A TRANSFER LEARNING APPROACH FOR AUTOMATIC MAPPING OF RETROGRESSIVE THAW SLUMPS (RTSS) IN THE WESTERN CANADIAN ARCTIC

YIWEN LIN

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Degree of Master of Science in Geography

Department of Geography, Environment and Geomatics
Faculty of Arts
University of Ottawa

Abstract

Retrogressive thaw slumps (RTSs) are thermokarst landforms that develop on slopes in permafrost regions when thawing permafrost causes the land surface to collapse. RTSs are an indicator of climate change and pose a threat to human infrastructure and ecosystems in the affected areas. As the availability of ready-to-use high-resolution satellite imagery increases, automatic RTS mapping is being explored with deep learning methods. We employed a pre-trained Mask-RCNN model to automatically map RTSs on Banks Island and Victoria Island in the western Canadian Arctic, where there is extensive RTS activity. We tested the model with different settings, including image band combinations, backbones, and backbone trainable layers, and performed hyper-parameter tuning and determined the optimal learning rate, momentum, and decay rate for each of the model settings. Our final model successfully mapped most of the RTSs in our test sites, with F1 scores ranging from 0.61 to 0.79. Our study demonstrates that transfer learning from a pre-trained Mask-RCNN model is an effective approach that has the potential to be applied for RTS mapping across the Canadian Arctic.

Acknowledgements

First and foremost, this endeavor would not have been possible without my supervisor, Dr. Anders Knudby, who is knowledgeable, friendly and helpful. He encouraged me to think independently for my thesis project and assisted me with a complete research plan. After that, he continued to support my research and guided me regarding any obstacles I encountered through extensive meetings and discussions with me. Apart from that, he sets an example of good time management for me, which has led me to develop the ability to multitask better.

I would like to extend my sincere thanks to the rest of my thesis committee, Dr. Michael Sawada and Dr. Antoni Lewkowicz. Michael is considered as the most proficient at deep learning in the department. Although busy, he passed on the best he knows about deep learning to me and sent me lots of materials for reference. His insightful comments and hard questions truly motivate me to progress. Antoni is an expert in thaw slumps study, and therefore he shared his knowledge about retrogressive thaw slumps with me which helped me build a solid foundation for conducting the research.

In addition to the people mentioned above who have given me a great amount of academic support, I am also grateful for researcher Lingcao Huang at the University of Colorado Boulder. Inspired by his work, I started this research. He has always been willing to help and actively respond to questions while I was conducting the research. My thanks also go to all my lab mates for sharing their experience and academic opinions, which have been very useful.

I would like to express my deepest gratitude to my parents who have been supporting me both materially and spiritually to finish my graduate study in Canada. Their love and understanding have kept me upbeat and positive whenever things have been difficult over the past two years.

Lastly, I'd like to acknowledge my roommates and friends for their continued support and encouragement. Among them, I am especially thankful to my friend Pu Sun for his attentive listening when I was expressing my worries and his sincere advice when I sought help.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Changing Climate and Thermokarst Landforms

Compared to lower latitude regions, a greater warming trend accompanied by increasing precipitation has been observed over the last two decades in the Arctic. Ever since 1980, the increase of average annual air temperature in the Arctic has been more than twice that of the global average (AMAP, 2012; IPCC, 2019). Significantly, there was a 3.1°C mean annual Arctic air temperature increase from 1971 to 2019 (AMAP, 2021). A modest increase in precipitation over the Arctic for the past seven decades has also been recorded (AMAP, 2012). According to observation and modelling, the total annual precipitation including rainfall and snowfall in the Arctic has increased at least 9% for the period of 1971 to 2019 (AMAP, 2021).

The changing climate has driven drastic changes in the Arctic permafrost environments (AMAP, 2012; Arenson et al., 2015). Permafrost is the frozen ground that remains below 0°C for at least two consecutive years (Van Everdingen, 1998), and it underlies around 24% of the land surface in the Northern Hemisphere and nearly 50% of land mass in Canada (Couture & Cruden, 2010). Permafrost terrain is highly susceptible to climate change (Grosse et al., 2011). Particularly, the Canadian Arctic Archipelago has been shown to be one of the most climate-sensitive permafrost regions in the world (Fraser et al., 2018). The increasing air temperature over the Arctic has propelled the rise of ground temperature over terrestrial permafrost. Between 2000 and 2018, the annually average temperature over the Canadian Arctic region has risen roughly 0.7°C to 0.9° C (Overland et al., 2019). In higher latitudes and colder parts of the Canadian Arctic, climate observation data has shown an up to 4°C ground temperature increase for the period of 1990 to 2016 (Romanovsky et al., 2019). The rapid climate change over the Canadian Arctic region brings about the acceleration of permafrost thaw, thereby driving the formation of thermokarst, which is one of the most problematic mass-wasting permafrost landforms.

1.2 RTS Background

Among the diverse forms of thermokarst, retrogressive thaw slumps (RTS) are the most active and widespread in the western Canadian Arctic (Burn & Lewkowicz, 1990). RTSs are dynamic

permafrost features comprising an upslope ice-rich headwall, a downslope muddy slump floor, and a debris tongue behind the floor. Melting of ground ice caused by permafrost thaw results in the occurrence of RTSs, and the rise of air temperature speeds up the process. Precipitation, terrain factors and ground ice conditions all influence RTS development. RTSs are generally found along rivers, lakes, coastlines, and on slopes (Northwest Territories Geological Survey, 2010), and can stay active for three to five decades (French, 2018). The headwall of an RTS can reach a height up to 25 m (Northwest Territories Geological Survey, 2010) and typically slopes at 20° to 80° while the slope of the slump floor usually ranges from 1° to 10° (Burn & Lewkowicz, 1990).

RTSs are highly dynamic as their development has a cyclical nature. RTSs are initiated when ground ice is exposed as a result of thermal disturbance or erosion of ice-rich permafrost (Burn & Lewkowicz, 1990). As exposed ground ice ablates during the thawing season, typically early July to late August, RTSs start to develop by retrogression, in which soil thaws and falls from the headwall to the floor, thus exposing new frozen soil as the headwall retreats away from the slump floor (Burn & Lewkowicz, 1990; French, 2018; Lacelle et al., 2010; Lewkowicz & Way, 2019). When temperatures drop below 0 °C in the fall and ice ceases to melt, RTS development is suspended. Then, next summer, the re-ablation of ground ice triggers RTS development again (Lewkowicz & Way, 2019).

RTSs enter a stable phase when the ground ice is exhausted, or when the surface that the headwall retreats into is no longer higher than that of the slump floor (Burn & Lewkowicz, 1990). During stabilization, sedimentary material on the slump floor gradually dries out and vegetation begins to re-colonize (Lantz et al., 2009). Therefore, RTSs can be categorised as either active or stable. Usually, active RTSs have a well-defined headwall and are covered by newly exposed sediments as well as poorly developed vegetation. On the contrary, stable RTSs are characterized by their vegetated slump floors. Since the mudflow of RTSs can cover ground ice and thus preserve it, a single regression of RTS may result in incomplete permafrost thaw. If the preserved ice is subsequently exposed, a stable RTS may re-activate or a new RTS may occur on the slump floor of the stable one. Consequently, RTS activity can influence an area repeatedly (Lewkowicz & Way, 2019).

RTSs have caused serious problems in the Canadian Arctic region, impacting overlying infrastructure, soil and vegetation, sedimentation process, water quality and thus the entire aquatic ecosystems, and carbon exchange. RTS-induced uneven surface subsidence has caused the embankment collapse of adjacent roads. For example, at Dempster Highway, a relatively small RTS influenced its stability, which required additional gravel fill (van der Sluijs et al., 2018). It is also possible for coastal RTSs to trigger erosion and flood hazards near infrastructure (Radosavljevic et al., 2016). RTS activity can give rise to the formation of ion-rich slump soils by releasing formerly sequestered materials from permafrost (Kokelj et al., 2005). The altered soil conditions exert an adverse effect on vegetation, as the sediment on the slump floor, having high Na+ concentration, is toxic to plants and therefore unfavorable for vegetation growth (Kokelj et al., 2002). Besides, the changed soil environment poses a threat to water bodies by modifying water quality. In summer, owing to the melt of relict ground ice, active RTSs deliver hundreds of thousands of cubic meters of sediments full of soluble materials into adjacent rivers, lakes or coastal zones via surface runoff. The discharged materials raise the concentrations of soluble ions, change the pH level, conductivity and total sedimentation rates, lower dissolved organic carbon (DOC) content, and affect water color and turbidity in recipient water bodies, indicating a substantial change of water quality (Eickmeyer, 2013; Kokelj et al., 2005; Northwest Territories Geological Survey, 2010). For example, nearly 300 lakes across Banks Island have undergone substantial changes in water color that are visually detectable on satellite imagery (Lewkowicz & Way, 2019). RTSs may also disturb aquatic food webs via altering the amount and quality of biomass and habitat structures of aquatic invertebrate communities (Mesquita, 2008), causing adverse consequences on aquatic ecosystems. The modified sedimentation process has also negatively impacted fluvial sedimentary systems, which resulted in the transfiguration of the summer hydrological regime (Rudy et al., 2017). For nearby rivers, the enlargement of slump floor and debris tongue has caused the change of river base-level and strengthened the valley-side erosion (Lacelle et al., 2015). RTSs are also involved in a positive feedback loop related to carbon emissions. Carbon stored in the permafrost is liberated by RTS activity, increasing carbon dioxide and methane emissions that in turn intensify global warming. RTS-affected lakes may also experience a change in net carbon flux, leading to increased emissions (Eickmeyer, 2013).

Over the past few decades, RTS activity has intensified both in terms of the number of active RTSs, their size, and their retreat rate. A study conducted across several regions in the western Canadian Arctic (the Jesse Moraine, the Tuktoyaktuk Coastlands, the Bluenose Moraine and the Peel Plateau) shows that, over the past 60 years, the average RTS size has increased for all these regions, and their growth rate is accelerating (Segal et al., 2016). On Banks Island, more than 4500 RTSs have been initiated since 1984, with a 60-fold growth in the number of active RTSs, and the total area of RTSs on the island has increased by nearly 60000 m^2 during the period 1984–2015 (Lewkowicz & Way, 2019). In other studies, an increase in RTS retreat rate to a maximum of 15 m/y has been reported (Lacelle et al., 2015), and a widespread increase in the number and size of RTS since the mid-1980s have been observed as well (Kokelj et al., 2015; Rudy et al., 2017). These trends are not limited to Canadian Arctic; a study on the northern Qinghai-Tibetan plateau documents that the number of RTSs increased by 80.0% from 1969 to 2017, and the total active RTS impacted area increased by 69.4% from 1997 to 2017. The areal growth rate of RTSs during 2015–2017 is 156.7 m^2/y, which is higher than 62.9 m^2/y in 1997–2009 and 60.0 m^2/y in 2009–2015, indicating that the development of RTSs is rapidly growing in the alpine permafrost regions as well (Mu et al., 2020). RTS longevity is also likely to increase as global warming strengthens (Lewkowicz & Way, 2019). The increase of RTS activity leads to intensification of its environmental impact on a regional or even global scale as the carbon cycle is being affected, but the dynamic nature and widespread distribution of RTSs make their ecological impacts on Arctic environments difficult to investigate. Consequently, it is important to better understand RTS distribution and development across the Canadian Arctic regions, as well as the relationship between RTS development and impacts on the surrounding ecosystems for assistance in environmental protection and infrastructure planning.

1.3 RTS Mapping

As RTSs are widely scattered over relatively inaccessible high latitude cold regions, and can cover areas ranging from less than 1500 m^2 to more than 520000 m^2, it is a challenge to map and monitor RTSs at a regional scale. Remote sensing, as a large-scale earth surface monitoring tool that is easy to access at a relatively low cost, can be useful for this task. Its capacity to capture landscape images over large, remote and cold Arctic areas without the need for human presence

in the field makes it potentially very cost-effective for mapping RTSs. In remote sensing, images with different resolutions are available. Mid-resolution satellite imagery, such as Landsat imagery, can provide regional coverage that has been used for large spatial scale RTS mapping (Nitze et al., 2018). Nonetheless, its spatial resolution of 30 m (15m pan-chromatic) means that many smaller RTSs cannot be mapped. Imagery at higher spatial resolution is better suited to detect smaller RTSs, and to map RTS shape and size, but may be costly and more difficult to acquire over large spatial extents. However, both types of satellite imagery are limited to the period of time for which they are available. Landsat imagery at 30 m spatial resolution has been available since 1984, with relatively consistent (annual) acquisitions over much of the Arctic. Higher-resolution imagery (4m resolution and better) has been available since 1999 but is acquired on an ad hoc basis when a customer pays, meaning that imagery is rarely consistently available. Including aerial remote sensing, RTSs can be mapped on historic aerial photographs dating back to the mid-20th century, or using high-resolution unmanned aerial vehicles (UAV), neither of which can produce regional image coverage. Therefore, lots of studies have coupled mid-resolution satellite imagery with high resolution satellite imagery, aerial photographs and/or UAV imagery for mapping RTSs (Lantuit & Pollard, 2008; Luo et al., 2019; Mu et al., 2020; Rudy et al., 2017; Segal et al., 2016; Ward Jones et al., 2019). This approach can combine the advantages of different kinds of images, but the procedures for image collection, co-registration, and processing to combine these disparate data sources are relatively complicated.

