverse forests, the model was trained using a large dataset containing $>$ 1000 images per genus class that characterise a wide range of trees found within a genera, located in differing urban environments. The model could be used to analyse the distribution of genera with a high number of training images ($>$ 50 images in the test dataset) and good model performance ($>$ 70% accuracy) in other urban environments with sufficient street-level imagery taken in summer, including but not restricted to: *Acer*, *Prunus*,

Tilia, *Fraxinus*, *Carpinus*, *Ulmus*, *Betula*, *Magnolia*, *Platanus*, *Thuja*, *Pinus*, *Pseudotsuga*, *Sorbus*, *Cercidiphyllum*, *Cedrus*, *Metasequoia*, *Malus* and *Quercus*.

GSV imagery, processed for training the model, is currently predominantly available for the months of April to October, when most tree genera display leaves. An expansion of the tree genera training data to different seasons would increase the generality of the model as it could be used to classify tree images taken at any time. Even though bio-surveillance management focuses on campaign planning for leaf-on tree conditions, when harmful pest and pathogens are the most active and the most likely to be detected, other smart urban forest management activities, i.e. the health assessment of allergy potential from urban trees, might benefit from information about tree diversity before spring starts (s. section 4.2).

In addition to the current imagery being constrained to certain times of the year, the process of generating training data was restricted to Vancouver's public street and boulevard trees, which limits the ability to train the model for rare tree genera found only on private property. For an exhaustive urban tree diversity assessment, future work should focus on the development of tools for or the collection of imagery for genera that cannot be assessed through the presented training data generation workflow, including tree genera on private property and parks (s. section 4.4).

3.5 Conclusion

I successfully analysed tree genera distribution across the Metro Vancouver area, using DL for the classification of over 2 million street-level images. To facilitate CNN training, I presented a method to rapidly and semi-automatically collect a large training dataset. I trained a fine-grained tree genera classification model with a mean average precision of 83% for 41 different tree genera and one "other" class including a total of 125 genera in the analysis. Integrated into smart urban forest management, the presented workflow and model for analysing tree genera distributions in urban environments presents the opportunity to aid bio-surveillance campaign planning to detect invasive pests and insects early. The approach, coupled with publicly available street-level imagery, could enhance urban forest diversity assessments through more detailed information on trees located on private prop-

erty and has the potential to generate information on tree genera more rapidly and over large areas than has been possible to date through to manual data collection to update tree inventories. Depending on the genera of interest, the workflow can be reproduced to retrain the model on new genera classes or the model can directly be transferred to other urban environments.

Chapter 4

Conclusions

4.1 Key findings

This thesis presented exploratory research in developing a method for automatically detecting, locating and classifying urban trees. In the context of smart urban forest management, a combination of novel DL architectures and cost-efficient street-level imagery was used to generate urban tree inventory data over a large urban spatial extent. The developed method relied solely on street-level imagery as a data input instead of more costly or less detailed aerial or satellite imagery that many other models require (s. chapter 1 and section 1.2.1). The novelty of this method was enhanced in that monocular depth estimation and triangulation were used to predict tree locations without a dependence on complementary information or aerial imagery (s. chapter 2). Finally, a reproducible and fast approach to generate a tree genera classification dataset was presented and maps of urban tree genera distributions for the Metro Vancouver area were created (s. chapter 3).

4.1.1 Tree detection with Mask R-CNN

Trees were detected through training and using the MASK R-CNN architecture for instance segmentation. Experimental results for performance and transferability of tree instance segmentation were demonstrated for four cities (Vancouver, Surrey, Coquitlam and Pasadena) and two data sources (GSV and Mapillary). MASK R-CNN was successfully trained with a minimal amount of training images (48 images) and

a layered training approach integrating open source imagery datasets (i.e. COCO Stuff) to identify fuzzy objects like trees to a high precision. The experimental results of this study demonstrated that a layered training approach allowed for more accurate instance segmentation of trees, compared to using only transfer learning. Tree instance segmentation results (0.6-0.7 AP_{50}) were consistent with current tree or plant semantic segmentation performances found in other studies, with the added value that this work also distinguished between different tree objects [18]. The combination of DL and street-level imagery showed promising results for the detection of different tree shapes and sizes in various urban ecosystems and urban management regimes and was not limited to the use of the same sensor or dataset, without the need for extensive retraining.

4.1.2 Tree geolocation with monocular depth estimation

Trees were located using one or multiple street-level photographs, combining monocular depth estimations generated with the monoDepth model, with tree detection masks and location and bearing information of each photograph. Initial tree location predictions were enhanced using triangulation that required no additional or contextual information. Tree detection with MASK R-CNN in combination with monocular depth estimation was able to provide a basis for street tree location prediction that is comparable to manually conducted ground truth measurements with hand held GPS devices in urban environments. Over 70% of trees, measured on the ground, were successfully located for four different plots (Surrey, Vancouver, Coquitlam urban center and residential area). The geolocation of street trees with a mean of around 4 meters, mainly found in the Vancouver area, was approximately 2 meters more accurate than for private trees (6 meter), predominantly recorded in Surrey and Coquitlam, seen from the street. The presented method allowed for the assessment of trees on private property, a part of the urban forest for that cities are still lacking information.

4.1.3 Tree genus classification

Fine-grained classification across different tree genera from imagery is a challenging task even for humans [13]. To facilitate tree genera classification in urban

environments, a method for rapid sampling of tree genera training images from GSV was presented. A tree genera dataset of 40,000 images compiled for 41 fine-grained tree genera and one "other genera" class, including a total of 80 different tree genera, was created. This dataset was used to train a CNN for tree genera classification with 83% mean average precision. The model was applied to generate tree genera distribution maps over the Metro Vancouver area and could be used in the future for other urban areas provided that sufficient street-level imagery can be acquired to use as training sample.

4.2 Implications

4.2.1 Deep learning for bio-surveillance planning

The goal of this research was to create and assess a methodology that has the potential to improve the consistency and availability of urban tree inventory data across different regional authorities and scales. New data can help inform decision makers for bio-surveillance efforts and urban forest management. For example, an open and reproducible DL approach resulting in more accurate and detailed tree inventories could add significant value to identifying and targeting areas of high infestation risk in existing bio-surveillance investigations, particularly in cases where infestation risk and impact is predicated on species composition and forest structure [40]. Improving urban forest inventories and subsequently identifying trees with high infestation risk using DL techniques can allow decision makers to proactively prevent, monitor and manage forest invasive alien species outbreaks in higher temporal resolution than currently possible [88].

4.2.2 A new baseline for risk assessment

Diversity in structure and function is crucial to urban forest resilience, as exemplified by the outbreaks that have devastated monocultures of elms (*Ulmus*) and ash (*Fraxinus*) trees across cities in Canada and the United States. Urban tree biodiversity and the connectivity of tree canopy supports wildlife habitat, contributes to the protection of native ecosystems, and enhances the ability for urban ecosystems and people to adapt to climate change. A detailed urban tree inventory can be

used to quantify the monetary value of and manage the suite of ecosystem services provided by biodiverse urban forests, including ecological, health, recreational and aesthetic benefits [126]. Tree clusters or groups of trees, for example, will generate more services (such as cooling) compared to a single tree. Large trees will generate greater ecosystem services value than smaller trees.

Weighting the cost of managing and protecting urban trees against the benefits or services they provide is often used as a baseline for risk assessment and decision making. In combination with other spatial information derived from LIDAR or high resolution satellite RS data (e.g. tree health, tree structure), the trained NN can, for example, improve bio-surveillance efforts through implementation into a decision support pipeline for FIS risk analysis. The presented tool could also help map tree genera that are more drought-tolerant contributing to climate adaptation strategies of cities that are expected to be affected by more frequent heat waves.

4.2.3 Smart urban forest management

The research goals identified for this thesis (s. section 1.3) are of interest to many industry and government participants. According to Östberg et al. [155], tree species or genus and location of urban trees have been named some of the most needed urban tree inventory parameters by various city officials and researchers. Municipalities often do not have the resources nor capacity to carry out complete inventories of their urban forest resource, not to mention consistent updates once a baseline inventory has been completed. Efficient, cost-effective, and reliable urban tree inventory techniques are sorely needed to provide cities with the tools for strategic urban forest planning and management. This research also highlights a novel way in which technology can be used to monitor urban forests and enable more proactive decision making about urban biodiversity, which could be considered a contribution to smart urban forestry.

4.2.4 A novel method for environmental research

Lastly, this study represents a project at the forefront of introducing state-of-the-art DL frameworks to environmental management and decision making. It is expected to not only produce a cost-efficient and openly available tree inventory generation

framework, but also to inform research needs for other fields of study. By introducing and showcasing how new AI concepts can be leveraged for environmental RS and object detection, I intend to inspire their application to generate new solutions and expect far-reaching future implications for the fields of environmental management and global change studies.

4.3 Limitations

Detecting, mapping and classifying urban trees from street-level imagery is a complex and challenging task. As a novel approach to generate tree inventory data, this project encountered critical limitations that require further thought for application and research in future (s. section 4.4).

4.3.1 Tree visibility on street-level imagery

The presented methodological workflow is limited to assessing trees that can be seen on street-level imagery. This often excludes parts of the urban forest that are not visible from the street, for example trees found in backyards and trees in parks. Furthermore, as outlined in chapter 2, 70% of all ground-truth measurements were matched in the analysis. Thus, 30% of front yard and street trees were not recorded in our detection and location predictions, either due to erroneous localization or due to detection errors, i.e. through occlusion from other trees (s. chapter 2). Additionally, the performance to detect and locate trees in other parts of the urban forest (e.g., urban woodlands and parks) was not assessed. Even though the developed tool helps to gain insights about the urban forest going beyond street-trees, it is still unclear what proportion of the urban forest can be recorded.

4.3.2 Availability of street-level imagery

Another key limitation for the future application of the developed software and workflow is the availability of spatially coherent street-level imagery. Increasingly, street-level imagery providers update their terms of use to prohibit the large-scale processing and extraction of information from the provided data source (this is in particular the case for GSV, which updated terms of use in September 2018). The purpose of closing many geospatial data sources is predominantly to avoid costly

law suits in case street-level imagery were used to collect sensitive and private information (correspondence I. Seiferling, December 2018, MIT). Other providers still cannot generate large spatial coverage within cities and over different countries and regions (Mapillary).

Lastly, standards among service providers differ. As a result, different services provide street-level imagery of varying quantity and quality. GSV spaces their camera positions of image capture roughly every 15 meters, whereas Mapillary provides imagery spaced one meter or lower apart. While GSV makes it a requirement to collect data with high resolution panoramic cameras only, Mapillary imagery ranges from low resolution smart phone cameras to also very high resolution panoramic cameras and lenses. Similarly to GSV, Bing StreetSide or OpenStreetCam have certain standards in place before an image is made publicly available on their platform, however, image resolution is, to date, still lower than GSV. The presented methodology requires processing of panoramic or high resolution street-level imagery. It has not been tested for compatibility and performance with lower-resolution imagery from providers other than GSV and Mapillary. Lastly, the methodology is restricted to assessing areas with sufficient, high-resolution street-level imagery coverage only.