After acquiring remote sensing images, the traditional and the most commonly adopted RTS mapping approach is manual digitization (Lantuit & Pollard, 2008; Luo et al., 2019; Mu et al., 2020; Rudy et al., 2017; Segal et al., 2016; Ward Jones et al., 2019). Manual digitization is considered as an accurate and easy-to-implement approach over small areas, but given the complex geometries, scattered distribution, varying sizes, and different types of RTSs, it is also laborious and time-consuming. Besides, it may yield omission errors as a result of fatigue, boredom and inattention from the operators when the study area becomes very large and covers hundreds or thousands of RTSs (Healy et al., 2004). Moreover, highly dynamic RTSs vary from year to year, requiring a fast and efficient approach with limited need for expensive human labor to regularly monitor RTS development. Therefore, alternative methodologies for mapping RTSs have been developed. Brooker et al. (2014) calculated the Tasselled Cap transformations of

brightness, greenness and wetness based on a dense stack of Landsat imagery and proved it effective not only in mapping active RTSs but also differentiating between active and stable slumps. Additionally, based on the fact that the slump floor of a RTS is characteristically lower in elevation than the surrounding landscape due to ground subsidence of thaw, Bernhard et al. (2020) mapped RTSs by detecting the significant elevation change of RTSs on a Digital Elevation Model (DEM) derived from Single-Pass TanDEM-X Observations using a simple threshold method.

Other emerging techniques have also been used to map RTSs. The increasingly popular machine learning has proven its outstanding performance in computer vision tasks, inspiring the integration of machine learning and remote sensing for various land cover mapping applications, including RTS mapping. A study has built and applied a Random Forest (RF) model to Eastern Canada, Alaska, eastern Siberia and western Siberia for classification of lakes, RTSs and wildfires (Nitze et al., 2018), while others have employed both RF and Support Vector Machine (SVM) methods for RTS mapping in Northern Canada (Bernhard et al., 2020). In addition, recent advances in deep learning, a branch of machine learning, have facilitated landscape mapping on remote sensing imagery via multiple novel convolutional neural networks (CNN). Huang et al. (2018, 2020, 2022) utilized DeepLabv3+, one of the existing CNN architectures to map RTSs in the Tibetan Plateau and the Canadian Arctic. Another RTS mapping study (Nitze et al. (2021)) not only leveraged DeepLabv3+, but also explored two other state-of-art CNN architectures, UNet and UNet++. The above-mentioned architectures are all types of semantic segmentation. However, the other segmentation type, instance segmentation, has not been tested yet in the task of RTS mapping. As instance segmentation is able to identify every instance under a class, which is different from semantic segmentation that exclusively focuses on class-to-class identification, it has the potential to discriminate adjacent objects without merging them into one segment, as well as produce accurate outlines of such very close objects. This coincides with the goals of RTS mapping, which are to precisely calculate the number of individual RTS that might be very close to each other within an area and to precisely delineate the boundaries of the RTSs to allow monitoring changes over time. Therefore, this study explores the applicability of instance segmentation to RTS mapping.

1.4 Thesis Objective

The objectives of this study are: (1) to train an instance segmentation model (Mask R-CNN) to automatically map individual RTSs in satellite imagery, (2) to apply the trained model to several areas with different RTS distribution and landscapes in the western Canadian Arctic region, and (3) to evalute the model performance.

1.5 Thesis Format

This thesis presents the paper “A transfer learning approach for automatic mapping of Retrogressive Thaw Slumps (RTSs) in the western Canadian Arctic”, which is intended for publication in a peer-reviewed journal, as Chapter 2. The conclusion is presented in Chapter 3, and the references for each chapter are presented all together in Chapter 4.

Chapter 2: A transfer learning approach for automatic mapping of Retrogressive Thaw Slumps (RTSs) in the western Canadian Arctic

2.1 Introduction

The Canadian Arctic is a glacially conditioned ice-rich permafrost environment, intrinsically vulnerable to a changing climate (Fraser et al., 2018; Lewkowicz & Way, 2019). Compared to lower latitude regions, a greater warming trend accompanied by increasing precipitation has been observed over the last two decades in the Arctic (AMAP, 2012; IPCC, 2019). Rapid climate change over the Canadian Arctic region is accelerating permafrost thaw, driving the formation of thermokarst, which is one of the most problematic mass-wasting permafrost landforms.

Among the diverse forms of thermokarst, retrogressive thaw slumps (RTS) are the most active and widespread in the western Canadian Arctic (Burn & Lewkowicz, 1990). RTSs are dynamic permafrost features comprising an upslope ice-rich headwall, a downslope muddy slump floor, and a debris tongue behind the floor. Melting of ground ice caused by permafrost thaw results in the occurrence of RTSs and increases in air temperature speeds up the process. Precipitation, terrain factors and ground ice conditions all influence RTS development. RTSs are generally found along rivers, lakes, coastlines, and on slopes (Northwest Territories Geological Survey, 2010), and can stay active for three to five decades (French, 2018). When an RTS stabilizes, sedimentary material on the slump floor gradually dries out and vegetation begins to re-colonize (Lantz et al., 2009). Based on this change, RTSs can be categorised as either active or stable: active RTSs have a well-defined headwall and are covered by newly exposed sediments as well as poorly developed vegetation, while stable RTSs are characterized by their vegetated slump floors.

RTS activity influences local ecosystems as well as infrastructure, destabilizing highways and pipelines, affecting soil conditions, accelerating sediment transportation to the downstream environments (Kokelj et al., 2015; Rudy et al., 2017), changing water composition and thereby impacting aquatic ecosystems (Kokelj et al., 2005), as well as enhancing ground-air carbon

exchange (Eickmeyer, 2013). In particular, RTSs pose a threat to a large number of Canadian Arctic lakes. These lakes are experiencing significant modification in water color and quality as a result of the release of sediment by RTS activity (Lewkowicz & Way, 2019), which indicates a potential disruption of ecological integrity.

Due to global warming, RTSs are undergoing a substantial increase in abundance, size and retreat rate (Kokelj et al., 2015; Lacelle et al., 2015; Lewkowicz & Way, 2019; Rudy et al., 2017; Segal et al., 2016). The increase of RTS activity leads to intensification of its environmental impact on a regional or even global scale as the carbon cycle is being affected, but the dynamic nature and widespread distribution of RTSs make their combined ecological impacts on Arctic environments difficult to investigate. Consequently, it is important to understand RTS development and distribution across the Canadian Arctic region, as well as the relationship between RTS development and impacts on the surrounding ecosystems. Better understanding of the spatial distribution and temporal variation of RTS activity will contribute to a more comprehensive understanding of the impacts of climate change on Arctic environments and people and facilitate infrastructure planning in the affected areas. However, it is challenging to map and monitor RTSs at a regional scale because they are widely scattered over relatively inaccessible terrain, and because they vary greatly in size, ranging from less than 1500 m^2 to more than 520000 m^2 (Günther et al., 2015; Kokelj et al., 2015; Lacelle et al., 2015; Ramage et al., 2017; Runge et al., 2022). Earth observation technology, with its ability to produce relevant data covering large areas, repeatedly and at relatively low cost, is therefore ideally suited to facilitate our understanding of RTS distribution and development across the areas where they occur.

Many satellite images have spatial resolutions that allow identification of most RTSs. With the large number of satellite launches and the growth in data volume, satellite images are becoming more accessible than ever. With these images, researchers can conduct manual digitization for land cover mapping, which has been commonly adopted in the field of RTS research (Lantuit & Pollard, 2008; Luo et al., 2019; Mu et al., 2020; Rudy et al., 2017; Segal et al., 2016; Ward Jones et al., 2019). This approach, although easy to implement over small areas, becomes laborious, time-consuming and may yield omission errors as a result of fatigue, boredom and inattention

from the operators when the study area becomes very large and covers hundreds or thousands of RTSs (Healy et al., 2004). Moreover, highly dynamic RTSs vary from year to year, requiring a fast and efficient approach with limited need for expensive human labor to regularly monitor RTS development.

As machine learning has gained increasing popularity because of its outstanding performance in the field of computer vision, including classification, segmentation and non-image applications, its use for satellite image interpretation is being explored and extensively applied in various land cover mapping tasks (de Lima & Marfurt, 2020; Ghosh et al., 2018; Hu et al., 2020; Ma et al., 2019), including for RTS mapping. A Random Forest (RF) model has been built and applied to Eastern Canada, Alaska, eastern Siberia and western Siberia for classification of lakes, RTSs and wildfires (Nitze et al., 2018). Likewise, a RF and a Support Vector Machine (SVM) were employed for distinguishing RTSs from other land cover in Northern Canada (Bernhard et al., 2020). In addition, recent advances in deep learning, a branch of machine learning, have greatly facilitated land cover mapping using Convolutional Neural Networks (CNNs). As one of the most promising image processing tools, a CNN is adept at handling image data and extracting information in a timely and efficient manner. Some studies (Huang et al., 2020) leveraged DeepLabv3+, a CNN architecture, for mapping RTSs in Tibet and the Canadian Arctic. Nitze et al. (2021) also explored other state-of-art CNN architectures (UNet, UNet++), in addition to DeepLabv3+, for RTS mapping. They applied their approach across a range of Arctic permafrost regions and tried to construct a model with high transferability at a larger scale. The architectures used in the above-mentioned studies are all based on the concept of semantic segmentation, which generates image segments by assigning a class to each pixel of an image. While semantic segmentation exclusively focuses on class-to-class identification, instance segmentation, on the other hand, enables identifying every individual under a class (i.e., pixels can be identified not just as “RTS”, but as “RTS 1” or “RTS 2”). The difference between these two segmentation methods suggests that semantic segmentation models might fail to discriminate objects very close to each other (and instead merge them into one object) or to generate and track accurate outlines of such very close objects. Being able to identify individual RTSs with accurate outlines can be helpful because it allows users to precisely count the number of RTSs within a specific area, and investigate the change of RTS boundaries over time, thus allowing better estimation of

RTS development. While no instance segmentation model has been tested for RTS mapping, we expect that such models generally have the potential to perform well for RTS mapping, especially in areas with RTSs that are close to each other. Therefore, the objectives of this study are: (1) to train an instance segmentation model (Mask R-CNN) to automatically map RTSs in satellite imagery, (2) to apply the trained model to several areas with different RTS distribution and landscapes in the western Canadian Arctic region, and (3) to evalute the model performance.

2.2 Study Area

Our study spans two islands in the Canadian Arctic Archipelago – Banks Island and Victoria Island. Banks Island, the fourth biggest island in Canada, lies at the westernmost end of the Canadian Arctic Archipelago, it is topographically undulating and composed of poorly lithified sand and shale bedrock with Quaternary sediments at the bottom (French, 2017). As a typical cold region with long cold winters and short cool summers, the climate record from Sachs Harbour, located at the south-western tip of Banks Island (71°57′N, 124° 44′W), shows an average annual air temperature of -12.8 °C, and a summer average of 5.2 °C from 1981 to 2010. The average annual precipitation is 151.5 mm, the majority of which falls as snow, and around 37% of which is rainfall during summer (Environment and Climate Change Canada, 2022b). Victoria Island is the second largest island in Canada, sitting in the western Canadian Arctic, and is composed of mostly sedimentary rocks (Marsh, 2008). Most of Victoria Island is stratified by dolomite, with some limestone, sandstone and shale of Early Paleozoic origin (Saarela et al., 2020). Climate data from Cambridge Bay, located on the south coast of Victoria Island (69°09′N, 105°08′W), show that its annual average temperature during 1981-2010 was -13.9 °C, with a summer average of 7.9 °C. The average annual precipitation in Cambridge Bay is 141.7 mm, about 47% of which falls as rain, mostly during the months of July and August (Environment and Climate Change Canada, 2022a).

Banks Island and Victoria Island are both located within the continuous permafrost zone with medium-to-high ground ice content, and therefore have long been known as regions with widespread thermokarst landscape features, such as ice-wedge polygons, thermokarst lakes and ponds, and RTSs (French, 2017; Kokelj et al., 2020; Rudy et al., 2017). RTSs, as one of the most active and significant geographic features on both islands, are primarily located along the south

and east coast of Banks Island and the northwestern part of Victoria Island (Kokelj et al., 2017). Recently, intensified RTS activity on both islands has been reported as the regional temperature has increased (Kokelj et al., 2017; Segal et al., 2016). On Banks Island specifically, RTSs have increased more than 60-fold in number during 1984–2013, and at least 20-fold in volume during 1984-2015 (Lewkowicz & Way, 2019).

We selected three sites from eastern Banks Island and one site from western Victoria Island as our study area, as shown in Figure 1. These sites serve as representatives of eastern and southern Banks Island and northwestern Victoria Island. While extensive RTS activity is occurring at each of these sites, they vary with respect to landscapes and RTS distribution:

(1) S1 is located on the eastern side of Banks Island, where the land transitions from beach to mud, sand and gravel cliffs. It is situated between the latitudes of 72.93 °N and 73.16 °N and longitudes of 117.88 °W and 118.72 °W, covering an area of around 685 km^2. The vegetation of this site is mainly prostrate dwarf-shrub, herb tundra (Walker et al., 2005). There is a large number of water bodies in this area. Most RTSs at this site are distributed along lakes and rivers, with a few scattered over the hillslopes. There are no coastal RTSs in this area due to its inland location.
(2) S2 is situated on eastern Banks Island between the latitudes of 72.79 °N and 72.86 °N and longitudes of 118.82 °W and 119.11 °W, covering an area of around 80 km^2. Its landscapes and RTS distribution are similar to S1.
(3) S3 is situated on the Prince Albert Peninsula, the northwest region of Victoria Island, with latitude ranging from 72.46 °N to 72.56 °N and longitude ranging from 117.86 °W to 118.17 °W. This site covers 31 km^2. The RTS distribution in this site is similar to S1 while the vegetation is slightly different, primarily covered by sedge and prostrate dwarf-shrub (Walker et al., 2005).
(4) S4 is located in the southern part of Banks Island. Its latitude ranges from 71.39 °N to 71.45 °N and its longitude ranges from 122.09 °W to 122.21 °W. The total area of this site is around 112 km^2. This site is mostly hilly terrain and dominantly vegetated by sedge and dwarf-shrub (Walker et al., 2005). It contains only RTSs that are close to rivers and slopes.

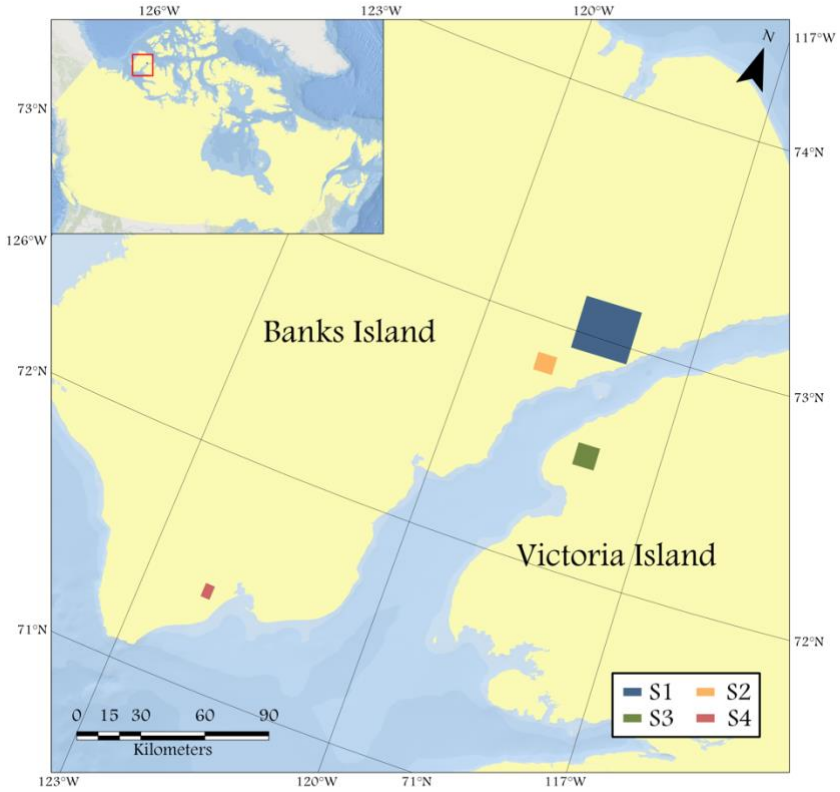

Figure 1. Study Area.

The number of different types of RTS distribution and the percentage of RTS area per site are shown in Table 1.

Table 1. The number of different RTS types and RTS area percentage of each site.

Site	Lake RTSs	River RTSs	Slope RTSs	RTS area percentage
S1	135	78	29	12.13%
S2	14	36	2	12.52%
S3	25	13	5	6.14%
S4	0	21	7	20.81%

The selection of the size of these sites was determined by the quality of the available satellite imagery, the density of RTSs in different areas and the number of RTSs that can be manually digitized in a limited time.

2.3 Methods

The methods used in this study include satellite imagery collection and preprocessing, RTS and non-RTS data collection, model preparation and training, followed by prediction, post-processing and evaluation. The flowchart of the whole procedure is shown in Figure 2.

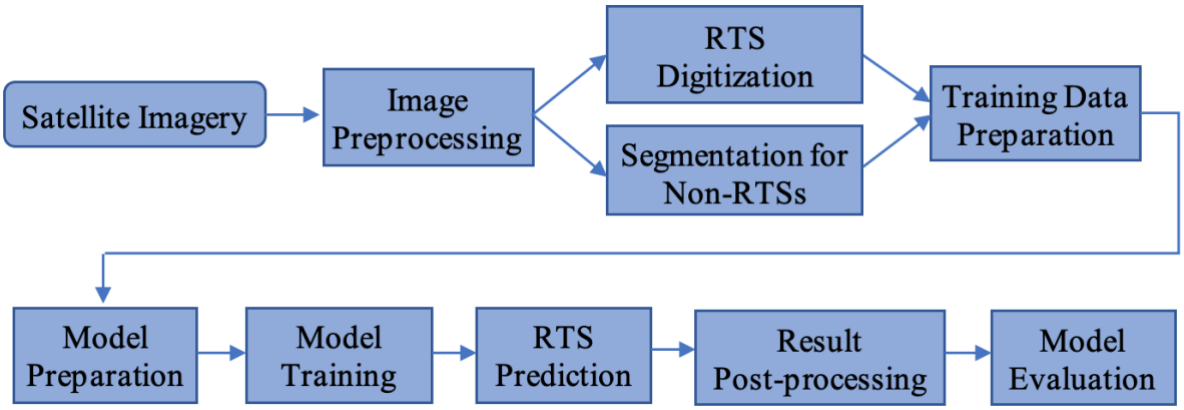

Figure 2. Flowchart of the automatic RTS mapping method based on Mask R-CNN.

2.3.1 Image Collection and Preprocessing

We obtained our image data from PlanetScope satellite imagery. Deployed in a multi-satellite constellation, PlanetScope satellites are capable of acquiring global land surface images of both high spatial and temporal resolution. Currently, there are over 180 PlanetScope satellites operating in sun-synchronous orbit, collecting imagery in four bands – blue, green, red and near infrared (NIR) – with a ground resolution of approximately 3 meters. As a result of the high temporal resolution of PlanetScope imagery – often multiple images are available daily – cloud-free imagery is typically available across our study area on a weekly or at least monthly basis. At the same time, its high spatial resolution allows for precise digitization of landforms, enabling us to confidently map small RTSs that would have been omitted by using satellite imagery from other sources.

Through its Education and Research Program, Planet provides limited free access to daily PlanetScope imagery (Planet Team, 2017). This imagery is provided as image tiles of 10 km x 30 km, orthorectified and calibrated to surface reflectance with 16-bit depth. All sites have daily imagery, and most of the site is covered by multiple images daily. For this study, we selected only images from 2017, which is the first year PlanetScope imagery is available. Imagery from the four sites was acquired as follows: (1) A shapefile of the site was uploaded to Planet Explorer to define the spatial scope and browse the available imagery, limiting the search to imagery from the months of July and August, during which Banks Island and Victoria Island are mostly free of surface ice and snow. (2) From the many images available through Planet Explorer, for each site a single day was identified that provides 100% image coverage and the least amount of cloud-cover. This date may be different from site to site. (3) The individual tiles from one day can have partial cloud cover and may overlap. From all available tiles, the smallest set of tiles that in combination provides complete cloud-free coverage of each site was therefore determined, and the necessary tiles downloaded. The imagery dates of each site are shown in Table 2.

Table 2. Imagery dates of each site.

Site	Date
S1	2017-08-18
S2	2017-08-18
S3	2017-08-18
S4	2017-07-19

Next, the downloaded tiles from each site were mosaicked into one single image. For images that have large water bodies, we applied a threshold of 0.1 surface reflectance in the NIR band and kept only values greater than 0.1 to mask out water bodies. After that, according to the input requirements of the CNN model to be used, the 16-bit image with 4 bands was converted into an 8-bit, 3-band image by rescaling the pixel DN values from the range of 0-4000 to 0-255 and then extracting the desired three bands. Here, we considered three combinations of bands for creation of the 3-band image: (1) red, green, blue (RGB), (2) near-infrared, red, green (NRG), and (3) the first three principal components derived from a Principal Component Analysis (PCA). We

implemented PCA transformation on the image in ArcGIS Pro 2.8.0 with the Principal Components tool and selected the first three principal components as the final 3 bands. We included the RGB combination as it is the most commonly chosen band combination in RTS mapping (Huang et al., 2020; Luo et al., 2019; Mu et al., 2020; Rudy et al., 2017; Segal et al., 2016). There are two reasons for also choosing NRG combination as a candidate: (1) the NIR band is effective at contrasting the low-vegetation condition of active RTSs with the surrounding vegetated terrain, and (2) these bands suffer the least from atmospheric scattering and thus tend to have greater pixel-to-pixel contrast than the blue band. For PCA, we selected it because of its ability to optimize the amount of information present within three bands. Each of these three band combinations were used as the basis for model training (see section 2.3.5).

2.3.2 Data Collection

We created training, validation and test data by digitizing RTS and non-RTS polygons from the PlanetScope imagery. RTSs can be divided into two categories: active and stable. We included only active RTSs in our data, due to their influence on the surrounding environment and the risk they pose to infrastructure. The digitization of RTSs was conducted only for RTSs larger than 900 m^2 as visual identification of smaller RTSs could not be done confidently with the available imagery. The digitization process was implemented in QGIS 3.10 using digitization tools. We tested different band combinations to use for digitization and found visual identification of RTSs was easiest from the NRG image. The procedure was as follows: (1) RTSs were visually identified on the image according to the presence of distinct headwalls and the color contrast between slump floors and surrounding terrain. On the NRG image, the active RTSs look brighter and less red than stable ones, as shown in Figure 3. (2) RTS polygons were drawn to outline active RTSs, including the headwalls and their extended debris tongues. RTSs that had initially developed at separate locations, but appeared merged in the imagery from 2017, were digitized as a single RTS.

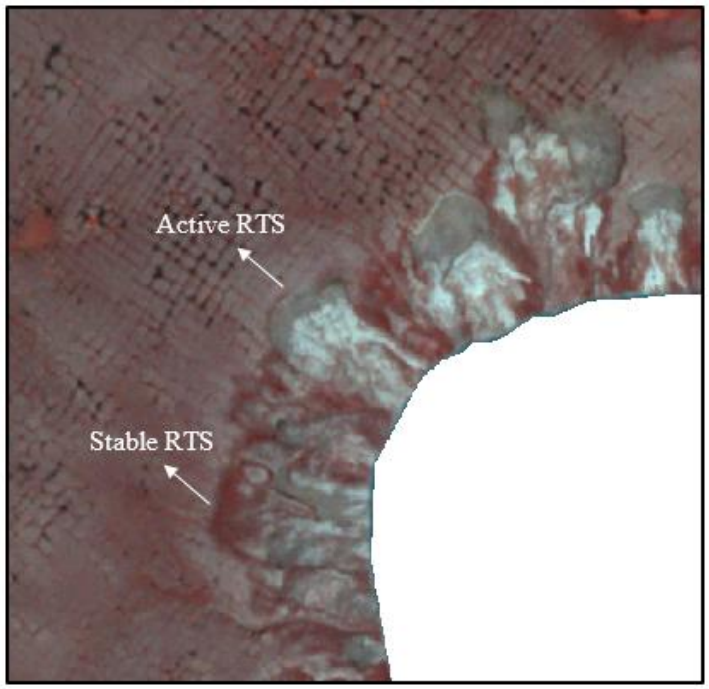

Figure 3. Examples of active RTSs and stable RTSs. The active RTSs appear to be brighter while the stable RTSs appear to be darker and more red. The white part in the image (bottom right) is a lake that has been masked out.