4.3.3 Limited tree genera training data

Related to the above, the generated tree genera dataset and tree genera classification model are mainly targeted to identify planted trees and trees on boulevards rather than trees found in local parks, greenbelts, backyards or other local natural forests. Planted trees on developed sites that can be identified well and are abundant in Metro Vancouver include: *Acer*, *Prunus*, *Quercus*, *Tilia*, *Platanus*, *Fagus*, *Thuja*, *Malus*, *Carpinus* and *Magnolia*. Due to the genera dataset being developed mainly on the basis of Vancouver's existing street tree inventory, classification accuracy for native trees is generally lower than for planted trees. This raises the question of how far the developed tree genera classification model is applicable to native trees in urban woodlands. Hence, the chosen training and evaluation methodology is limited in that it does not assess urban tree classification accuracy for the largest number of trees, namely native trees, because the number of forest trees are vastly

more than planted trees on streets, in yards and in cultivated park areas: *Alnus* and *Populus* dominate the deciduous list in abundance and size; *Thuja*, *Tsuga* and *Pseudotsuga* dominate the coniferous list in abundance and size. Improving training data for native trees is particularly important in suburbs outside of the City of Vancouver (Langley, Maple Ridge, Surrey etc.), as many of the roadside trees are native trees growing on road allowance or near the road on private property.

4.4 Future research directions

This research has demonstrated the value of using street-level imagery and DL architectures for smart urban forestry management. Current limitations open up a range of avenues for future methodological development, testing applications and research.

4.4.1 Assessing different data sources

A promising research avenue to assess the urban forest as a whole is the combination of data sources from different perspectives, such as the side view from street-level imagery and an aerial perspective from aerial imagery or LIDAR data [13, 144]. Mapping urban trees from multiple perspectives could have the potential to overcome the limitation in missing tree instances of street-level imagery presented in 4.3. While street-level imagery remains the most valuable data source for fine-grained urban tree species classification, a combination with aerial imagery or LIDAR data has the potential to provide more accurate localization of trees. Furthermore, a baseline count of urban trees from aerial data could help quantify the percentage of urban trees that were not detected on street-level images. Knowing how many park and private trees are not included in the genera classification could provide insights into how applicable the developed tree genera maps are for biosurveillance efforts. The difference in LIDAR and street-level imagery tree counts could further help to identify areas where a denser street-level imagery coverage could be needed. Additionally, a combination with aerial imagery and LIDAR data could contribute other information, such as tree health (multispectral aerial imagery) or tree structure (LIDAR), to allow for a more comprehensive assessment of the urban forest state in the future.

Similarly, using video imagery instead of street-level imagery could provide more detail of urban scenes leading to more accurate tree location predictions and tree detection of private trees otherwise occluded. MASK R-CNN as well as Yolo are DL architectures that also perform well for instance segmentation in video datasets. Novel methods such as Optical flow, Structure from Motion (SFM) or Simultaneous localisation and mapping (SLAM) could provide the basis for an improved geolocation module. Such research could be beneficial for smart urban forest management applications that require a higher level of detail but only need assessments for smaller urban areas, such as genera classification for areas directly located at ports or commercial zones of very high FIS risk.

4.4.2 Crowd-sourcing and street-level imagery collection

To overcome limitations caused by low or no availability of spatially coherent, high resolution street-level imagery, new avenues of imagery collection in urban areas could be evaluated. Data could, for example, be crowd-sourced through mobile phone applications or citizen scientist campaigns, engaging citizens in smart urban forest management. Including citizens in the acquisition of imagery could add the benefit of also covering private areas or parks with imagery. Adding these images in the presented workflow could allow urban forest managers to include otherwise undetected private and park trees in the tree inventory (s. section 4.3).

Alternatively, street-level imagery could be captured through professional campaigns using the same cameras, sensors, techniques and standards given by GSV. This could give urban forest managers more control over the season, spacing or frequency of image acquisition. Collected data through citizen engagement or smart urban forest management campaigns could be hosted for free through services like Mapillary and processed as presented in this research. Future research questions could include: How does the cost of acquiring tree inventory data through manual measurements, LIDAR or hyperspectral surveys compare to data generated through street-level images? How do tree detection and geolocation results change with different spacing between camera positions of image capture? How does seasonality, image resolution and other external circumstances like weather influence tree detection and classification results? How frequently can tree inventories be updated

using the proposed workflow from data caption to the final map product?

4.4.3 Methodological adaption for bio-surveillance

The achieved results for fine-grained tree genus classification raise the question: What are future possibilities and where are limits for urban tree assessments with DL architectures and street-level imagery? Tree genera identified in this research could be extended in number and characteristics, e.g. through inclusion of more native genera as outlined above (s. section 4.3) or rare genera found in urban environments. A next step after assessing tree classification at the genus level could be a more detailed tree species classification, focusing on species particularly endangered by arriving FIS or vulnerable to climate change impacts, such as droughts and rising temperatures. Additionally, specifically for bio-surveillance and early detection of tree pests and pathogens, it would be interesting to directly assess tree health from street-level imagery. Impacts of tree pests and pathogens, such as large holes in the bark for trees caused by ALB can already be visually identified by professionals from photographs. Generating training datasets and developing a model to detect different FIS impacts, for example defoliation, discolouring, holes, in combination with regular imagery captures could open up opportunities to detect and contain the spread of FIS at an early stage.

4.4.4 Green smart cities of the future

Finally, the generated dataset opens up a range of avenues for future research in smart urban forest management. Accurate masking of trees and the position of generated tree masks could provide the basis for a variety of quantitative urban forest assessments. For example, a detailed analysis of tree masks to extract information about tree structure could allow estimating ecosystem function using proxies. Qualitatively comparing different generated tree masks could provide valuable information on aesthetic appeal of different trees types and shapes, providing insights into cultural ecosystem services provided by street trees. Location information of urban trees in combination with a genus label can help answer questions such as: What effect do urban trees have on the livability and resilience of our cities [53]? What is the value and range of ecosystem services provided by urban forests [60]?

How can urban trees help adapt and mitigate impacts of climate change on cities [38]? How do urban trees contribute to human health and well being [137]? In conclusion, street-level imagery in combination with DL brings a new perspective to assessing urban forests.

Bibliography

[1] nightrome/cocostuff: The official homepage of the COCO-Stuff dataset. URL https://github.com/nightrome/cocostuff. → page 111

[2] COCO - Common Objects in Context, December 2018. URL http://cocodataset.org/#download. → pages xiii, 111

[3] Martın Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: A system for large-scale machine learning. page 21. → page 117

[4] Waleed Abdulla. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow: matterport/Mask_rcnn, 2017. URL https://github.com/matterport/Mask_RCNN. original-date: 2017-10-19T20:28:34Z. → pages 35, 36

[5] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz, and R. Szeliski. Reconstructing Rome. *Computer*, 43(6):40–47, June 2010. ISSN 0018-9162. doi:10.1109/MC.2010.175. → page 29

[6] Michael Alonzo, Bodo Bookhagen, and Dar A. Roberts. Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sensing of Environment*, 148:70–83, May 2014. ISSN 0034-4257. doi:10.1016/j.rse.2014.03.018. URL http://www.sciencedirect.com/science/article/pii/S0034425714001047. → pages 7, 27

[7] Alexis A. Alvey. Promoting and preserving biodiversity in the urban forest. *Urban Forestry & Urban Greening*, 5(4):195–201, December 2006. ISSN 1618-8667. doi:10.1016/j.ufug.2006.09.003. URL

http://www.sciencedirect.com/science/article/pii/S1618866706000732. → pages 2, 54

[8] Mark Ambrose, Frank Koch, and Denis Yemshanov. Modeling Urban Host Tree Distributions for Invasive Forest Pests Using a Multi-Step Approach. *World Conference on Natural Resource Modeling*, June 2016. URL https://scholarexchange.furman.edu/rma/all/presentations/26. → page 5

[9] Myla FJ Aronson, Christopher A. Lepczyk, Karl L. Evans, Mark A. Goddard, Susannah B. Lerman, J. Scott MacIvor, Charles H. Nilon, and Timothy Vargo. Biodiversity in the city: key challenges for urban green space management. *Frontiers in Ecology and the Environment*, 15(4): 189–196, 2017. ISSN 1540-9309. doi:10.1002/fee.1480. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/fee.1480. → page 6

[10] Josselin Aval, Jean Demuynck, Emmanuel Zenou, Sophie Fabre, David Sheeren, Mathieu Fauvel, Karine Adeline, and Xavier Briottet. Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:197–210, December 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2018.09.016. URL http://www.sciencedirect.com/science/article/pii/S0924271618302594. → page 27

[11] Karen Bakker and Max Ritts. Smart Earth: A meta-review and implications for environmental governance. *Global Environmental Change*, 52: 201–211, September 2018. ISSN 0959-3780. doi:10.1016/j.gloenvcha.2018.07.011. URL http://www.sciencedirect.com/science/article/pii/S0959378017313730. → page 3

[12] Nina Bassuk and Thomas Whitlow. Environmental Stress In Street Trees. *Arboricultural Journal*, 12(2):195–201, May 1988. ISSN 0307-1375, 2168-1074. doi:10.1080/03071375.1988.9746788. URL https://www.tandfonline.com/doi/full/10.1080/03071375.1988.9746788. → page 2

[13] Adam Berland and Daniel A. Lange. Google Street View shows promise for virtual street tree surveys. *Urban Forestry & Urban Greening*, 21: 11–15, January 2017. ISSN 1618-8667. doi:10.1016/j.ufug.2016.11.006. URL

http://www.sciencedirect.com/science/article/pii/S1618866716303181. → pages 7, 8, 9, 28, 76, 81, 117

[14] Zhou Bolei. COCO + Places 2017 Challenge, 2017. URL https://places-coco2017.github.io. → pages 33, 35, 36

[15] Steve Branson, Jan Dirk Wegner, David Hall, Nico Lang, Konrad Schindler, and Pietro Perona. From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135:13–30, January 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2017.11.008. URL http://www.sciencedirect.com/science/article/pii/S0924271617303453. → pages 9, 17, 28, 34, 62, 116

[16] Eckehard G. Brockerhoff, Andrew M. Liebhold, Brian Richardson, and David M. Suckling. Eradication of invasive forest insects: concepts, methods, costs and benefits. *New Zealand Journal of Forestry Science. 40 suppl.: S117-S135.*, 40(suppl):S117–S135, 2010. URL https://www.fs.usda.gov/treesearch/pubs/34736. → page 5

[17] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, October 2018. ISSN 08936080. doi:10.1016/j.neunet.2018.07.011. URL http://arxiv.org/abs/1710.05381. arXiv: 1710.05381. → page 62