To assist with RTS identification, the RTS point data gathered by Lewkowicz & Way (2019) before 2016 were used by overlaying it on the images to quickly locate possible positions of RTSs. This dataset has attributes including start and end dates of RTS activity, from which stable RTSs can be identified and excluded. We also inspected the stable RTSs in our study area to ensure none of them reactivated after 2016. However, the existing RTS point positions were collected before 2016, whereas PlanetScope satellite only provides images starting from 2017, leaving the possibility that RTSs active in 2016 had stabilized by 2017. This dataset was therefore only used to speed up location of potential active RTSs, and the active/passive status of each RTS was visually verified in the imagery.

There are several challenges to digitization: (1) due to the limited image resolution and complicated landforms, it is often difficult to precisely locate where an RTS started to develop and delineate their boundaries, and (2) the RTSs that occur along slopes (henceforth: “slope”

RTSs) are hard to identify owing to the influence of similar surrounding landforms. Two additional strategies were used to solve these problems:

(1) Inspecting year-to-year change of RTSs on Google Earth Timelapse.
This method has been used for RTS change detection and proven to be advantageous in terms of determining actual RTS activity instead of inferring activity from headwall appearance or lack of vegetation, which is relatively less precise (Lewkowicz & Way, 2019). Google Earth Timelapse is a global, zoomable video based on a series of mostly annual cloud-free satellite image mosaics that enable users to see the change of the entire Earth surface dated back to 1985. Timelapse is based on imagery from Landsat 4, 5, 7 and 8, and Sentinel-2A, resulting in a 30 m or smaller spatial resolution. By viewing the fast-changing image sequence, the initial locations of RTSs can be visually located and some of the less obvious "slope" RTSs can more easily be identified. Moreover, we can further view later years, such as 2018 and 2019, which helps us confirm whether the RTSs identified in 2017 were still active. Nevertheless, due to the limited resolution, this is only feasible for larger, detectable RTSs in Google Earth Timelapse. This strategy is also beneficial for excluding other stable landforms with RTS-like appearances.

(2) Using PlanetScope images from other years as reference on Planet Explorer.
Checking the RTSs on PlanetScope images from 2018 is supplementary to using Google Earth Timelapse. We employed Planet Explorer to conduct this task. Planet Explorer not only provides daily PlanetScope images but also offers high-quality mosaicked imagery at weekly, monthly and quarterly intervals. If the boundary of an RTS is not clear in the individual downloaded image, we checked the same RTS in the 2018 monthly mosaic of the same month to see if it shows a clearer boundary. This method may also contribute to the identification of "slope" RTSs if they present a clear headwall shadow in images from other years. In addition, as a complement to Google Earth Timelapse, the 2018 monthly mosaics were also used to determine the active status of the RTSs in 2017 that could be detected in Google Earth Timelapse.

We obtained 242, 52, 43 and 28 RTS polygons for S1, S2, S3, S4, respectively, from digitization. The size of the obtained RTSs ranges from 939 m^2 to 1381102 m^2. Non-RTS polygons are also vital for RTS model training (Huang et al., 2020; Nitze et al., 2021), as it helps the model

distinguish between RTSs and other similar landforms. We conducted object-based image analysis segmentation on training images to generate non-RTS polygons using the Segmentation function under the Classification tools in ArcGIS Pro 2.8.0. After segmentation, we manually selected segments that are similar to RTSs in appearance (but are not actually RTSs) as non-RTS polygons. With this method, we created 683 non-RTS polygons for S1.

2.3.3 Training Data Preparation

In this study we treated S1, the largest site on Banks Island, as the training site, and we treated the rest – S2, S3 and S4 – as test sites. The imagery of the training site was used to train the Mask R-CNN model, while the imagery of the testing sites was used for testing model generalizability.

The satellite image of S1 (8935 x 8522 pixels) is large compared to traditional CNN input images, and this significantly increases the computational burden of model training. Therefore, for training data preparation, we used a grid to cut the image of S1 into smaller images patches, each one 256 x 256 pixels in size, and then overlayed the RTS data layer on each patch, as shown in Figure 4. Only patches that intersected an RTS or non-RTS polygon were retained for training. We label patches that contains an RTS polygon RTS patches, whether or not it also contains any non-RTS polygons. The rest are labeled non-RTS patches. RTS polygons that cover more than one patch were split in this process; we label these RTS pieces.

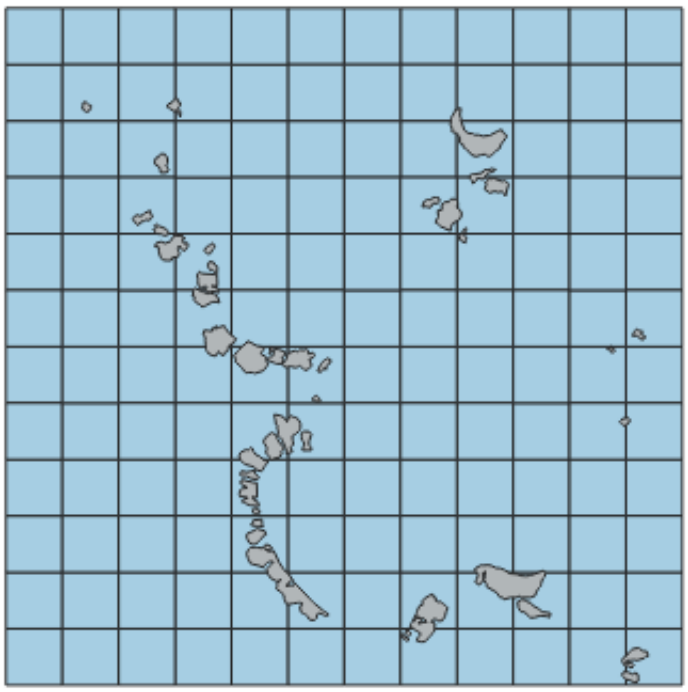

Figure 4. Concept diagram showing a subset of the grid (blue squares) used to create image patches from the training image, with RTS polygons overlaid (grey).

Next, we randomly split the RTS patches and non-RTS patches into training and validation sets with a ratio of 9:1. The splitting was performed only once, so the training and validation sets were kept unchanged for all the following experiments. We then applied image augmentation techniques to increase the size of the training set. Only RTS patches in the training set underwent augmentation in an effort to balance the number of RTS and non-RTS patches. The augmentation techniques we applied from library imgaug 0.4.0 (https://github.com/aleju/imgaug) are as follows : (1) horizontal flipping, (2) vertical flipping with a probability of 0.5 (i.e. each patch had a 0.5 probability of being vertically flipped), (3) random cropping by 0% to 10% of image height/width or a random padding by 0% to 5% of image height/width, (4) a combination of rotating and scaling with a probability of 0.5, (5) one or two changes from adding noise, blurring, changing contrast and changing brightness. Eventually, we obtained 366 RTS and 360 non-RTS patches for training (total 726 training patches), and 21 RTS and 40 non-RTS patches for validation (total 61 validation patches). The number of different RTS types of training and validation sets are shown in Table 3.

Table 3. The number of different RTS types of training and validation sets.

Dataset	Lake RTSs	River RTSs	Slope RTSs
Training sets	124	70	27
Validation sets	21	13	2

For CNN model training, we also need the labels that indicate the locations of RTSs. For each patch, as required by instance segmentation, each polygon in the label should be assigned a unique ID. We first removed any RTSs pieces smaller than 900 m^2, and then gave an instance ID (1, 2, 3…) to the RTS polygons/pieces. We then rasterized the RTS polygons/pieces according to their instance ID(s) and the extent of the corresponding image as the RTS labels (Figure 5). The non-RTS labels were rasterized with "0" values everywhere, indicating no RTS present.

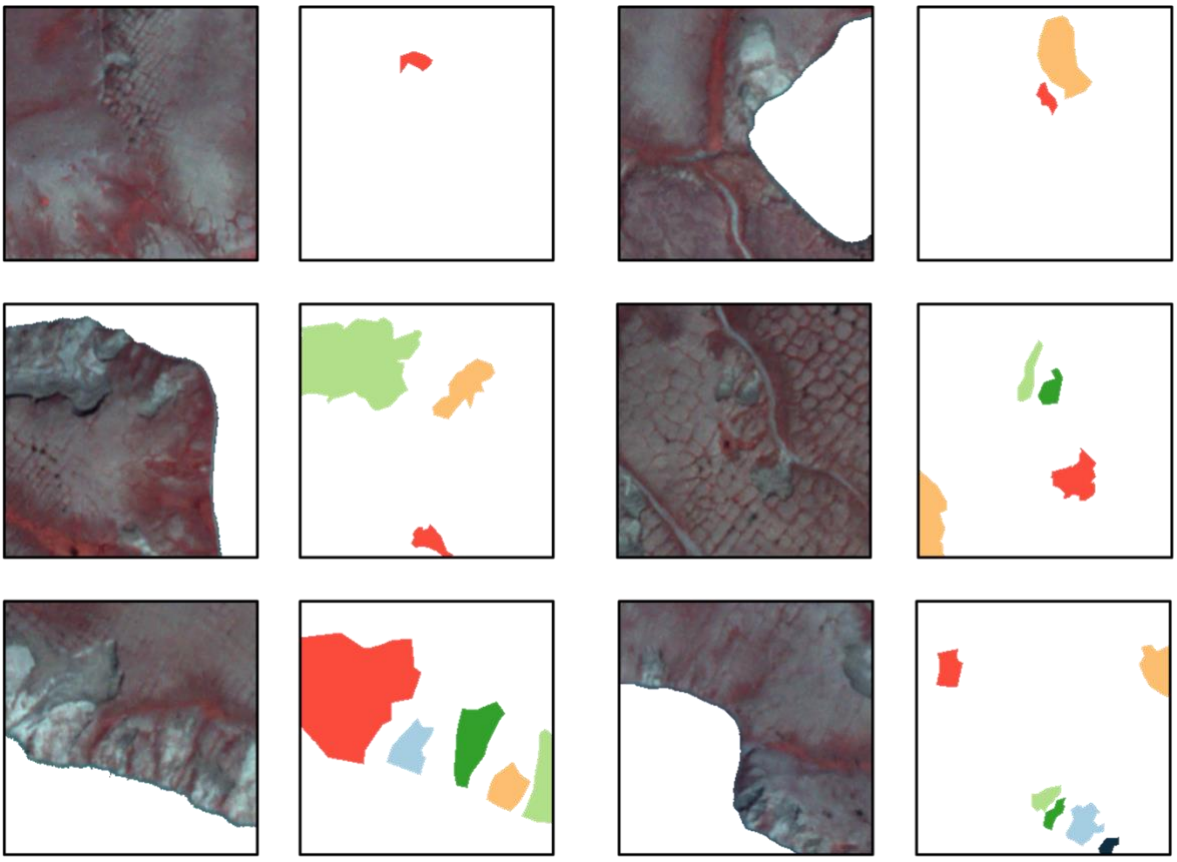

Figure 5. RTS patches (the first and third columns) and their corresponding RTS labels (the second and last columns). The white part in RTS patches are parts of lakes that have been

masked out. The different colors in labels represent different RTS instances with different IDs, and the white part in labels is pixels of value “0”.

2.3.4 Model Preparation

2.3.4.1 Mask R-CNN

In this study, we utilized Mask R-CNN for RTS mapping. Mask R-CNN is a state-of-the-art model designed for instance segmentation. As an extention of Faster R-CNN that is used for object detection, Mask R-CNN performs segmentation on images through two stages. In the first stage, images are fed into a backbone network for feature extraction and subsequently a region proposal network (RPN) that generates region proposals of objects. The second stage utilizes the information from region proposals as well as the extracted feature maps in the first stage to produce the bounding boxes of objects. After that, a fully connected network is concatenated for predicting the class and refining the bounding boxes, while on the other hand, several convolutional operations are performed to generate the masks for all the image objects. A detailed descripition of Mask R-CNN is provided by He at al. (2017).

2.3.4.2 Pre-trained backbones

Using a CNN model that has been trained to solve a problem in one domain for other tasks in the same or different domains is called transfer learning. This is a robust and efficient approach that helps alleviate training data deficiency that may lead to overfitting, and can potentially increase the model accuracy (Barman et al., 2019). Given the small amount of training data available, we employed two pre-trained backbones for our model, ResNet-50 and ResNet-101, both of which are commonly used backbones for image segmentation tasks (Feng, 2017; He et al., 2016). The main difference between these two backbones is the different number of neural network layers; ResNet-50 has 50 layers and ResNet-101 has 101 layers. In the original development of Mask R-CNN, incorporating a Feature Pyramid Network (FPN), which extracts image features at different scales by exploiting the multi-scale characteristic of the deep convolutional networks (Li et al., 2017), with ResNet-50 and ResNet-101 (henceforth ResNet-50-FPN and ResNet-101-FPN) as the model backbone, achieved a notable increase in both accuracy and speed (He et al., 2017). Therefore, we chose ResNet-50-FPN and ResNet-101-FPN, both pre-trained on the

ImageNet dataset, as the backbones for our model. Hereafter, the model with pre-trained backbone is called the pre-trained model.

The optimal number of convolutional layers that are set to be trainable can vary for different pre-trained backbones and different tasks. ResNet variants contain several layers of repetitive convolution blocks, each one comprised of three convolutional layers. The architecture of ResNet is shown in Figure 6. For the Conv3 layer, ResNet-50 has 6 convolution blocks while ResNet-101 has 23 convolution blocks. The higher layers, such as Conv5, contain more abstract information, whereas the lower layers, such as Conv1, extract specific features. Thus, to determine the optimal trainable layers of the backbones, starting with the last layer (Conv5), we unfroze that layer to make it trainable, ran the model and recorded the accuracy, then moved to the second-last layer (Conv4) and so on, until there was no obvious improvement in model performance. This optimization was done by setting a batch size of 2 and a learning rate of 0.001 for 40 epochs for each of the pre-trained backbones.

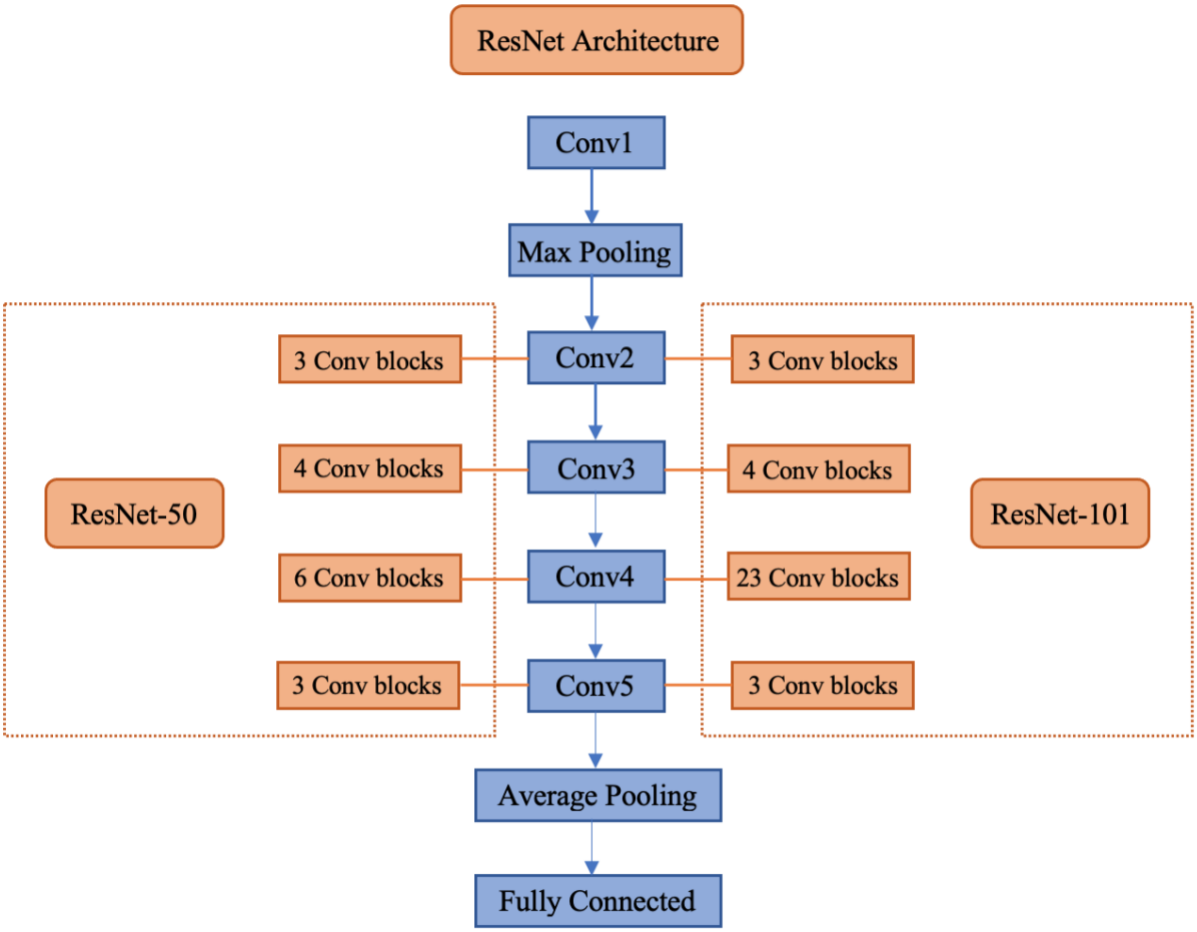

Figure 6. ResNet Architecture.

2.3.5 Model Training

The batch size, learning rate, momentum and weight decay of each model setting were fine-tuned using Ray 1.12.1 (Liaw et al., 2018). The fine-tuned hyperparameter settings are shown in Table 4. We chose these four hyper-parameters considering that they are often some of the most important parameters to be tested and the amount of time it may take to test all the hyper-parameters. We fine-tuned the batch sizes of different model settings from 1, 2 and 4 due to the large network architecture of Mask R-CNN and limited GPU memory. Learning rates were selected from 0.0001, 0.0005, 0.001, 0.005, 0.01, and 0.05, while momentum and weight decay were selected from 0.7, 0.8, 0.9, and 0.99, and 0.00001, 0.0001, and 0.001, respectively.

Table 4. Fine-tuned hyperparameters of each model setting.

Band Combination	Backbone	Batch size	Learning rate	Momentum	Weight decay
RGB	ResNet-50-FPN	2	0.005	0.7	1e-03
	ResNet-101-FPN	2	0.005	0.7	1e-05
PCA	ResNet-50-FPN	2	0.005	0.7	1e-03
	ResNet-101-FPN	2	0.005	0.7	1e-04
NRG	ResNet-50-FPN	2	0.005	0.8	1e-04
	ResNet-101-FPN	2	0.005	0.8	1e-04

The model was implemetned using Pytorch, an open source deep learning framework. We trained the model for a minimum of 15 and a maximum of 35 epochs on one NVIDIA Tesla V100 GPU with 16GB memory. For the first 1000 iterations, a learning rate warm-up (slowly increasing the learning rate until reaching the defined learning rate) with a 0.001 factor was used to prevent the model from being suddenly exposed to RTS data that is different from most RTS data. We also utilized an early stopping strategy to prevent overfitting, which terminates the training when model performance does not improve on the validation images for five consecutive epochs. It took 25-35 minutes for each model to converge. We used the cross-entropy loss function during model training.

2.3.6 Prediction and post-processing

After we obtained the best trained model, we applied it to the three test sites for RTS prediction. We followed the procedure in section 2.3.3 for training data preparation to process the images of the test sites and generate the corresponding labels. The output of prediction are the masks of detected RTSs along with confidence scores - the predicted probabilities that a mask contains an RTS. Each mask consists of pixels, each of which comes with the probability that it is part of an RTS. As mentioned before, the RTSs that stretch across one or more patches were split up into RTS pieces in the image gridding process. In order to recover the original shape of RTSs and therefore acquire the complete RTS objects, we modified the preprocessing of the test images as well as the post-processed segmentation masks: First, we added a 64-pixel overlap between the

patches of each test image when applying the gridding. After the model outputs the predicted masks of the test images with overlap, we filtered out the masks that have a confidence score lower than 0.5, which we have tested and proven to be a optimal value. Then, we thresholded the probabilties of image pixels in each mask and reallocated a value to them: Pixels with a probability greater than 0.5 were reassigned as 1 (the "RTS" class), while the others were reassigned as 0 (the "background" class). Lastly, all the masks were mosaiced together and polygonized into a shapefile that contains the predicted RTS polygons. For some RTSs the model generated more than one mask, in which case a "union" process was used to retain the RTS extent included in at least one mask. Finally, we removed all RTS polygons smaller than 900 m^2.

2.3.7 Evaluation

We adopted the following metrics for model evaluation: precision, recall, F1 score and average precision (AP). We used AP to compare different model settings and implement hyperparameter fine-tuning and early stopping, and calculated precision, recall and F1 score for the evaluation of the post-processed predictions. Precision, recall and AP metrics are based on the Intersection over Union (IoU), defined as

$$IoU\ (A, B) = area\ (A \cap B) / area\ (A \cup B) \quad (1)$$

where A refers to a predicted RTS polygon and B refers to a digitized RTS polygon. IoU is a practical metric for assessing how the shape of an object is captured by the model, with larger IoU indicating better ability of to capture the shape. Each polygon is then classified into a true positive (TP) or a false positive (FP) based on whether its IoU is greater or smaller than a predefined threshold, and an unpredicted polygon whose IoU is smaller than the threshold is considered as a false negative (FN). In this study, this threshold is set to the commonly used value of 0.5 (Everingham et al., 2010; Lin et al., 2014). After acquiring TP, FN and FP, we calculated precision, recall and F1 score following equations (2), (3) and (4), respectively. Precision reflects how robust the model is at not misidentifying non-RTSs as RTSs, whereas recall is a measure of the degree to which the model correctly detects existing RTSs. The F1

score is an overall measurement that takes both precision and recall into account. The higher the value of the F1 score, the better the performance of the model.

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$\text{F1 score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \quad (4)$$

AP is the average of the precision corresponding to each unique recall level. AP is calculated using this process: (1) sort the predicted RTS polygons by their confidence scores from largest to smallest, (2) categorize each polygon into TP or FP, (3) calculate the cumulative TP and FP in score order and subsequently the corresponding precision and recall, (4) define a set of recall levels that are of same intervals and interpolate the precision at each level, (5) take the mean of all precision values as the AP. In this study, we computed AP following the 101-point interpolation method for calculating AP_{50} (subscript 50 stands for IoU = 0.5) in standard COCO (Common Objects in Context) metrics (Dollar & Lin, 2014).

2.4 Results

2.4.1 Backbone trainable layers

The influence of the number of trainable layers on AP, in both ResNet-101-FPN and ResNet-50-FPN, is shown in Figure 7. For ResNet-101-FPN, AP increases when Conv5 and then Conv4 are unfrozen, but not when additional layers are, so the trainable layers for ResNet-101-FPN are set to Conv5 and Conv4. For ResNet-50, unfreezing Conv3 also increases AP slightly, so the trainable layers for ResNet-50-FPN are set to Conv5, Conv4, and Conv3. For different band combinations, we obtained the same results and hence used the same trainable layer settings.

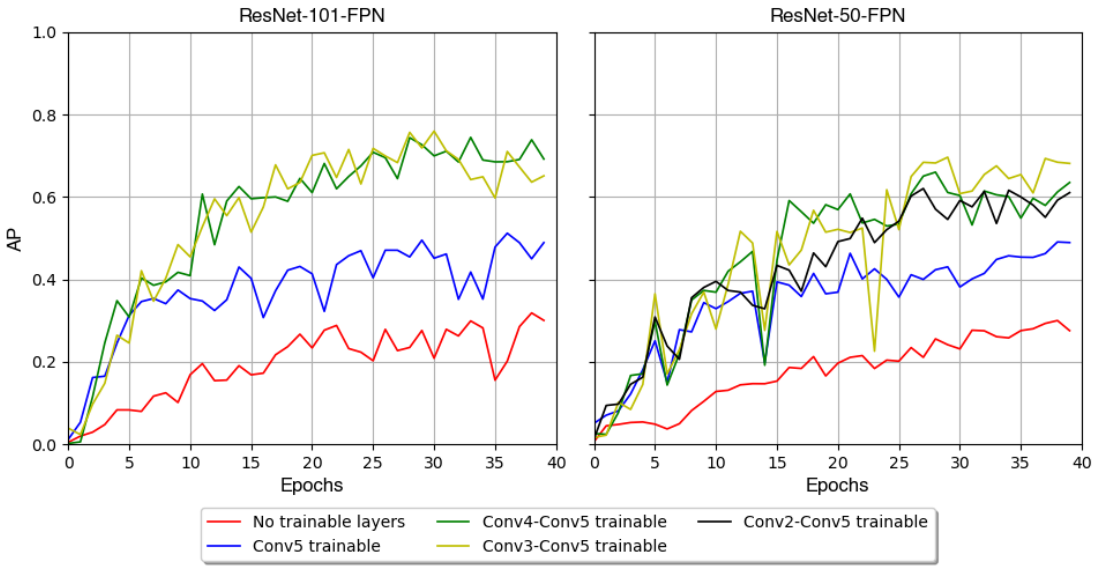

Figure 7. Tests of different trainable layers of ResNet-101-FPN and ResNet-50-FPN.