[18] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. COCO-Stuff: Thing and Stuff Classes in Context. *arXiv:1612.03716 [cs]*, December 2016. URL http://arxiv.org/abs/1612.03716. arXiv: 1612.03716. → pages xiii, 17, 19, 33, 35, 37, 41, 44, 52, 76, 111, 119

[19] Bill Yang Cai, Xiaojiang Li, Ian Seiferling, and Carlo Ratti. Treepedia 2.0: Applying Deep Learning for Large-scale Quantification of Urban Tree Cover. August 2018. URL https://arxiv.org/abs/1808.04754. → pages 9, 18, 35

[20] K. Cai, W. Shao, X. Yin, and G. Liu. Co-segmentation of aircrafts from high-resolution satellite images. In *2012 IEEE 11th International Conference on Signal Processing*, volume 2, pages 993–996, October 2012. doi:10.1109/ICoSP.2012.6491746. → pages 17, 31

[21] Joëlle Salomon Cavin and Christian A. Kull. Invasion ecology goes to town: from disdain to sympathy. *Biological Invasions*, 19(12):3471–3487, December 2017. ISSN 1387-3547, 1573-1464.

doi:10.1007/s10530-017-1588-9. URL https://link.springer.com/article/10.1007/s10530-017-1588-9. → page 4

[22] Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer M. Rao, Troy D. Kelley, Dave Braines, Murat Sensoy, Christopher J. Willis, and Prudhvi Gurram. Interpretability of deep learning models: A survey of results. In *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1–6, August 2017. doi:10.1109/UIC-ATC.2017.8397411. ISSN: null. → page 72

[23] X. Chen, S. Xiang, C. Liu, and C. Pan. Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 11(10):1797–1801, October 2014. ISSN 1545-598X. doi:10.1109/LGRS.2014.2309695. → pages 17, 31

[24] G. Cheng, P. Zhou, and J. Han. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12): 7405–7415, December 2016. ISSN 0196-2892. doi:10.1109/TGRS.2016.2601622. → pages 17, 31

[25] Liang Cheng, Yi Yuan, Nan Xia, Song Chen, Yanming Chen, Kang Yang, Lei Ma, and Manchun Li. Crowd-sourced pictures geo-localization method based on street view images and 3d reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141:72–85, July 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2018.04.006. URL http://www.sciencedirect.com/science/article/pii/S0924271618301102. → page 29

[26] Francois Chollet. *Deep Learning with Python*. Manning Publications Co., Greenwich, CT, USA, 1st edition, 2017. ISBN 978-1-61729-443-3. → pages xii, 11, 12, 13, 14, 17, 35, 118, 119, 120

[27] Sylvain Christin, Éric Hervet, and Nicolas Lecomte. Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10):1632–1644, 2019. ISSN 2041-210X. doi:10.1111/2041-210X.13256. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13256. → page 10

[28] Dan Cireşan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338, August 2012. ISSN 0893-6080. doi:10.1016/j.neunet.2012.02.023. URL http://www.sciencedirect.com/science/article/pii/S0893608012000524. → page 14

[29] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3213–3223, 2016. URL http://openaccess.thecvf.com/content_cvpr_2016/html/Cordts_The_Cityscapes_Dataset_CVPR_2016_paper.html. → page 38

[30] Paul J. Crutzen. The "Anthropocene". In Eckart Ehlers and Thomas Krafft, editors, *Earth System Science in the Anthropocene*, pages 13–18. Springer, Berlin, Heidelberg, 2006. ISBN 978-3-540-26590-0. doi:10.1007/3-540-26590-2_3. URL https://doi.org/10.1007/3-540-26590-2_3. → page 1

[31] Jifeng Dai, Kaiming He, and Jian Sun. Convolutional Feature Masking for Joint Object and Stuff Segmentation. pages 3992–4000, 2015. URL https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Dai_Convolutional_Feature_Masking_2015_CVPR_paper.html. → page 111

[32] Adam G. Dale and Steven D. Frank. The Effects of Urban Warming on Herbivore Abundance and Street Tree Condition. *PLOS ONE*, 9(7): e102996, July 2014. ISSN 1932-6203. doi:10.1371/journal.pone.0102996. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0102996. → page 3

[33] Paul M. Dare. Shadow Analysis in High-Resolution Satellite Imagery of Urban Areas. *Photogrammetric Engineering & Remote Sensing*, 71(2): 169–177, February 2005. ISSN 00991112. doi:10.14358/PERS.71.2.169. URL http://openurl.ingenta.com/content/xref?genre=article&issn=0099-1112&volume=71&issue=2&spage=169. → page 27

[34] Jesse Davis and Mark Goadrich. The Relationship Between Precision-Recall and ROC Curves. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 233–240, New York, NY, USA, 2006. ACM. ISBN 978-1-59593-383-6.

doi:10.1145/1143844.1143874. URL http://doi.acm.org/10.1145/1143844.1143874. event-place: Pittsburgh, Pennsylvania, USA. → page 37

[35] Tahia Devisscher, Lorien Nesbitt, and Adrina C. Bardekjian. Main Findings and Trends of Urbanization | Forestry in the Midst of Global Changes | Taylor & Francis Group. In *Forestry in the Midst of Global Changes*, page 446. Taylor & Francis, December 2018. ISBN 978-1-315-28237-4. URL https://www.taylorfrancis.com/. → page 1

[36] Fábio Duarte and Carlo Ratti. What Big Data Tell Us About Trees and the Sky in the Cities. In Klaas De Rycke, Christoph Gengnagel, Olivier Baverel, Jane Burry, Caitlin Mueller, Minh Man Nguyen, Philippe Rahm, and Mette Ramsgaard Thomsen, editors, *Humanizing Digital Reality: Design Modelling Symposium Paris 2017*, pages 59–62. Springer Singapore, Singapore, 2018. ISBN 978-981-10-6611-5. doi:10.1007/978-981-10-6611-5_6. URL https://doi.org/10.1007/978-981-10-6611-5_6. → pages 9, 10, 28

[37] T Elmqvist, H Setälä, SN Handel, S van der Ploeg, J Aronson, JN Blignaut, E Gómez-Baggethun, DJ Nowak, J Kronenberg, and R de Groot. Benefits of restoring ecosystem services in urban areas. *Current Opinion in Environmental Sustainability*, 14:101–108, June 2015. ISSN 1877-3435. doi:10.1016/j.cosust.2015.05.001. URL http://www.sciencedirect.com/science/article/pii/S1877343515000433. → page 2

[38] Theodore A. Endreny. Strategically growing the urban forest will improve our world. *Nature Communications*, 9(1):1–3, March 2018. ISSN 2041-1723. doi:10.1038/s41467-018-03622-0. URL https://www.nature.com/articles/s41467-018-03622-0. → pages 1, 2, 3, 4, 84

[39] Environment Canada. *An invasive alien species strategy for Canada*. Environment Canada, Ottawa - Ontario, September 2004. → page 55

[40] Rebecca S. Epanchin-Niell, Robert G. Haight, Ludek Berec, John M. Kean, and Andrew M. Liebhold. Optimal surveillance and eradication of invasive species in heterogeneous landscapes. *Ecology Letters*, 15(8):803–812, August 2012. ISSN 1461-0248. doi:10.1111/j.1461-0248.2012.01800.x. URL http:

//onlinelibrary.wiley.com/doi/10.1111/j.1461-0248.2012.01800.x/abstract. → page 77

[41] Gianluca Falco, Marco Pini, and Gianluca Marucco. Loose and Tight GNSS/INS Integrations: Comparison of Performance Assessed in Real Urban Scenarios. *Sensors*, 17(2):255, February 2017. doi:10.3390/s17020255. URL https://www.mdpi.com/1424-8220/17/2/255. → page 51

[42] Marcin Feltynowski, Jakub Kronenberg, Tomasz Bergier, Nadja Kabisch, Edyta Łaszkiewicz, and Michael W. Strohbach. Challenges of urban green space management in the face of using inadequate data. *Urban Forestry & Urban Greening*, 31:56–66, April 2018. ISSN 1618-8667. doi:10.1016/j.ufug.2017.12.003. URL http://www.sciencedirect.com/science/article/pii/S1618866717304569. → page 2

[43] Mirijam Gaertner, John R. U. Wilson, Marc W. Cadotte, J. Scott MacIvor, Rafael D. Zenni, and David M. Richardson. Non-native species in urban environments: patterns, processes, impacts and challenges. *Biological Invasions*, 19(12):3461–3469, December 2017. ISSN 1387-3547, 1573-1464. doi:10.1007/s10530-017-1598-7. URL https://link.springer.com/article/10.1007/s10530-017-1598-7. → page 55

[44] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv:1704.06857 [cs]*, April 2017. URL http://arxiv.org/abs/1704.06857. arXiv: 1704.06857. → page 117

[45] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, 114 (50):13108–13113, December 2017. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1700035114. URL https://www.pnas.org/content/114/50/13108. → page 9

[46] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11): 1231–1237, September 2013. ISSN 0278-3649.

doi:10.1177/0278364913491297. URL https://doi.org/10.1177/0278364913491297. → page 38

[47] S.E Gill, J.F Handley, A.R Ennos, and S Pauleit. Adapting Cities for Climate Change: The Role of the Green Infrastructure. *Built Environment*, 33(1):115–133, March 2007. ISSN 02637960. doi:10.2148/benv.33.1.115. URL http://openurl.ingenta.com/content/xref?genre=article&issn=0263-7960&volume=33&issue=1&spage=115. → page 6

[48] Edward L. Glaeser, Scott Duke Kominers, Michael Luca, and Nikhil Naik. Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life. *Economic Inquiry*, 56(1):114–137, 2018. ISSN 1465-7295. doi:10.1111/ecin.12364. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.12364. → page 9

[49] Tobias Glasmachers. Limits of End-to-End Learning. page 16. → page 58

[50] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Unsupervised Monocular Depth Estimation with Left-Right Consistency. *arXiv:1609.03677 [cs, stat]*, September 2016. URL http://arxiv.org/abs/1609.03677. arXiv: 1609.03677. → pages 29, 37, 38, 51

[51] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, November 2016. ISBN 978-0-262-03561-3. Google-Books-ID: Np9SDQAAQBAJ. → pages xii, 12, 13, 14, 16, 17, 35, 36, 119

[52] Google. Street View – Where We've Been & Where We're Headed Next, 2018. URL https://www.google.ca/streetview/understand/. → page 8

[53] Nancy B. Grimm, Stanley H. Faeth, Nancy E. Golubiewski, Charles L. Redman, Jianguo Wu, Xuemei Bai, and John M. Briggs. Global Change and the Ecology of Cities. *Science*, 319(5864):756–760, February 2008. ISSN 0036-8075, 1095-9203. doi:10.1126/science.1150195. URL https://science.sciencemag.org/content/319/5864/756. → pages 1, 2, 83