2.4.2 Model Comparison

We tested different model settings (band combinations and backbones) five times and compared their average APs on the validation data set; the results of each model setting are displayed in Table 5. Among all the band combinations, the NRG band combination shows the best AP (model setting 5 and model setting 6), with PCA in second place (model 3 and model 4) and RGB imagery performing the worst (model 1 and model 2). The AP obtained with RGB imagery is much lower than that of NRG and PCA, while the difference between NRG and PCA is smaller. When comparing the two backbones, we find that models with ResNet-101-FPN perform better than the models with ResNet-50-FPN irrespective of band combinations. Among all the settings, model setting 6 achieved the highest AP. The loss curve of the model that achieved the highest AP from model setting 6 is presented in Figure A1.

Table 5. Validation results of different model settings.

Model Setting	Band Combination	Backbone	AP
1	RGB	ResNet-50-FPN	0.44 ± 0.03
2		ResNet-101-FPN	0.53 ± 0.02
3	PCA	ResNet-50-FPN	0.72 ± 0.01
4		ResNet-101-FPN	0.73 ± 0.02
5	NRG	ResNet-50-FPN	0.74 ± 0.01
6		ResNet-101-FPN	0.77 ± 0.01

2.4.3 Prediction

For model prediction, we applied the model with the best validation AP from model setting 6 to all the test sites to test the model generalizability. The results are shown in Table 6.

Table 6. Model test results for different sites (at IoU of 0.5).

Test site	TP	FP	FN	Precision	Recall	F1 score
S2	39	8	13	0.83	0.75	0.79
S3	37	42	6	0.47	0.86	0.61
S4	16	5	12	0.76	0.57	0.65

At an IoU level of 0.5, among all the test sites, the model achieved the best precision, recall and F1 score on S2, which has the most similar landscapes to S1 (the training site). For S3, the number of false positives was high, leading to low precision, whereas recall remains high. For S4, the precision is high, and the recall is relatively low. These results indicate that the model has a balanced performance on S2, being able to find a large portion of RTSs while in most cases not misidentifying other landforms. Nevertheless, the model tends to overestimate the RTSs in S3 and underestimate the RTSs in S4. Overall, the model generalizes well to sites other than S1.

We also present the IoU histogram of the digitized RTSs of each site, as shown in Figure A2, Figure A3 and Figure A4. We can infer from the figure that, the IoU values of digitized RTSs

against the predicted RTSs on S2 and S4 mostly exceeded 0.5. For S3, except for the RTSs showing very low IoU values (below 0.1), the IoU values of the other RTSs are also concentrated between 0.5 and 1.0.

Some predicted RTSs from S2 are shown in Figure 8. Here most of the RTSs have been mapped by the model, and the shape of the mapped RTSs generally matches well with the digitized RTSs. However, some small RTSs were not found by the model (Figure 8a). Also, the model can distinguish some RTSs that are close to each other (Figure 8b). For some RTSs, their headwalls were mapped well but there is some disagreement of the slump floor and debris tongue (Figure 8b), which also leads the model to treat a merged RTS as two separate RTSs (Figure 8c). We found one false positive that covers the area of a stable RTS (Figure 8d). It is also worth noting that the model found an RTS that was initially missed during digitization (Figure 8e); we did not add such finds to the test data to recalculate the metrics, leading the numbers in Table 2 to slightly underestimate the model's actual recall values.

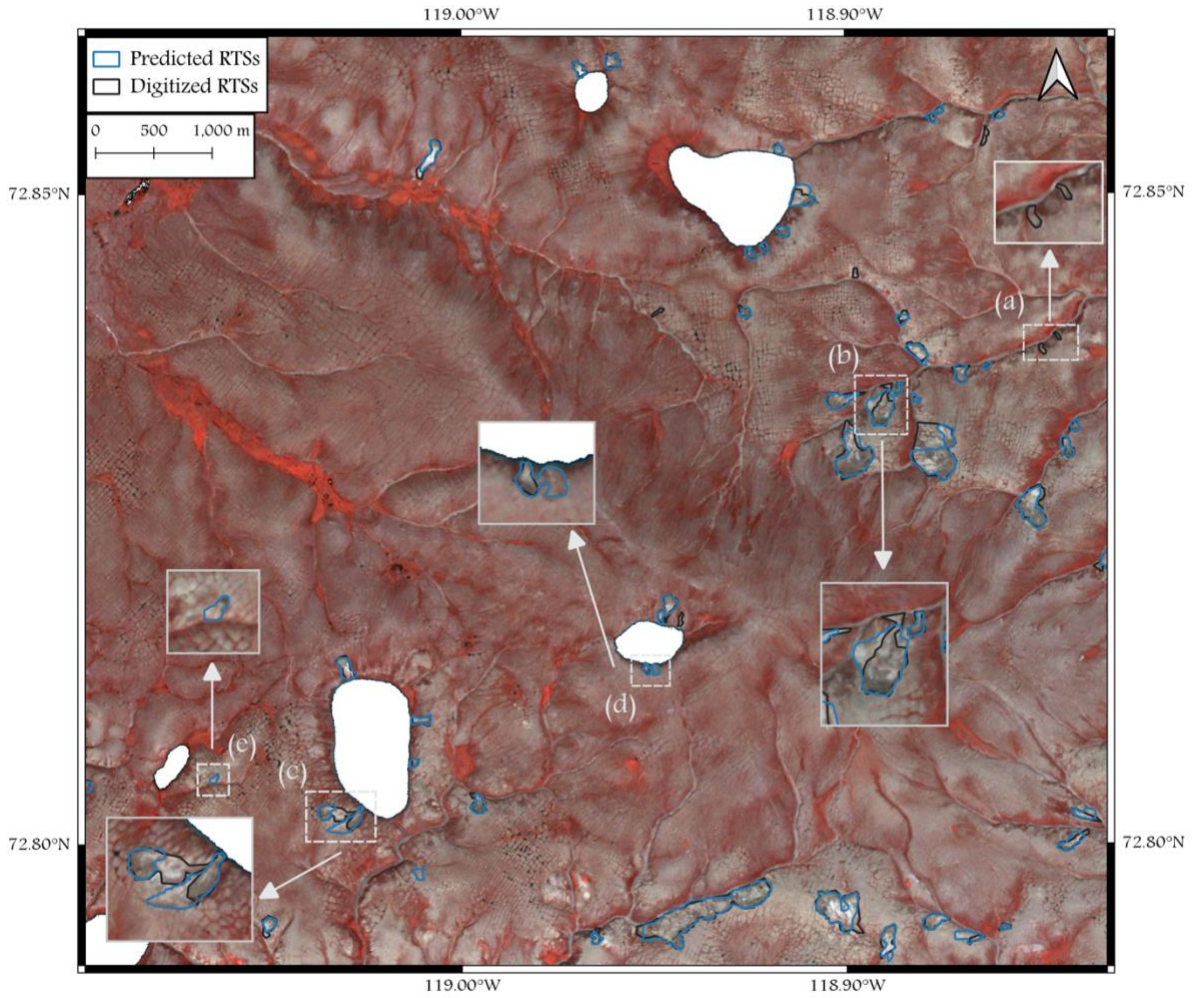

Figure 8. Predicted RTS polygons at S2. The white parts in the image are the water bodies that have been masked out.

Figure 9 presents some of the predictions from S3. The predicted RTSs are mostly aligned with the shapes of the digitized RTSs. However, on the right-hand side, where there are some bright landscape elements that look like (but are not) RTSs, the model generates several false positives (Figure 9a). On the other hand, Figure 10 shows a part of the predictions from S4, where no similar false positives exist. Here, the model tends to only map the front part of some RTSs which is visually darker and therefore the debris tongues are missing (Figure 10a).

For model ability of predicting different RTS types, we found that the model performs well in mapping RTSs along lakes and slopes, but less well in detecting the RTSs or mapping the full shape of the RTSs near rivers.

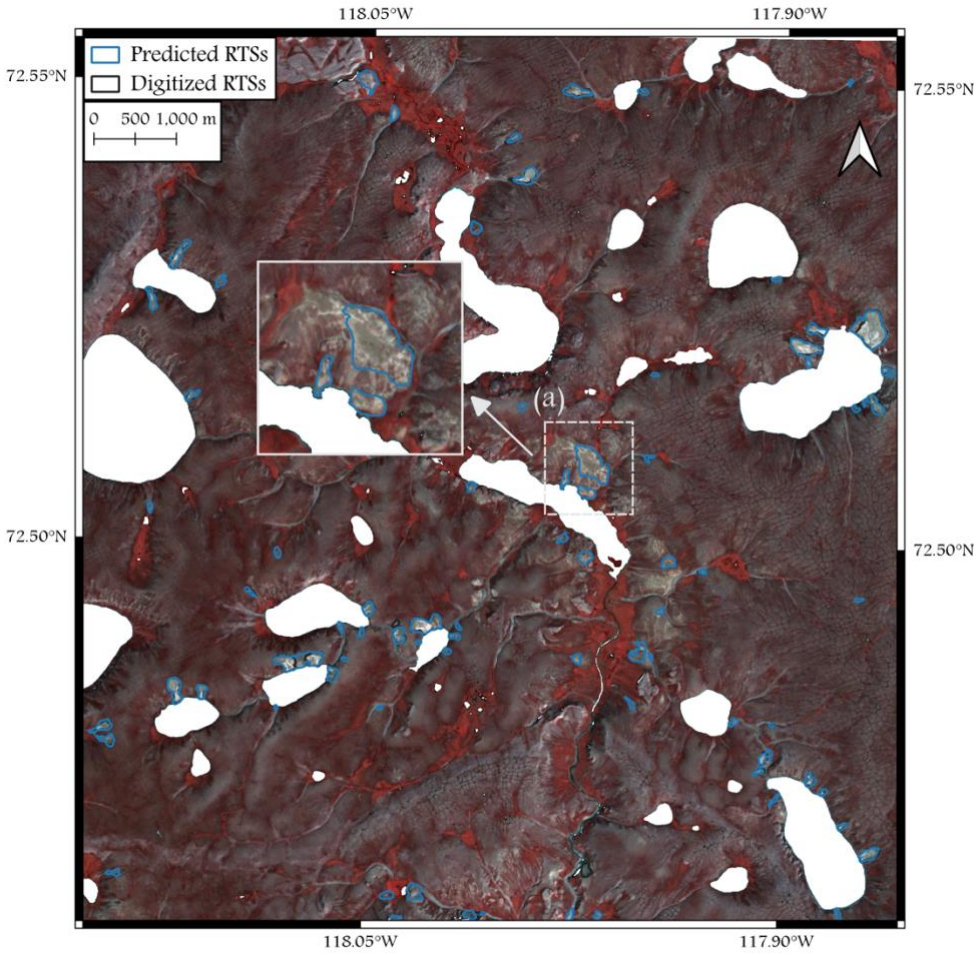

Figure 9. Predicted RTS polygons of S3. The white parts in the image are the water bodies that have been masked out.

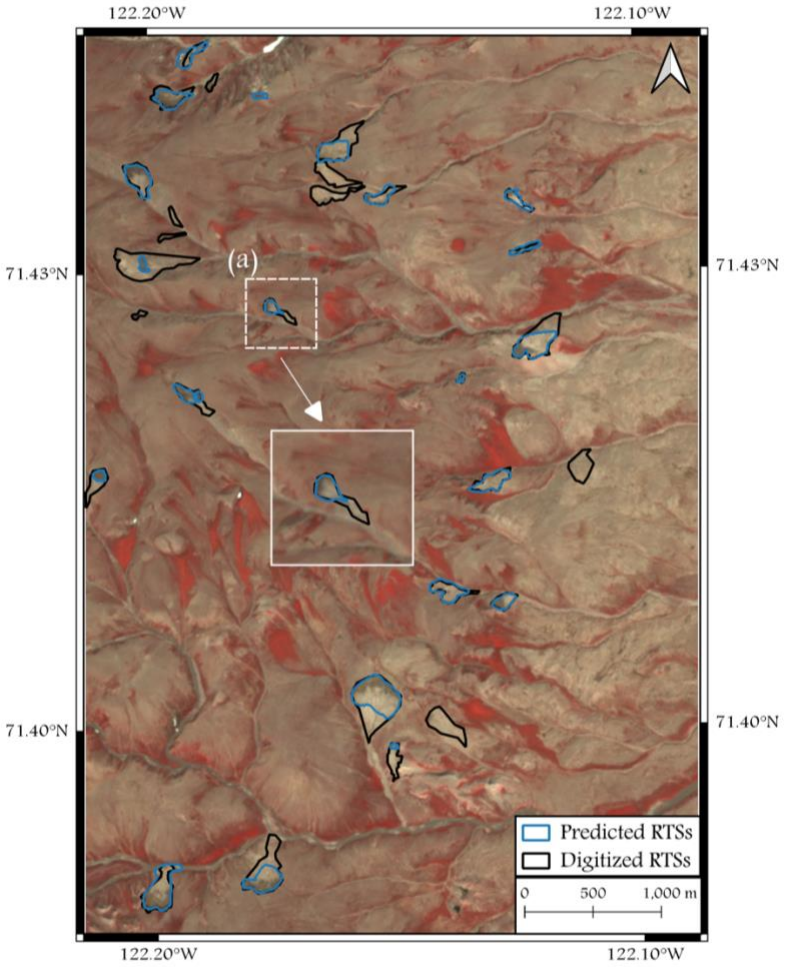

Figure 10. Predicted RTS polygons of S4.