[54] Raia Hadsell, Pierre Sermanet, Jan Ben, Ayse Erkan, Marco Scoffier, Koray Kavukcuoglu, Urs Muller, and Yann LeCun. Learning long-range vision for autonomous off-road driving. *Journal of Field Robotics*, 26(2): 120–144, February 2009. ISSN 1556-4967. doi:10.1002/rob.20276. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.20276. → page 16

[55] Terry Hartig and Peter H. Kahn. Living in cities, naturally. *Science*, 352 (6288):938–940, May 2016. ISSN 0036-8075, 1095-9203. doi:10.1126/science.aaf3759. URL https://science.sciencemag.org/content/352/6288/938. → page 1

[56] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Unsupervised Learning. In Trevor Hastie, Robert Tibshirani, and Jerome Friedman, editors, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, pages 485–585. Springer New York, New York, NY, 2009. ISBN 978-0-387-84858-7. doi:10.1007/978-0-387-84858-7_14. URL https://doi.org/10.1007/978-0-387-84858-7_14. → page 11

[57] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, October 2017. doi:10.1109/ICCV.2017.322. → pages xiii, 9, 18, 23, 35, 36, 116, 117

[58] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. pages 770–778, 2016. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html. → pages 62, 117

[59] H. Hirschmuller. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, February 2008. ISSN 0162-8828. doi:10.1109/TPAMI.2007.1166. → page 29

[60] Ngaio Hotte, Lorien Nesbitt, Sara Barron, Judith Cowan, and Zhaohua Cindy Cheng. *The Social and Economic Values of Canada's Urban Forests: A National Synthesis*. UBC Faculty of Forestry University of British Columbia, April 2015. URL http://urbanforestry.sites.olt.ubc.ca/files/2016/09/The-Social-and-Economic-Values-of-Canada\OT1\textquoterights-Urban-Forests-A-National-Synthesis-2015.pdf. → pages 1, 5, 83

[61] Ronghang Hu, Piotr Dollar, Kaiming He, Trevor Darrell, and Ross Girshick. Learning to Segment Every Thing. page 9, November 2017. → page 117

[62] Katherine Isaac, Lee Beaulieu, Cameron Owen, Angela Danyluk, and Ben Mulhall. Urban Forest Strategy. page 60, 2018. → pages 5, 31, 54, 56

[63] Xianyan Jia, Shutao Song, Wei He, Yangzihao Wang, Haidong Rong, Feihu Zhou, Liqiang Xie, Zhenyu Guo, Yuanzhou Yang, Liwei Yu, Tiegang Chen, Guangxiao Hu, Shaohuai Shi, and Xiaowen Chu. Highly Scalable Deep Learning Training System with Mixed-Precision: Training ImageNet in Four Minutes. *arXiv:1807.11205 [cs, stat]*, July 2018. URL http://arxiv.org/abs/1807.11205. arXiv: 1807.11205. → page 64

[64] Nadia Kabisch, Horst Korn, Jutta Stadler, and Aletta Bonn, editors. *Nature-Based Solutions to Climate Change Adaptation in Urban Areas: Linkages between Science, Policy and Practice*. Theory and Practice of Urban Sustainability Transitions. Springer International Publishing, 2017. ISBN 978-3-319-53750-4. doi:10.1007/978-3-319-56091-5. URL https://www.springer.com/gp/book/9783319537504. → page 1

[65] Michael Kampffmeyer, Arnt-Borre Salberg, and Robert Jenssen. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. pages 1–9, 2016. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w19/html/Kampffmeyer_Semantic_Segmentation_of_CVPR_2016_paper.html. → page 18

[66] Jian Kang, Marco Körner, Yuanyuan Wang, Hannes Taubenböck, and Xiao Xiang Zhu. Building instance classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:44–59, November 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2018.02.006. URL http://www.sciencedirect.com/science/article/pii/S0924271618300352. → pages 9, 17, 31

[67] Yinghai Ke and Lindi J. Quackenbush. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *International Journal of Remote Sensing*, 32(17):4725–4747, September 2011. ISSN 0143-1161. doi:10.1080/01431161.2010.494184. URL https://doi.org/10.1080/01431161.2010.494184. → pages 6, 27

[68] Julie Kjeldsen-Kragh Keller and Cecil C Konijnendijk. Short Communication: A Comparative Analysis of Municipal Urban Tree Inventories of Selected Major Cities in North America and Europe. page 8, 2012. → pages 5, 6, 7

[69] Maggi Kelly, Qinghua Guo, Desheng Liu, and David Shaari. Modeling the risk for a new invasive forest disease in the United States: An evaluation of five environmental niche models. *Computers, Environment and Urban Systems*, 31(6):689–710, November 2007. ISSN 0198-9715. doi:10.1016/j.compenvurbsys.2006.10.002. URL http://www.sciencedirect.com/science/article/pii/S0198971506000913. → pages 26, 52, 57

[70] Mate Kisantal, Zbigniew Wojna, Jakub Murawski, Jacek Naruniec, and Kyunghyun Cho. Augmentation for small object detection. *arXiv:1902.07296 [cs]*, February 2019. URL http://arxiv.org/abs/1902.07296. arXiv: 1902.07296. → pages 37, 43, 45

[71] Frank H. Koch, Mark J. Ambrose, Denys Yemshanov, P. Eric Wiseman, and F. D. Cowett. Modeling urban distributions of host trees for invasive forest insects in the eastern and central USA: A three-step approach using field inventory data. *Forest Ecology and Management*, 417:222–236, May 2018. ISSN 0378-1127. doi:10.1016/j.foreco.2018.03.004. URL http://www.sciencedirect.com/science/article/pii/S0378112717316961. → page 55

[72] Vladimir A. Krylov, Eamonn Kenny, and Rozenn Dahyot. Automatic Discovery and Geotagging of Objects from Street View Imagery. *Remote Sensing*, 10(5):661, May 2018. doi:10.3390/rs10050661. URL https://www.mdpi.com/2072-4292/10/5/661. → pages 44, 47

[73] Labelbox Inc. Labelbox: The best way to create and manage training data. URL https://labelbox.com/. → page 33

[74] Yann LeCun, Yoshua Bengio, and T Bell Laboratories. Convolutional Networks for Images, Speech, and Time-Series. In *The handbook of brain theory and neural networks*, pages 255–258. MIT Press, October 1998. → page 14

[75] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 0028-0836, 1476-4687. doi:10.1038/nature14539. URL http://www.nature.com/articles/nature14539. → pages 12, 42

[76] S. Lefèvre, D. Tuia, J. D. Wegner, T. Produit, and A. S. Nassar. Toward Seamless Multiview Scene Analysis From Satellite to Street Level. *Proceedings of the IEEE*, 105(10):1884–1899, October 2017. ISSN 0018-9219. doi:10.1109/JPROC.2017.2684300. → page 29

[77] Songnian Li, Suzana Dragicevic, Francesc Antón Castro, Monika Sester, Stephan Winter, Arzu Coltekin, Christopher Pettit, Bin Jiang, James Haworth, Alfred Stein, and Tao Cheng. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:119–133, May 2016. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2015.10.012. URL http://www.sciencedirect.com/science/article/pii/S0924271615002439. → pages 8, 28

[78] Xiaojiang Li and Carlo Ratti. Using Google Street View for Street-Level Urban Form Analysis, a Case Study in Cambridge, Massachusetts. In Luca D'Acci, editor, *The Mathematics of Urban Morphology*, Modeling and Simulation in Science, Engineering and Technology, pages 457–470. Springer International Publishing, Cham, 2019. ISBN 978-3-030-12381-9. doi:10.1007/978-3-030-12381-9_20. URL https://doi.org/10.1007/978-3-030-12381-9_20. → page 9

[79] Xiaojiang Li, Chuanrong Zhang, Weidong Li, Robert Ricard, Qingyan Meng, and Weixing Zhang. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3):675–685, January 2015. ISSN 1618-8667. doi:10.1016/j.ufug.2015.06.006. URL http://www.sciencedirect.com/science/article/pii/S1618866715000874. → page 9

[80] Xiaojiang Li, Carlo Ratti, and Ian Seiferling. Mapping Urban Landscapes Along Streets Using Google Street View. In Michael P. Peterson, editor, *Advances in Cartography and GIScience*, Lecture Notes in Geoinformation and Cartography, pages 341–356. Springer International Publishing, 2017. ISBN 978-3-319-57336-6. → pages 9, 28

[81] Xiaojiang Li, Carlo Ratti, and Ian Seiferling. Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landscape and Urban Planning*, 169:81–91, January 2018. ISSN 0169-2046. doi:10.1016/j.landurbplan.2017.08.011. URL http://www.sciencedirect.com/science/article/pii/S0169204617301950. → page 9

[82] Xun Li, Wendy Y. Chen, Giovanni Sanesi, and Raffaele Lafortezza. Remote Sensing in Urban Forestry: Recent Applications and Future Directions. *Remote Sensing*, 11(10):1144, January 2019.

doi:10.3390/rs11101144. URL https://www.mdpi.com/2072-4292/11/10/1144. → page 27

[83] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common Objects in Context. *arXiv:1405.0312 [cs]*, May 2014. URL http://arxiv.org/abs/1405.0312. arXiv: 1405.0312. → pages xii, 9, 18, 111

[84] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 936–944, Honolulu, HI, July 2017. IEEE. ISBN 978-1-5386-0457-1. doi:10.1109/CVPR.2017.106. URL http://ieeexplore.ieee.org/document/8099589/. → page 117

[85] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42: 60–88, December 2017. ISSN 1361-8415. doi:10.1016/j.media.2017.07.005. URL http://www.sciencedirect.com/science/article/pii/S1361841517301135. → page 16

[86] David M. Lodge, Paul W. Simonin, Stanley W. Burgiel, Reuben P. Keller, Jonathan M. Bossenbroek, Christopher L. Jerde, Andrew M. Kramer, Edward S. Rutherford, Matthew A. Barnes, Marion E. Wittmann, W. Lindsay Chadderton, Jenny L. Apriesnig, Dmitry Beletsky, Roger M. Cooke, John M. Drake, Scott P. Egan, David C. Finnoff, Crysta A. Gantz, Erin K. Grey, Michael H. Hoff, Jennifer G. Howeth, Richard A. Jensen, Eric R. Larson, Nicholas E. Mandrak, Doran M. Mason, Felix A. Martinez, Tammy J. Newcomb, John D. Rothlisberger, Andrew J. Tucker, Travis W. Warziniack, and Hongyan Zhang. Risk Analysis and Bioeconomics of Invasive Species to Inform Policy and Management. *Annual Review of Environment and Resources*, 41(1):453–488, 2016. doi:10.1146/annurev-environ-110615-085532. URL https://doi.org/10.1146/annurev-environ-110615-085532. → page 4

[87] A. Lopes, S. Oliveira, M. Fragoso, J.A. Andrade, and P. Pedro. Wind Risk Assessment in Urban Environments: The Case of Falling Trees During Windstorm Events in Lisbon. In Katarína Střelcová, Csaba Mátyás, Axel

Kleidon, Milan Lapin, František Matejka, Miroslav Blaženec, Jaroslav Škvarenina, and Ján Holécy, editors, *Bioclimatology and Natural Hazards*, pages 55–74. Springer Netherlands, Dordrecht, 2009. ISBN 978-1-4020-8875-9 978-1-4020-8876-6. doi:10.1007/978-1-4020-8876-6_5. URL http://link.springer.com/10.1007/978-1-4020-8876-6_5. → page 2

[88] Gary M. Lovett, Marissa Weiss, Andrew M. Liebhold, Thomas P. Holmes, Brian Leung, Kathy Fallon Lambert, David A. Orwig, Faith T. Campbell, Jonathan Rosenthal, Deborah G. McCullough, Radka Wildova, Matthew P. Ayres, Charles D. Canham, David R. Foster, Shannon L. LaDeau, and Troy Weldy. Nonnative forest insects and pathogens in the United States: Impacts and policy options. *Ecological Applications*, 26(5):1437–1455, July 2016. ISSN 1939-5582. doi:10.1890/15-1176. URL http://onlinelibrary.wiley.com/doi/10.1890/15-1176/abstract. → pages 4, 5, 77

[89] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152: 166–177, June 2019. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2019.04.015. URL http://www.sciencedirect.com/science/article/pii/S0924271619301108. → pages 8, 28

[90] Daniel W. McKenney, John H. Pedlar, Denys Yemshanov, D. Barry Lyons, Kathy Campbell, and Kate Lawrence. Estimates of the Potential Cost of Emerald Ash Borer (Agrilus planipennis Fairmaire) in Canadian Municipalities. 2012. → page 54

[91] Emily K. Meineke, Robert R. Dunn, Joseph O. Sexton, and Steven D. Frank. Urban Warming Drives Insect Pest Abundance on Street Trees. *PLOS ONE*, 8(3):e59687, March 2013. ISSN 1932-6203. doi:10.1371/journal.pone.0059687. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0059687. → page 2

[92] Jeff Michels, Ashutosh Saxena, and Andrew Y. Ng. High Speed Obstacle Avoidance Using Monocular Vision and Reinforcement Learning. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 593–600, New York, NY, USA, 2005. ACM. ISBN 978-1-59593-180-1. doi:10.1145/1102351.1102426. URL

http://doi.acm.org/10.1145/1102351.1102426. event-place: Bonn, Germany. → page 29

[93] Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed Precision Training. *arXiv:1710.03740 [cs, stat]*, October 2017. URL http://arxiv.org/abs/1710.03740. arXiv: 1710.03740. → page 64

[94] Ariane Middel, Jonas Lukasczyk, Sophie Zakrzewski, Michael Arnold, and Ross Maciejewski. Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landscape and Urban Planning*, 183:122–132, March 2019. ISSN 0169-2046. doi:10.1016/j.landurbplan.2018.12.001. URL http://www.sciencedirect.com/science/article/pii/S0169204618313550. → page 9

[95] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015. ISSN 0028-0836, 1476-4687. doi:10.1038/nature14236. URL http://www.nature.com/articles/nature14236. → page 12

[96] J. Morgenroth, J. Östberg, C. Konijnendijk van den Bosch, A. B. Nielsen, R. Hauer, H. Sjöman, W. Chen, and M. Jansson. Urban tree diversity—Taking stock and looking ahead. *Urban Forestry & Urban Greening*, 15:1–5, January 2016. ISSN 1618-8667. doi:10.1016/j.ufug.2015.11.003. URL http://www.sciencedirect.com/science/article/pii/S1618866715001557. → page 55

[97] Michael J. Mortimer and Brian Kane. Hazard tree liability in the United States: Uncertain risks for owners and professionals. *Urban Forestry & Urban Greening*, 2(3):159–165, January 2004. ISSN 16188667. doi:10.1078/1618-8667-00032. URL https://linkinghub.elsevier.com/retrieve/pii/S1618866704700321. → page 2

[98] Giorgos Mountrakis, Jun Li, Xiaoqiang Lu, and Olaf Hellwich. Deep learning for remotely sensed data. *ISPRS Journal of Photogrammetry and*

Remote Sensing, 145:1–2, November 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2018.08.011. URL http://www.sciencedirect.com/science/article/pii/S0924271618302302. → pages 17, 30

[99] Nikhil Naik, Scott Duke Kominers, Ramesh Raskar, Edward L Glaeser, and César A Hidalgo. Do People Shape Cities, or Do Cities Shape People? The Co-evolution of Physical, Social, and Economic Change in Five Major U.S. Cities. page 38, October 2015. → page 9

[100] Nikhil Naik, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and César A. Hidalgo. Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29): 7571–7576, July 2017. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1619003114. URL https://www.pnas.org/content/114/29/7571. → page 9

[101] Lorien Nesbitt, Ngaio Hotte, Sara Barron, Judith Cowan, and Stephen R. J. Sheppard. The social and economic value of cultural ecosystem services provided by urban forests in North America: A review and suggestions for future research. *Urban Forestry & Urban Greening*, 25:103–111, July 2017. ISSN 1618-8667. doi:10.1016/j.ufug.2017.05.005. URL http://www.sciencedirect.com/science/article/pii/S1618866717300456. → page 2

[102] Anders B Nielsen, Johan Östberg, and Tim Delshammar. Review of Urban Tree Inventory Methods Used to Collect Data at Single-Tree Level. page 17, 2014. → pages 5, 6, 7, 26, 27

[103] Sophie Nitoslawski and Peter Duinker. Managing Tree Diversity: A Comparison of Suburban Development in Two Canadian Cities. *Forests*, 7, May 2016. doi:10.3390/f7060119. → page 32

[104] Sophie A. Nitoslawski, Peter N. Duinker, and Peter G. Bush. A review of drivers of tree diversity in suburban areas: Research needs for North American cities. *Environmental Reviews*, 24(4):471–483, December 2016. ISSN 1181-8700, 1208-6053. doi:10.1139/er-2016-0027. URL http://www.nrcresearchpress.com/doi/10.1139/er-2016-0027. → pages 54, 55

[105] Sophie A. Nitoslawski, Nadine J. Galle, Cecil Konijnendijk Van Den Bosch, and James W. N. Steenberg. Smarter ecosystems for smarter

cities? A review of trends, technologies, and turning points for smart urban forestry. *Sustainable Cities and Society*, 51:101770, November 2019. ISSN 2210-6707. doi:10.1016/j.scs.2019.101770. URL http://www.sciencedirect.com/science/article/pii/S2210670719307644. → page 3

[106] David J. Nowak, Satoshi Hirabayashi, Allison Bodine, and Eric Greenfield. Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*, 193:119–129, October 2014. ISSN 0269-7491. doi:10.1016/j.envpol.2014.05.028. URL http://www.sciencedirect.com/science/article/pii/S0269749114002395. → pages 2, 26

[107] NVIDIA. Deep Learning SDK Documentation, October 2019. URL https://docs.nvidia.com/deeplearning/sdk/index.html. → page 64

[108] Vancouver Board of Parks and Recreation. Biodiversity strategy. Technical report, Vancouver Board of Parks and Recreation, Vancouver, 2016. URL https://vancouver.ca/files/cov/biodiversity-strategy.pdf. → page 54

[109] Trudy Paap, Treena I. Burgess, and Michael J. Wingfield. Urban trees: bridge-heads for forest pest invasions and sentinels for early detection. *Biological Invasions*, 19(12):3515–3526, December 2017. ISSN 1387-3547, 1573-1464. doi:10.1007/s10530-017-1595-x. URL https://link.springer.com/article/10.1007/s10530-017-1595-x. → pages 3, 5

[110] Ashlyn L. Padayachee, Ulrike M. Irlich, Katelyn T. Faulkner, Mirijam Gaertner, Şerban Procheş, John R. U. Wilson, and Mathieu Rouget. How do invasive species travel to and through urban environments? *Biological Invasions*, 19(12):3557–3570, December 2017. ISSN 1387-3547, 1573-1464. doi:10.1007/s10530-017-1596-9. URL https://link.springer.com/article/10.1007/s10530-017-1596-9. → pages 2, 4, 26

[111] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010. ISSN 1041-4347. doi:10.1109/TKDE.2009.191. URL http://ieeexplore.ieee.org/document/5288526/. → page 35

[112] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep Face Recognition. In *Procedings of the British Machine Vision Conference 2015*, pages 41.1–41.12, Swansea, 2015. British Machine Vision

Association. ISBN 978-1-901725-53-7. doi:10.5244/C.29.41. URL http://www.bmva.org/bmvc/2015/papers/paper041/index.html. → pages 56, 64

[113] Mason F. Patterson, P. Eric Wiseman, Matthew F. Winn, Sang-mook Lee, and Philip A. Araman. Effects of photographic distance on tree crown atributes calculated using urbancrowns image analysis software. *Arboriculture & Urban Forestry 37(4):173-179*, 37(4):173–179, 2011. URL https://www.fs.usda.gov/treesearch/pubs/39433. → page 7

[114] Jill D Pokorny. Urban Tree Risk Management: A Community Guide to Program Design and Implementation. Technical report, 2003. → page 2

[115] Therese M. Poland and Deborah G. McCullough. Emerald Ash Borer: Invasion of the Urban Forest and the Threat to North America's Ash Resource. *Journal of Forestry*, 104(3):118–124, March 2006. ISSN 0022-1201. doi:10.1093/jof/104.3.118. URL https://academic.oup.com/jof/article/104/3/118/4598702. → page 5

[116] Rassati Davide, Faccoli Massimo, Petrucco Toffolo Edoardo, Battisti Andrea, Marini Lorenzo, and Clough Yann. Improving the early detection of alien wood-boring beetles in ports and surrounding forests. *Journal of Applied Ecology*, 52(1):50–58, September 2014. ISSN 0021-8901. doi:10.1111/1365-2664.12347. URL https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2664.12347. → pages 5, 55

[117] M. K. Ridd. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities†. *International Journal of Remote Sensing*, 16(12): 2165–2185, August 1995. ISSN 0143-1161. doi:10.1080/01431169508954549. URL https://doi.org/10.1080/01431169508954549. → page 6

[118] Jérôme Rousselet, Charles-Edouard Imbert, Anissa Dekri, Jacques Garcia, Francis Goussard, Bruno Vincent, Olivier Denux, Christelle Robinet, Franck Dorkeld, Alain Roques, and Jean-Pierre Rossi. Assessing Species Distribution Using Google Street View: A Pilot Study with the Pine Processionary Moth. *PLOS ONE*, 8(10):e74918, October 2013. ISSN 1932-6203. doi:10.1371/journal.pone.0074918. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074918. → page 9

[119] Stuart J. Russell and Peter Norvig. *Artificial Intelligence : A Modern Approach*. Malaysia; Pearson Education Limited,, 2016. URL http://thuvienso.thanglong.edu.vn/handle/DHTL_123456789/4010. → page 11