2.5 Discussion

We tested the ability of a Mask R-CNN model to map retrogressive thaw slumps in PlanetScope satellite imagery for several areas on Banks and Victoria Islands in the western Canadian Arctic. Different from traditional CNN model training, we selected one representative site for model training, and then used other sites as the test sites for model prediction to test how well the model can generalize from one part of the Canadian Arctic to nearby areas with different landscape types.

To determine the best model settings, we tested a range of image band combinations and pre-trained backbones. The results demonstrate that the near-infrared band is crucial for

discriminating between RTSs and surrounding land covers, presumably because the NIR band provides better ability to distinguish the poorly vegetated slump floors of active RTSs from their vegetated surroundings – and from vegetated slump floors of stable RTSs. Usually, including with PlanetScope imagery, satellite images contain many bands, and for many applications using more than three bands would be beneficial. However, pre-trained backbones, including the one used in our study, are typically trained on a large amount of three-band (RGB) imagery, limiting application that rely on them to three bands. As a result, we only input three bands to the model. However, as the extra bands contain useful information that could contribute to model training, we also tested a band combination consisting of the first three principal components derived from PCA. Nevertheless, our results show that this approach was not superior to using NRG bands. Besides, generating the 3-band images using PCA requires extra work, whereas NRG bands can be easily extracted from the original 4-band images. Consequently, in situations where a test of different band combinations is undesirable, we suggest using the NRG band combination in the context of mapping RTSs using a backbone pre-trained on three-band imagery.

We also compared two different pre-trained backbones under all possible settings. Whether ResNet-101 could have a better performance than ResNet-50 or not is task dependent (Bressem et al., 2020; Nitze et al., 2021). Neural networks with more convolutional layers do not necessarily improve the overall performance if the classification task to be solved is not complicated, but it does always increase the training time. In this study, our models with ResNet-101-FPN did show higher average APs than ResNet-50-FPN for all band combinations, indicating that utilizing more convolutional layers contributes to the model performance in this case. However, this is only obvious for NRG and RGB band combination. For PCA, the difference in model performance resulting from the use of ResNet-101-FPN compared to the use of ResNet-50-FPN is less pronounced.

The final object-level mapped results are promising, with F1 scores of 0.79, 0.61 and 0.65 for test sites S2, S3 and S4, respectively, which is comparable to other RTS mapping studies utilizing deep learning approaches (Huang et al., 2018, 2020, 2022; Nitze et al., 2021). The recall of each site is 0.75 (S2), 0.86 (S3) and 0.57 (S4). This proves that our model is capable of

detecting most RTSs as well as mostly capturing the whole shape of the RTSs, not only for S2 where the landscapes is similar to S1, but also for S3 and S4 which are less similar to S1. It is also important to note that the model found a small RTS that was overlooked during the manual digitization, but was subsequently visually validated by following the approaches in section 2.3.2. This illustrates another aspect of how deep learning has the potential to contribute to RTS mapping, not only because it can be deployed over large areas quickly and cheaply, but also because it can reduce the number of omission errors from manual digitization. Nevertheless, if manual digitization is done with enough care, we argue that it will still achieve a higher mapping accuracy than deep learning methods. We obtained high precisions of 0.83 for S2 and 0.76 for S4, demonstrating that our model is also able to distinguish between RTSs and other landforms in S2 and S4. However, we obtained a precision of 0.47 for S3, where there are many bright landforms with shapes similar to RTSs. Nevertheless, envisioning that the Mask R-CNN model can be used in the context of a larger system, where a human operator can find and remove false positives, the precision less important than the recall in the case of RTS mapping.

From the visualized results, we can infer that the model sometimes does not detect the slump floors or the debris tongues of RTSs, or include stable RTSs or parts thereof, which could be due to the complexity of RTS structure and the unclear distinction between active and stable RTSs. In terms of how well the model works for different RTS distributions, we find that it is better at mapping RTSs along lakes or on slopes, whereas it is less good at detecting the small RTSs or capturing the entire shape of RTSs along rivers. We also note that the Mask R-CNN model is able to map some RTSs that are very close to each other (Figure 8b), which is a documented limitation of the model developed by Huang et al. (2020). In addition, there are some disagreements about the RTS boundaries between manual digitization and model prediction. It should be noted that our digitized RTS and non-RTS polygons have not been validated in the field, and that the different parts of an RTS are not always easily distinguishable in PlanetScope imagery. As is also illustrated by the small RTSs overlooked in manual digitization, these issues highlight the problems associated with assessing the performance of a model against test data that are digitized from imagery and therefore themselves imperfect. Nevertheless, the precision, recall and F1 score metrics, supported by visual inspection of predicted and digitized RTS

polygons overlain on imagery, support the general idea that the model performs and generalizes well to map RTSs in different geographic areas.

The low precision for S3 shows that the model has learned the main characteristics of RTSs but is not fully able to differentiate RTSs from similar-looking landforms that exist in some other areas. This also further indicates one shortcoming of our model, which is the limited generalizability. We attribute this primarily to a lack of training data. For neural network training, the amount of training data is a significant factor for model performance, and as a rule of thumb more training data is better. Even though we employed transfer learning, which reduces the need for large amounts of training data, our data set is still small (total 787 images) compared to ImageNet (over 1 million images), the original dataset for the pre-trained model. Thus, increasing the amount and diversity of training data could likely further improve model performance. This could be done by (1) adding more RTSs with greater diversity of shape, size, and landscape context, with a focus on RTSs of smaller size and/or along rivers, and (2) add more non-RTSs covering various landforms, especially those that look similar to RTSs. In the case of RTS mapping where the distribution of RTSs is usually scattered and the land on which RTSs are located is vast and composed of diverse landscapes, non-RTSs are of particular importance for improving the model precision. Moreover, according to the predictions for S2, where some stable RTSs become false positives, incorporating stable RTS data into the non-RTS data would likely also improve model performance. Thanks to the collective efforts of the RTS research community, there has been an increased amount of RTS data available to the public for neural network training in recent years (Nitze et al., 2021). In the future, as more RTS data across various Arctic landscapes are employed for CNN model training, we anticipate an RTS mapping model of better generalizability and greater potential to be applied semi-automatically for large-scale RTS mapping.

In this study, the RTSs cover 12.13%, 12.52%, 6.14% and 20.81% of the area of each site, respectively. This indicates that our RTS mapping model enables mapping relatively small, scattered RTSs over relatively large geographic areas. This is a distinct advantage in assisting large-scale RTS research, thus eliminating the need for researchers to spend much time and

effort collecting RTS data. As a result, the methodology used in this study is also applicable to other land cover mapping tasks that involve mapping small landforms from large areas.

2.6 Conclusion

Our study utilized Mask R-CNN, a deep learning model based on instance segmentation, to map RTSs in the western Canadian Arctic using PlanetScope satellite imagery. Our results demonstrated that Mask R-CNN is an effective CNN model that can be applied as an alternative to semantic segmentation models for the task of automatic RTS mapping in the Canadian Arctic. We showed that transfer learning can provide reliable results even with a small amount of training data, and can produce good model performance with appropriate pre-trained models and trainable layers. Our best Mask R-CNN model successfully mapped 39 of the 52 digitized RTS polygons at the S2 test site on eastern Banks Island, while producing 8 false positives, thereby producing a recall of 0.75 a precision of 0.83 and an F1 score of 0.79. Besides, we acquired an F1 score of 0.61 for the S3 test site on northwestern Victoria Island and an F1 score of 0.65 for the S4 test site on southern Banks Island, which demonstrated that the model is capable of correctly mapping most of the RTSs near lakes, rivers and slopes, even in areas where the landscape is relatively different from the training site. Mask R-CNN is also a promising alternative to manual digitization of RTSs, which, while more accurate, may become unfeasible in cases where large-area and/or repeat RTS mapping is required. Nevertheless, the model generated a substantial number of false positives at the S3 test site, whose composition of landforms are distinct from the training site, suggesting relatively poor generalizability of the model and thus the need to increase the variety and the number of both RTS and non-RTS training data. With additional training data, Mask R-CNN has the potential to be applied to a larger spatial scope for regional-scale RTS mapping.

Chapter 3: Conclusion

This study utilized Mask R-CNN, a deep learning model based on instance segmentation, to map RTSs in the western Canadian Arctic using PlanetScope satellite imagery. The results demonstrated that Mask R-CNN is an effective CNN model that can be applied as an alternative to semantic segmentation models for the task of automatic RTS mapping in the Canadian Arctic. In this study, different three-band combinations were used and tested for model training: (1) red, green, blue (RGB), (2) near-infrared, red, blue (NRG), and (3) the first three principal components derived from a Principal Component Analysis (PCA). The results showed that, in the case where only three bands are selected, the NRG combination is the most suitable combination for RTS mapping, demonstrating the importance of the near-infrared band that extracts useful vegetation information for differentiating between RTSs and the surrounding land covers. The study also showed that transfer learning can provide reliable results even with a small amount of training data by utilizing pre-trained models from other domains. Good model performance can be achieved by appropriately choosing and examining the pre-trained backbones and trainable layers, the optimal selection of which is task-dependent. The best Mask R-CNN model successfully mapped 39 of the 52 digitized RTS polygons at the S2 test site on eastern Banks Island where the landscape is similar to the S1 training site, while producing 8 false positives, thereby producing a recall of 0.75, a precision of 0.83, and an F1 score of 0.79. Besides, this model acquired an F1 score of 0.61 for the S3 test site on northwestern Victoria Island and an F1 score of 0.65 for the S4 test site on southern Banks Island, which demonstrated that the model is capable of mapping most of the RTSs near lakes, rivers and slopes, even in areas where the landscape is relatively different from the training site. According to the visualized results, the predicted RTS polygons mostly match the shape of digitized RTS polygons, although there are some disagreements on the part of slump floor and debris tongue, as well as some disagreements between active RTSs and stable RTSs. Mask R-CNN is also a promising alternative to manual digitization of RTSs, which, while more accurate, may become unfeasible in cases where large-area and/or repeat RTS mapping is required. Nevertheless, the model generated a substantial number of false positives at the S3 test site, whose composition of landforms are distinct from the training site, suggesting insufficient generalizability of the model. Increasing the variety and the number of both RTS and non-RTS training data is needed to

improve model generalizability. With more training data available, it could be possible for a Mask R-CNN model without pre-trained backbones to generate good results. Since RTSs have lower elevation than the surrounding ground, the inclusion of DEM data, which were not used in this study, should also be tested. In addition, satellite imagery with higher resolution (< 3m) may also be helpful for building an RTS mapping model with better performance. Furthermore, an increasing number of studies have explored the transfer learning from three-band to multi-band imagery that is more suitable to satellite imagery application. With such improvement, the Mask R-CNN model has the potential to be applied to a larger spatial scope for regional-scale RTS mapping with high accuracy and generalizability.