[120] Jacques Régnière, Vince Nealis, and Kevin Porter. Climate suitability and management of the gypsy moth invasion into Canada. *Biological Invasions*, 11(1):135–148, January 2009. ISSN 1387-3547, 1573-1464. doi:10.1007/s10530-008-9325-z. URL https://link.springer.com/article/10.1007/s10530-008-9325-z. → page 57

[121] A. L. Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3):210–229, July 1959. ISSN 0018-8646. doi:10.1147/rd.33.0210. → page 11

[122] Juergen Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61:85–117, January 2015. ISSN 08936080. doi:10.1016/j.neunet.2014.09.003. URL http://arxiv.org/abs/1404.7828. arXiv: 1404.7828. → pages 17, 33

[123] Tracey-Lee Schwets and Robert D Brown. Form and structure of maple trees in urban environments. *Landscape and Urban Planning*, 46(4): 191–201, February 2000. ISSN 0169-2046. doi:10.1016/S0169-2046(99)00072-9. URL http://www.sciencedirect.com/science/article/pii/S0169204699000729. → page 71

[124] Ian Seiferling, Nikhil Naik, Carlo Ratti, and Raphäel Proulx. Green streets Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, 165:93–101, September 2017. ISSN 0169-2046. doi:10.1016/j.landurbplan.2017.05.010. URL http://www.sciencedirect.com/science/article/pii/S0169204617301147. → pages 9, 28

[125] Hansi Senaratne, Amin Mobasheri, Ahmed Loai Ali, Cristina Capineri, and Mordechai (Muki) Haklay. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1):139–167, January 2017. ISSN 1365-8816. doi:10.1080/13658816.2016.1189556. URL https://doi.org/10.1080/13658816.2016.1189556. → page 8

[126] USDA Forest Service. i-Tree Streets, October 2018. URL http://www.itreetools.org/streets/index.php. → pages 5, 78

[127] Andrew J. Shatz, John Rogan, Florencia Sangermano, Jennifer Miller, and Arthur Elmes. Modeling the risk of spread and establishment for Asian longhorned beetle (Anoplophora glabripennis) in Massachusetts from 2008-2009. *Geocarto International*, 31(8):813–831, September 2016. ISSN 1010-6049. doi:10.1080/10106049.2015.1086901. URL https://doi.org/10.1080/10106049.2015.1086901. → page 57

[128] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, October 2017. ISSN 1476-4687. doi:10.1038/nature24270. URL https://www.nature.com/articles/nature24270. → page 12

[129] C. Small. Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*, 22(7):1305–1334, January 2001. ISSN 0143-1161. doi:10.1080/01431160151144369. URL https://doi.org/10.1080/01431160151144369. → page 27

[130] Fiona Steele. Urban Forest Climate Adaptation Framework for Metro Vancouver. Technical report, Metro Vancouver, February 2016. → page 31

[131] Iain Stewart and Tim Oke. Thermal differentiation of local climate zones using temperature observations from urban and rural field sites. In *Urban Climate*, page 7, Keystone, Colorado, August 2010. → page 31

[132] Philip Stubbings, Joe Peskett, Francisco Rowe, and Dani Arribas-Bel. A Hierarchical Urban Forest Index Using Street-Level Imagery and Deep Learning. *Remote Sensing*, 11(12):1395, January 2019. doi:10.3390/rs11121395. URL https://www.mdpi.com/2072-4292/11/12/1395. → pages 8, 9, 28, 37

[133] Mapillary AB Sweden. Map data at scale from street-level imagery. URL https://www.mapillary.com/. → page 34

[134] Beau Tippetts, Dah Jye Lee, Kirt Lillywhite, and James Archibald. Review of stereo vision algorithms and their suitability for resource-limited systems. *Journal of Real-Time Image Processing*, 11(1):5–25, January 2016. ISSN 1861-8219. doi:10.1007/s11554-012-0313-2. URL https://doi.org/10.1007/s11554-012-0313-2. → page 29

[135] K. V. Tubby and J. F. Webber. Pests and diseases threatening urban trees under a changing climate. *Forestry: An International Journal of Forest Research*, 83(4):451–459, October 2010. ISSN 0015-752X. doi:10.1093/forestry/cpq027. URL https://academic.oup.com/forestry/article/83/4/451/546615. → page 54

[136] Alan M. Turing. Computing Machinery and Intelligence. In Robert Epstein, Gary Roberts, and Grace Beber, editors, *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, pages 23–65. Springer Netherlands, Dordrecht, 2009. ISBN 978-1-4020-6710-5. doi:10.1007/978-1-4020-6710-5_3. URL https://doi.org/10.1007/978-1-4020-6710-5_3. → page 11

[137] Jessica B. Turner-Skoff and Nicole Cavender. The benefits of trees for livable and sustainable communities. *PLANTS, PEOPLE, PLANET*, 1(4): 323–335, 2019. ISSN 2572-2611. doi:10.1002/ppp3.39. URL https://nph.onlinelibrary.wiley.com/doi/abs/10.1002/ppp3.39. → pages 1, 84

[138] Konstantinos Tzoulas, Kalevi Korpela, Stephen Venn, Vesa Yli-Pelkonen, Aleksandra Kaźmierczak, Jari Niemela, and Philip James. Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. *Landscape and Urban Planning*, 81(3):167–178, June 2007. ISSN 0169-2046. doi:10.1016/j.landurbplan.2007.02.001. URL http://www.sciencedirect.com/science/article/pii/S0169204607000503. → page 2

[139] DESA UN. World urbanization prospects: The 2018 revision. *United Nations Department of Economics and Social Affairs, Population Division: New York, NY, USA*, 2019. → page 1

[140] M. van den Bosch and Å Ode Sang. Urban natural environments as nature-based solutions for improved public health – A systematic review of reviews. *Environmental Research*, 158:373–384, October 2017. ISSN 0013-9351. doi:10.1016/j.envres.2017.05.040. URL http://www.sciencedirect.com/science/article/pii/S0013935117310241. → pages 2, 26

[141] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The iNaturalist Species Classification and Detection Dataset. *arXiv:1707.06642*

[cs], July 2017. URL http://arxiv.org/abs/1707.06642. arXiv: 1707.06642. → page 56

[142] Kathleen T. Ward and Gary R. Johnson. Geospatial methods provide timely and comprehensive urban forest information. *Urban Forestry & Urban Greening*, 6(1):15–22, February 2007. ISSN 1618-8667. doi:10.1016/j.ufug.2006.11.002. URL http://www.sciencedirect.com/science/article/pii/S161886670600080X. → page 7

[143] Michael Waskom, Olga Botvinnik, Drew O'Kane, Paul Hobson, Saulius Lukauskas, David C Gemperline, Tom Augspurger, Yaroslav Halchenko, John B. Cole, Jordi Warmenhoven, Julian de Ruiter, Cameron Pye, Stephan Hoyer, Jake Vanderplas, Santi Villalba, Gero Kunter, Eric Quintero, Pete Bachant, Marcel Martin, Kyle Meyer, Alistair Miles, Yoav Ram, Tal Yarkoni, Mike Lee Williams, Constantine Evans, Clark Fitzgerald, Brian, Chris Fonnesbeck, Antony Lee, and Adel Qalieh. mwaskom/seaborn: v0.8.1 (September 2017), September 2017. URL https://zenodo.org/record/883859#.XeMzlS2ZNTZ. → page 58

[144] Jan D. Wegner, Steven Branson, David Hall, Konrad Schindler, and Pietro Perona. Cataloging Public Objects Using Aerial and Street-Level Images - Urban Trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6014–6023, 2016. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Wegner_Cataloging_Public_Objects_CVPR_2016_paper.html. → pages 3, 6, 7, 8, 9, 10, 17, 18, 28, 29, 40, 50, 60, 81, 112

[145] Lynne M. Westphal. Social Aspects of Urban Forestry: Urban Greening and Social Benefits: a Study of Empowerment Outcomes. *Journal of Arboriculture 29(3):137-147*, 29(3), 2003. URL https://www.fs.usda.gov/treesearch/pubs/14148. → page 2

[146] Hanfa Xing, Yuan Meng, Zixuan Wang, Kaixuan Fan, and Dongyang Hou. Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141:237–251, July 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2018.04.025. URL http://www.sciencedirect.com/science/article/pii/S0924271618301333. → pages 17, 31

[147] Denys Yemshanov, Frank H. Koch, Mark Ducey, and Klaus Koehler. Trade-associated pathways of alien forest insect entries in Canada.

Biological Invasions, 14(4):797–812, April 2012. ISSN 1573-1464. doi:10.1007/s10530-011-0117-5. URL https://doi.org/10.1007/s10530-011-0117-5. → pages 4, 57

[148] Dameng Yin and Le Wang. How to assess the accuracy of the individual tree-based forest inventory derived from remotely sensed data: a review. *International Journal of Remote Sensing*, 37(19):4521–4553, October 2016. ISSN 0143-1161. doi:10.1080/01431161.2016.1214302. URL https://doi.org/10.1080/01431161.2016.1214302. → page 40

[149] Paul A. Zandbergen and Sean J. Barbeau. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *The Journal of Navigation*, 64(3):381–399, July 2011. ISSN 1469-7785, 0373-4633. doi:10.1017/S0373463311000051. URL https://www.cambridge.org/core/journals/journal-of-navigation/article/positional-accuracy-of-assisted-gps-data-from-highsensitivity-gpsenabled-mobile-phones/E1EE20CD1A301C537BEE8EC66766B0A9. → page 60

[150] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. *arXiv:1710.09412 [cs, stat]*, October 2017. URL http://arxiv.org/abs/1710.09412. arXiv: 1710.09412. → page 63

[151] L. Zhang, L. Zhang, and B. Du. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40, June 2016. ISSN 2168-6831. doi:10.1109/MGRS.2016.2540798. → pages 17, 31

[152] Xin Zhang, Gui-Song Xia, Qikai Lu, Weiming Shen, and Liangpei Zhang. Visual object tracking by correlation filters and online learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140:77–89, June 2018. ISSN 0924-2716. doi:10.1016/j.isprsjprs.2017.07.009. URL http://www.sciencedirect.com/science/article/pii/S0924271617301715. → pages 17, 30, 117

[153] Zhen Zhen, Lindi J. Quackenbush, and Lianjun Zhang. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sensing*, 8(4):333, April 2016. doi:10.3390/rs8040333. URL https://www.mdpi.com/2072-4292/8/4/333. → page 27

[154] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer. Deep Learning in Remote Sensing: A Comprehensive Review and List of

Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, December 2017. ISSN 2168-6831. doi:10.1109/MGRS.2017.2762307. → pages 10, 30

[155] Johan Östberg, Tim Delshammar, Björn Wiström, and Anders Busse Nielsen. Grading of Parameters for Urban Tree Inventories by City Officials, Arborists, and Academics Using the Delphi Method. *Environmental Management*, 51(3):694–708, March 2013. ISSN 1432-1009. doi:10.1007/s00267-012-9973-8. URL https://doi.org/10.1007/s00267-012-9973-8. → page 78

Appendix A

Additional publications and presentations

- [*oral presentation*] Lumnitz, S. (2018): Monitoring Urban Trees in a Changing World: Instance Segmentation with Deep Learning and Google Street View Imagery. BioSAFE Symposium. Vancouver, Canada.
- [*oral presentation*] Lumnitz, S. (2018): Bio-surveillance in a Changing World: Monitoring Urban Trees with Deep Learning and Street View Imagery. Entomology 2018 - ESA, ESC and ESBC Joint Annual Meeting. Vancouver, Canada.
- [*oral presentation*] Lumnitz, S. (2019): A Geographers Journey into AI: mapping Urban Trees from Scratch. Scipy. Austin, USA.
  URL: https://www.youtube.com/watch?v=pMiyZfhItZY
- [*oral presentation*] Lumnitz, S. (2019): Mapping urban trees with deep learning and street-level imagery - a story of geospatial open source software. UBC GIS day. Vancouver, Canada.
- Lafond, V., Lingua, F., Lumnitz, S., Paradis, G., Srivastava, V. and Griess, V. (2019): Challenges and opportunities in developing decision support systems for risk assessment and management of forest invasive alien species. Accepted: Environmental Reviews.

Appendix B

Theoretical background on developing Mask R-CNN

In this project, MASK R-CNN was used to detect and outline trees in imagery by generating bounding boxes and binary segmentation masks for each tree instance. These tree detections were then used for both tree location and genus classification (s. chapter 2 and 3). As the most central part of the proposed workflow, the MASK R-CNN architecture, it's training and evaluation process as well as the chosen training data is described in detail in this section.

B.1 Training, development and evaluation data generation

The presented workflow relied on two main data sources used to train, develop and evaluate the tree detection (MASK R-CNN) and the genus classification (RESNET50) models: 1) openly available benchmark datasets for training, i.e. the COCO Stuff annotated dataset, and second, generated street-level panoramas and corresponding tree labels. 2) Acquired street-level imagery for training, development, testing and inference purposes.

B.1.1 COCO Stuff dataset

The COCO Stuff dataset is the most expansive dataset with pixel-level annotations for ”stuff classes” to date [18]. Stuff classes describe classes of objects that are amorphous and consist of less to no distinct parts, e.g. sky, grass [31]. Stuff classes are opposed to ”thing classes”, e.g. car, human, which describe objects of a specific size and shape, and are made of identifiable parts, e.g. wheels and door for the thing class ”car” [83]. In the context of computer vision, the ”tree class” is commonly referred to as a stuff class, owing to a tree's amorphous structure composed of typically trunk, branches, leafs but also through shining background noise. COCO Stuff consists of 164,000 images from the original COCO dataset with pixel-level annotations for 172 classes: 91 stuff classes and 1 unlabeled class, additionally to the traditional 80 thing classes [18]. Figure B.1 provides an example of images and image annotations of images collected in COCO Stuff.

Figure B.1: COCO Stuff image and segmentation mask examples. The COCO Stuff dataset is the most expansive dataset with pixel-level annotations for stuff classes. The COCO Stuff dataset is used as training data in the presented workflow. These are three example images with corresponding Stuff segmentation masks retrieved from the COCO 2017 Stuff Segmentation Task Challenge [2, 18]

The dataset can be downloaded via the official COCO dataset website [2]. After downloading the image datasets, COCO Stuff data structures and annotations can be accessed, manipulated and integrated into our DL workflow using the official COCO Stuff API [1].

In the presented workflow, this extensive stuff dataset was used to extend the smaller annotated tree datasets for training the instance segmentation DNN model. All 36,500 images containing tree objects and pixel-level annotations were strate-

gically extracted from COCO Stuff. Extracted images are used as basic training and development datasets in the layered training approach for the tree detection model, and therefore complement the smaller development and test datasets used for fine-tuning (presented in the following section B.1.2).

B.1.2 Street-level panoramas and annotations

A developed data processing pipeline enabled a dense download of all GSV panoramas in a defined area of interest, and relied on the Python implementation of the official Google API. In order to download a single panorama the API required a set of coordinates of interest as main user input. The given user input was then compared to recorded GSV car or camera positions at time of image capture and metadata, including coordinate points, for the panorama closest to the user input coordinates was returned. In a subsequent step, metadata information was then used to download the actual panorama in various resolutions. Leveraging the API the proposed processing pipeline consisted on the following steps of operation:

1. **Generating a grid of coordinates over the area of interest**: A grid of coordinate points was generated to allow a dense download of all panoramas over a given area of interest. The upper left and lower right corner coordinate defined the size of the grid and allowed for a download of panoramas in the resulting square. A new coordinate point was generated every 6 meters to ensure all panoramas of an area were downloaded, since the common distance between two points of panorama capture was 15 meters [144].

2. **Crawling metadata of nearest panorama:** The generated coordinate grid was now used to crawl and download the metadata of all corresponding nearest panoramas.

3. **Downloading panorama data:** All panoramas of interest were densely downloaded based on the crawled metadata and the Google API. Panoramas were stored as a coherent dataset and further processed.

4. **Dataset annotation:** Training, development and testing datasets were build through semi-atuomatic and manual annotation of a subset of GSV images.

(a) For **tree detection model** development a small subset of 120 panoramas including roughly 1200 private and public trees located in Metro Vancouver, Canada and Pasadena, USA were annotated. Trees were manually annotated using Labelbox, software to annotate images and mask objects, in combination with a custom designed label template. Figures B.2 and B.3 show an annotated example panorama and details of annotation masks.

(b) For **tree genus classification model** development tree masks were automatically generated through inference of the trained tree detection model on all available imagery. Resulting masks were combined with information from existing tree inventories based on the spatial location of single trees. For a more detailed description see chapter 3.

Figure B.2: Manual image annotation with Labelbox. Trees on the full panorama are manually annotated using Labelbox, an open source software package. Masks are created by manually outlining tree instances. Labels are saved in *.json* format. (Imagery source: Google Street View 2018)

The presented data generation pipeline based on these four steps resulted in annotated datasets which were used as training, development and test datasets in training the tree detection and genus classification models. Step one to three can be used in combination with our fully trained model to update or generate urban tree inventories in future.

Figure B.3: Google Street View panorama and tree annotations. Trees on the full panorama are annotated using Labelbox. Original panoramas are of the size 6656x3328 pixels but will be downsized and halved for the purpose of training. The black stripes represent an automatic application of zero padding in order to transform images to the same format for training (Imagery source: Google Street View 2018).

Generated datasets are further preprocessed as part of both tree detection and tree genus classification models:

1. **Automated bounding box generation:** In addition to tree labels and masks,

bounding boxes were automatically computed by drawing a rectangular bounding box with the best fit around each separate mask. This approach allowed for an easy augmentation of tree detection datasets without the need to update each bounding box according to the respective image and mask transformation.

2. **Resizing images:** All images and corresponding labels in a dataset were resized to 1024x1024 pixels for tree detection and 256x256 pixels for tree classification. Zero padding was added at the sides when using the COCO Stuff dataset to preserve the original aspect ration. GSV panoramas were split in half and downsized from 6656x3328 pixels. Resizing images to the same size was vital to gradient descent.

3. **Resizing masks:** Individual instance masks were downsized to a resolution of 56x56 pixels in order to save memory space during one step of training and accelerate the training process. Figure B.4 gives an example of reshaped and normalized tree masks.

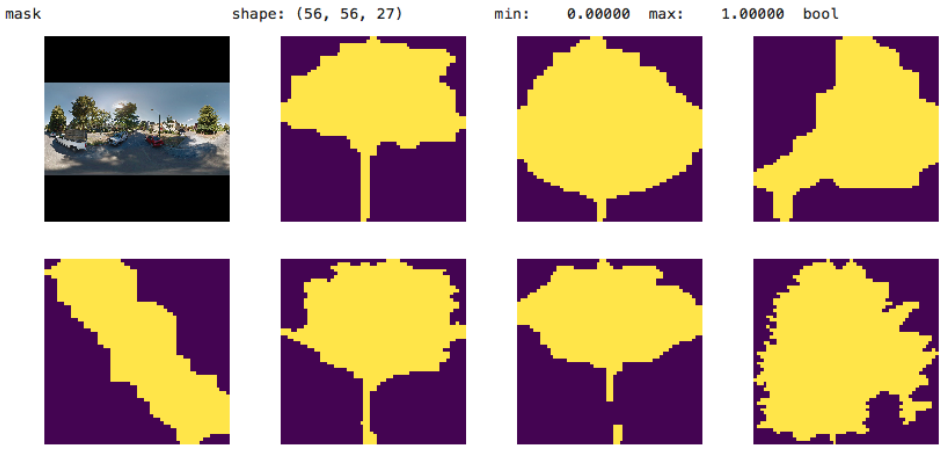


Figure B.4: Reshaped mask annotations. Annotations need to be reshaped to 56x56 pixels in order to reduce memory needed during training. This is an effective method to speed up the training process.

These preprocessing steps were particularly important to optimize the training process on high-resolution imagery for tree detection. All of the steps above sig-

nificantly decreased the time needed for training and allow me to use a single GPU with 11 GB of memory.

B.2 Mask R-CNN for tree detection

B.2.1 Mask R-CNN framework

MASK R-CNN is a framework specialized on instance segmentation, the computer vision task to detect and outline seperate objects in an image [57]. Figure B.5 shows the architecture of the MASK R-CNN framework.

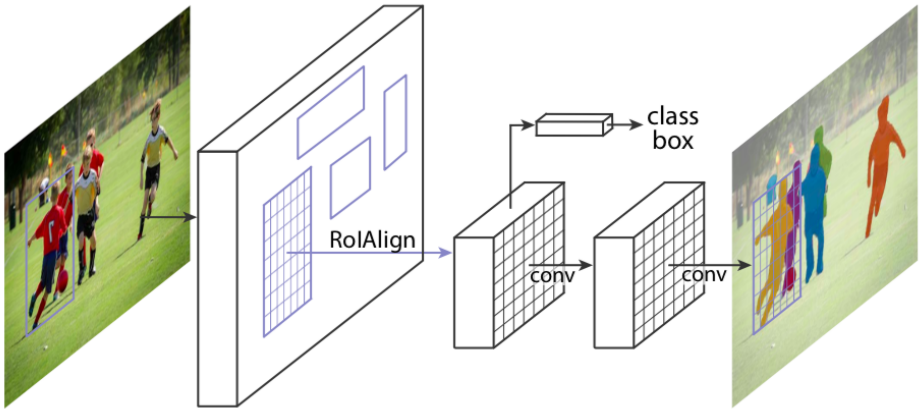


Figure B.5: The Mask R-CNN framework for instance segmentation. An input image is transformed using a Feature Pyramind Network (FPN) and a RESNET101 core. Resulting features are additionally realigned with the input image through a Region of Interest (ROI)Align operation which ensures that further generated masks are in the right position on the input image. Class and bounding box are predicted separately branching off from the first convolution layer. Binary masks are predicted in parallel for every ROI using an additional convolution layer [57].

MASK R-CNN extends FASTER R-CNN used by Branson et al. [15] by adding a segmentation mask prediction for each detected instance or object. It therefore allowed me to classify and detect single urban tree instances, create bounding boxes

and segmentation masks surrounding the individual trees [57]. MASK R-CNN can be conceptualizes as a two stage algorithm: 1) The first part is also referred to as a Region Proposal Network (RPN), predicting multiple ROI. A convolutional backbone architecture, RESNET101 coupled with a FPN, allows to generate multiple anchor boxes at different scales [58, 84]. 2) In the so called head of the model, features are then extracted from each ROI which can be associated with the relevant class. All in parallel, the head extracts class and bounding box values, leveraging a ROIPool operation, and creates a binary mask for each detected object using a ROIAlign operation [152].

The implemented network head of MASK R-CNN decouples the classification from the convolution mask prediction branch. Class, bounding box and binary mask are therefore predicted separately and in parallel for every ROI [57]. In comparison, the classification task usually depends on a previously predicted mask in other state-of-the-art instance segmentation frameworks [44]. Due to the separation of these different tasks, MASK R-CNN is performing faster and with higher accuracy compared to prior and other state-of-the-art systems and was chosen for this research project [61]. He et al. [57] provide more detailed information on the architecture of MASK R-CNN. The original architecture can be openly accessed online and is implemented in Keras with a Tensorflow backend [3].

B.3 Evaluation strategy

B.3.1 Architecture evaluation for tree detection

Detecting trees on street-level imagery is a binary classification and image segmentation problem. This means the CNN learns whether a pixel in an image is a tree or a non-tree pixel and assigns tree pixels to exactly one tree object, which is located at a specific position in the image. Owing to findings of Berland and Lange [13], where humans could identify trees and tree genera from the GSV image with 90% overlap to the field survey, and the promise of current CNN architectures to near human-level performance, MASK R-CNN was assumed to be able to preform this task. However, it was important to test whether or not the chosen MASK R-CNN

architecture was powerful enough for the task at hand.

A common test used to assess weather an architecture can perform a task is to train the architecture to over-fit a set of example images [26]. The goal of training a NN is to strike the balance between over-fitting to it's training data and introducing a bias due to inaccurate representation of the training data (under-fitting) and therefore to generalize the task at hand. Therefore, before a NN is trained to generalize it needs to be assessed if the architecture is suited to over-fit on a specific training set.

In a preliminary experiment, MASK R-CNN was trained with 20 annotated GSV images and inference was evaluated on the same images. Figure B.6 clearly indicates that MASK R-CNN correctly detects and masks all trees after only 8 epochs of training. This implied that MASK R-CNN could be used for the task of tree instance segmentation with the right training strategy and a good representation of the target class.

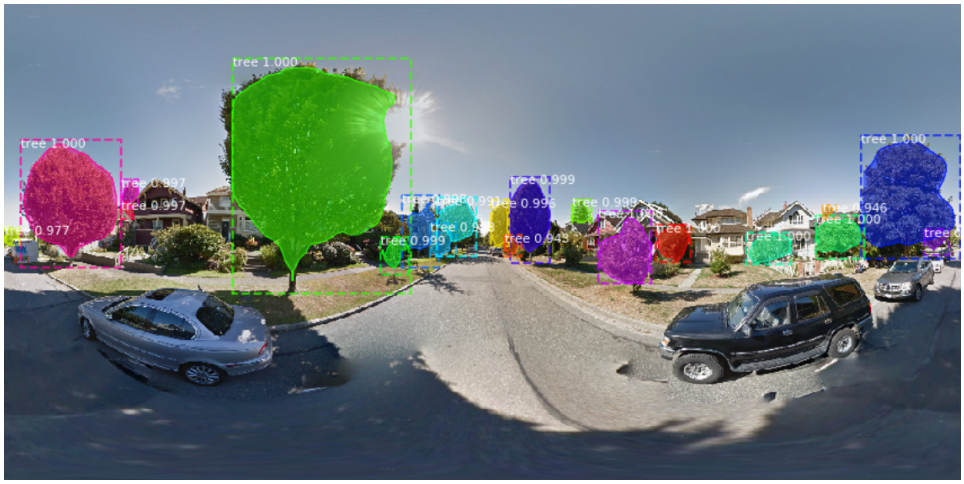


Figure B.6: Over-fitting Mask R-CNN. I purposely over-fit the model in order to test whether or not MASK R-CNN is powerful enough for the task of tree instance segmentation.

B.3.2 Evaluation metrics for tree detection

To evaluate the performance of tree detection and masking with MASK R-CNN the Mean Average Precision (*mAP*), Average Precision 50, and 75 (AP_{50}, AP_{75}) were computed.

A common way to determine if a detection proposal is right is to asses IOU [51]. A set of of proposed object pixels A and the set of true object pixels B were compared:

$$IoU(A,B) = \frac{\text{area of overlap}}{\text{area of union}} = \frac{A \cap B}{A \cup B} \tag{B.1}$$

Unless otherwise specified, *AP* is commonly averaged over multiple IOU values [18]. In this work, 10 IOU thresholds of .50:.05:.95 are used. This is a break from tradition, where *AP* is computed at a single IOU of .50 (which corresponds to the metric $APIoU$=.50). Averaging over IOUs rewards detectors with better localization evaluation.

B.4 Training strategy

B.4.1 Feature extraction and fine-tuning

For training MASK R-CNN two commonly used training strategies using pre-trained models were applied, 1) feature extraction and 2) fine-tuning [26].

In feature extraction new models are build upon already trained NN's, by downloading trained weights, i.e. mathematical features of representations (s. section 1.2.2). These weights derive from already learned very generic representations found in the first layers (1-80) of the RESNET101 core (s. section B.2.1). In contrast to the more complex representations learned in end layers or so called heads, these simple base representations can be reused. The level of re-usability of these representations or features decreases with the depth of layers in the architecture. The chosen implementation of the MASK R-CNN architecture allowed to download openly sourced weights with which can be used to initialize the training process. These weights are derived from the same MASK R-CNN architecture that has previously been trained on the official COCO dataset (which differs from the COCO stuff

dataset).

Fine-tuning describes the process in which MASK R-CNN is retrained for a new task, starting from transferred COCO weights. A best-practice fine-tuning workflow starts by only training the heads, followed by only training the last 4 RESNET101 layers and lastly completed by training only heads a second time. This approach optimizes and reduces the amount of training data and time needed [26]. Only the last layers of the RESNET101 core, the so called heads were adjusted in the training process. Generic layers were frozen, i.e. excluded in the updating process described in section 1.2.2), and only the later, top layers were trained to fit the problem of tree detection. Representations derived from the heads were therefore fine-tuned to tree shapes. Since tree shapes are very specific and in common benchmark datasets like COCO rather uncommon shapes, the last 4 layers in the RESNET101 core needed to be updated additionally to the classification head.

B.4.2 Training Mask R-CNN

In order to train MASK R-CNN training configurations and a specific training strategy had to be developed. Both of these are typically altered in order to optimize results achieved during the evaluation of the model. The most important training configurations that had to be set in MASK R-CNN for training were:

1. **Steps per epoch.** This determined how many gradient updates (updates of weights) were done before a new set of weights was saved at the end of an epoch.

2. **Images per GPU.** For a $1024x1024$ pixel image, the maximum batch size was two, using a GPU with 11-12 GB memory. Assessing more than two images per training step, required more than 12 GB of memory.

3. **Validation steps.** These ran at the end of an epoch to generate validation statistics. These default to 50 and should generally be much smaller than steps per epoch to not slow down the training process, due to additional memory requirements.

4. **Learning rate.** The smaller the number of the learning rate, the smaller the number of gradient updates. A default of 0.001 to train the heads was

used; a smaller value was more appropriate for training last layers of the RESNET101 backbone.

5. **Number of classes.** This indicates the number of classes to train for and was set to ”tree, no-tree”.

Appendix C

Comparing existing and generated tree genera distribution information

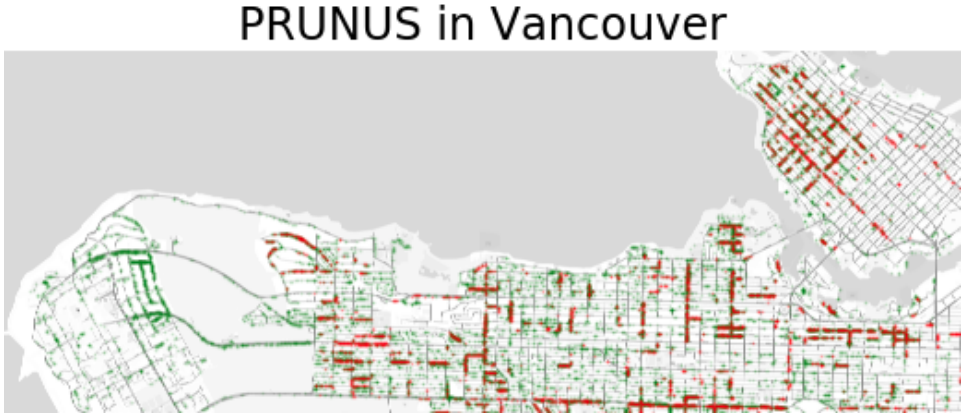


Figure C.1: Visual comparison of existing and generated tree inventory records for *Prunus*. Generated *Prunus* occurrence data (green) is more spread as existing records of *Prunus* (red) in Vancouver, owing to the inclusion of private trees and oversampling inherent to the methodology used. (Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.)

Appendix D

Tree genera detection

Table D.1: Tree genera detections in Metro Vancouver

genus	count	—	genus	count
FAGUS	609288		SORBUS	52896
ACER	492853		TILIA	49200
PRUNUS	360943		ULMUS	46560
PICEA	263775		LIQUIDAMBAR	28151
MALUS	240133		STYRAX	27796
CHAMAECYPARIS	232685		AESCULUS	26863
BETULA	214790		ILEX	25902
THUJA	174806		MAGNOLIA	20259
PARROTIA	140427		AMELANCHIER	18929
CRATAEGUS	122600		GLEDITSIA	14956
PINUS	120400		PLATANUS	10277
FRAXINUS	110167		LABURNUM	7708
QUERCUS	94854		ROBINIA	5226
PSEUDOTSUGA	90482		CERCIS	3675
CARPINUS	88762		ABIES	2727
CORNUS	79627		TSUGA	2498
CERCIDIPHYLLUM	74687		CATALPA	2182
CEDRUS	59897		METASEQUOIA	1323
LIRIODENDRON	55320		GINKGO	1176
PYRUS	54857		JUGLANS	564
			OTHER	41614