References

AMAP. (2012). Arctic climate issues 2011: Changes in Arctic snow, water, ice and permafrost. In *SWIPA 2011 Overview Report*. Arctic Monitoring and Assessment Programme (AMAP). http://www.amap.no/documents/doc/arctic-climate-issues-2011-changes-in-arctic-snow-water-ice-and-permafrost/129

AMAP. (2021). Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policy-makers. In *Arctic Monitoring and Assessment Programme (AMAP), Tromsø, Norway*. https://www.amap.no/documents/doc/arctic-climate-change-update-2021-key-trends-and-impacts.-summary-for-policy-makers/3508

Arenson, L. U., Colgan, W., & Marshall, H. P. (2015). *Chapter 2 - Physical, Thermal, and Mechanical Properties of Snow, Ice, and Permafrost* (J. F. Shroder, W. Haeberli, & C. B. T.-S. and I.-R. H. Whiteman Risks and Disasters (eds.); pp. 35–75). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-394849-6.00002-0

Barman, R., Deshpande, S., Agarwal, S., Inamdar, U., Devare, M., & Patil, A. (2019). Transfer Learning for Small Dataset. *Proceedings of National Conference on Machine Learning*, 132–137. https://www.researchgate.net/publication/333080572

Bernhard, P., Zwieback, S., Leinss, S., & Hajnsek, I. (2020). Mapping Retrogressive Thaw Slumps Using Single-Pass TanDEM-X Observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3263–3280. https://doi.org/10.1109/JSTARS.2020.3000648

Bressem, K. K., Adams, L. C., Erxleben, C., Hamm, B., Niehues, S. M., & Vahldiek, J. L. (2020). Comparing different deep learning architectures for classification of chest radiographs. *Nature*, *10*(1), 1–16. https://doi.org/10.1038/s41598-020-70479-z

Burn, C. R., & Lewkowicz, A. G. (1990). Canadian Landform Examples - 17 Retrogressive Thaw Slumps. *Canadian Geographer / Le Géographe Canadien*, *34*(3), 273–276. https://doi.org/10.1111/j.1541-0064.1990.tb01092.x

Couture, R., & Cruden, D. M. (2010). More comprehensive characterization of landslides in permafrost. *63rd Canadian Geotechnical Conference & 6th Canadian Permafrost Conference*, 855–861. https://members.cgs.ca/documents/conference2010/GEO2010/pdfs/GEO2010_110.pdf

de Lima, R. P., & Marfurt, K. (2020). Convolutional neural network for remote-sensing scene

classification: Transfer learning analysis. *Remote Sensing*, *12*(1), 86. https://doi.org/10.3390/rs12010086

Dollar, P., & Lin, T.-Y. (2014). *COCO API - Dataset*. https://github.com/cocodataset/cocoapi/blob/master/PythonAPI/pycocotools/cocoeval.py

Eickmeyer, D. (2013). *The Effects of Retrogressive Thaw Slump Development on Persistent Organic Pollutants in Lake Sediments of the Mackenzie River Delta* [University of Ottawa]. https://doi.org/10.20381/ruor-3179

Environment and Climate Change Canada. (2022a). *Canadian Climate Normals 1981–2010 Station Data - Cambridge Bay A.* https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?stnID=1786&prov=&lang=e&dCode=1&dispBack=1&StationName=Cambridge_Bay&SearchType=Contains&province=ALL&provBut=&month1=0&month2=12

Environment and Climate Change Canada. (2022b). *Canadian Climate Normals 1981–2010 Station Data - Sachs Harbour A.* https://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?searchType=stnProv&lstProvince=NT&txtCentralLatMin=0&txtCentralLatSec=0&txtCentralLongMin=0&txtCentralLongSec=0&stnID=1794&dispBack=01

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, *88*(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Feng, V. (2017). *An Overview of ResNet and its Variants*. Towards Data Science. https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

Fraser, R. H., Kokelj, S. V., Lantz, T. C., McFarlane-Winchester, M., Olthof, I., & Lacelle, D. (2018). Climate sensitivity of high arctic permafrost terrain demonstrated by widespread ice-wedge thermokarst on banks Island. *Remote Sensing*, *10*(6). https://doi.org/10.3390/rs10060954

French, H. M. (2017). The Banks Island Tundra. In O. Slaymaker (Ed.), *Landscapes and Landforms of Western Canada* (pp. 97–108). Springer International Publishing. https://doi.org/10.1007/978-3-319-44595-3_6

French, H. M. (2018). *The Periglacial Environment* (4th ed.). John Wiley & Sons. https://lccn.loc.gov/2017027903

Ghosh, A., Ehrlich, M., Shah, S., Davis, L., & Chellappa, R. (2018). Stacked U-nets for ground material segmentation in remote sensing imagery. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2018-Jun*, 252–256. https://doi.org/10.1109/CVPRW.2018.00047

Grosse, G., Romanovsky, V., Jorgenson, T., Anthony, K. W., Brown, J., & Overduin, P. P. (2011). Vulnerability and Feedbacks of Permafrost to Climate Change. *Eos, Transactions American Geophysical Union*, *92*(9), 73–74. https://doi.org/https://doi.org/10.1029/2011EO090001

Günther, F., Grosse, G., Wetterich, S., Jones, B. M., Kunitsky, V. V., Kienast, F., & Schirrmeister, L. (2015). The Batagay Mega Thaw Slump, Yana Uplands, Yakutia, Russia: Permafrost thaw Dynamics on Decadal Time Scale. *PAST Gateways - Palaeo-Arctic Spatial and Temporal Gateways - Third International Conference and Workshop*. https://epic.awi.de/id/eprint/38008/

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Healy, A., Kole, J., Buck-Gengler, C., & Bourne, L. (2004). Effects of Prolonged Work on Data Entry Speed and Accuracy. *Journal of Experimental Psychology: Applied*, *10*, 188–199. https://doi.org/10.1037/1076-898X.10.3.188

Hu, J., Li, L., Lin, Y., Wu, F., & Zhao, J. (2020). A Comparison and Strategy of Semantic Segmentation on Remote Sensing Images. *Advances in Intelligent Systems and Computing*, *1074*, 21–29. https://doi.org/10.1007/978-3-030-32456-8_3

Huang, L., Lantz, T. C., Fraser, R. H., Tiampo, K. F., Willis, M. J., & Schaefer, K. (2022). Accuracy, Efficiency, and Transferability of a Deep Learning Model for Mapping Retrogressive Thaw Slumps across the Canadian Arctic. *Remote Sensing*, *14*(12). https://doi.org/10.3390/rs14122747

Huang, L., Liu, L., Jiang, L., & Zhang, T. (2018). Automatic mapping of thermokarst landforms from remote sensing images using deep learning: A case study in the northeastern Tibetan

Plateau. *Remote Sensing*, *10*(12), 2067. https://doi.org/10.3390/rs10122067

Huang, L., Luo, J., Lin, Z., Niu, F., & Liu, L. (2020). Using deep learning to map retrogressive thaw slumps in the Beiluhe region (Tibetan Plateau) from CubeSat images. *Remote Sensing of Environment*, *237*(November 2019), 111534. https://doi.org/10.1016/j.rse.2019.111534

IPCC. (2019). *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. Cambridge University Press. https://doi.org/10.1017/9781009157964

Kokelj, S. V., Kokoszka, J., Van Der Sluijs, J., Rudy, A. C. A., Tunnicliffe, J., Shakil, S., Tank, S., & Zolkos, S. (2020). Permafrost thaw couples slopes with downstream systems and effects propagate through Arctic drainage networks. *The Cryosphere Discussion*, *September*, 1–43. https://doi.org/10.5194/tc-2020-218

Kokelj, S. V., Tunnicliffe, J., Lacelle, D., Lantz, T. C., Chin, K. S., & Fraser, R. (2015). Increased precipitation drives mega slump development and destabilization of ice-rich permafrost terrain, northwestern Canada. *Global and Planetary Change*, *129*, 56–68. https://doi.org/10.1016/j.gloplacha.2015.02.008

Kokelj, S. V., Jenkins, R. E., Milburn, D., Burn, C. R., & Snow, N. (2005). The influence of thermokarst disturbance on the water quality of small upland lakes, Mackenzie Delta region, Northwest Territories, Canada. *Permafrost and Periglacial Processes*, *16*(4), 343–353. https://doi.org/10.1002/ppp.536

Kokelj, S. V., Lantz, T. C., Tunnicliffe, J., Segal, R., & Lacelle, D. (2017). Climate-driven thaw of permafrost preserved glacial landscapes, northwestern Canada. *Geology*, *45*(4), 371–374. https://doi.org/10.1130/G38626.1

Kokelj, S. V., Smith, C. A. S., & Burn, C. R. (2002). Physical and chemical characteristics of the active layer and permafrost, Herschel Island, western Arctic Coast, Canada. *Permafrost and Periglacial Processes*, *13*(2), 171–185. https://doi.org/10.1002/ppp.417

Lacelle, D., Bjornson, J., & Lauriol, B. (2010). Climatic and geomorphic factors affecting contemporary (1950–2004) activity of retrogressive thaw slumps on the Aklavik Plateau, Richardson Mountains, NWT, Canada. *Permafrost and Periglacial Processes*, *21*(1), 1–15. https://doi.org/https://doi.org/10.1002/ppp.666

Lacelle, D., Brooker, A., Fraser, R. H., & Kokelj, S. V. (2015). Distribution and growth of thaw slumps in the Richardson Mountains–Peel Plateau region, northwestern Canada. *Geomorphology*, *235*, 40–51. https://doi.org/10.1016/j.geomorph.2015.01.024

Lantuit, H., & Pollard, W. H. (2008). Fifty years of coastal erosion and retrogressive thaw slump activity on Herschel Island, southern Beaufort Sea, Yukon Territory, Canada. *Geomorphology*, *95*(1–2), 84–102. https://doi.org/10.1016/j.geomorph.2006.07.040

Lantz, T. C., Kokelj, S. V., Gergel, S. E., & Henry, G. H. R. (2009). Relative impacts of disturbance and temperature: persistent changes in microenvironment and vegetation in retrogressive thaw slumps. *Global Change Biology*, *15*(7), 1664–1675. https://doi.org/10.1111/j.1365-2486.2009.01917.x

Lewkowicz, A. G., & Way, R. G. (2019). Extremes of summer climate trigger thousands of thermokarst landslides in a High Arctic environment. *Nature Communications*, *10*(1), 1329. https://doi.org/10.1038/s41467-019-09314-7

Li, X., Lai, T., Wang, S., Chen, Q., Yang, C., & Chen, R. (2017). Feature Pyramid Networks for Object Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944. https://doi.org/10.1109/CVPR.2017.106

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., & Stoica, I. (2018). Tune: A Research Platform for Distributed Model Selection and Training. *ICML 2018 AutoML Workshop*. https://doi.org/10.48550/arXiv.1807.05118

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision -- ECCV 2014* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Luo, J., Niu, F., Lin, Z., Liu, M., & Yin, G. (2019). Recent acceleration of thaw slumping in permafrost terrain of Qinghai-Tibet Plateau: An example from the Beiluhe Region. *Geomorphology*, *341*, 79–85. https://doi.org/10.1016/j.geomorph.2019.05.020

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *152*(March), 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Marsh, J. H. (2008). Victoria Island. In *The Canadian Encyclopedia*. https://www.thecanadianencyclopedia.ca/en/article/victoria-island

Mesquita, P. S. (2008). *Effects of Retrogressive Permafrost Thaw Slumping on Benthic Macrophyte and Invertebrate Communities of Upland Tundra Lakes* [University of Victoria]. http://hdl.handle.net/1828/1041

Mu, C., Shang, J., Zhang, T., Fan, C., Wang, S., Peng, X., Zhong, W., Zhang, F., Mu, M., & Jia, L. (2020). Acceleration of thaw slump during 1997–2017 in the Qilian Mountains of the northern Qinghai-Tibetan plateau. *Landslides*, *17*(5), 1051–1062. https://doi.org/10.1007/s10346-020-01344-3

Nitze, I., Grosse, G., Jones, B. M., Romanovsky, V. E., & Boike, J. (2018). Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic. *Nature Communications*, *9*(1), 1–11. https://doi.org/10.1038/s41467-018-07663-3

Nitze, I., Heidler, K., Barth, S., & Grosse, G. (2021). Developing and testing a deep learning approach formapping retrogressive thaw slumps. *Remote Sensing*, *13*(21). https://doi.org/10.3390/rs13214294

Northwest Territories Geological Survey. (2010). *Permafrost Thaw Slumps*. https://www.nwtgeoscience.ca/project/summary/permafrost-thaw-slumps

Overland, J. E., Hanna, E., Hanssen-Bauer, I., Kim, S.-J., Walsh, J. E., Wang, M., & Bhatt, U. S. (2019). Surface Air Temperature [in "State of the Climate in 2018"]. In *Bulletin of the American Meteorological Society: Vol. 100(9)* (Issue 7, pp. 142–143). https://doi.org/doi:10.1175/2019BAMSStateoftheClimate.1.

Planet Team. (2017). *Planet application Program interface: in space for life on earth*. https://api.planet.com

Radosavljevic, B., Lantuit, H., Pollard, W., Overduin, P., Couture, N., Sachs, T., Helm, V., & Fritz, M. (2016). Erosion and Flooding—Threats to Coastal Infrastructure in the Arctic: A Case Study from Herschel Island, Yukon Territory, Canada. *Estuaries and Coasts*, *39*(4), 900–915. https://doi.org/10.1007/s12237-015-0046-0

Ramage, J. L., Irrgang, A. M., Herzschuh, U., Morgenstern, A., Couture, N., & Lantuit, H. (2017). Terrain controls on the occurrence of coastal retrogressive thaw slumps along the Yukon Coast, Canada. *Journal of Geophysical Research: Earth Surface*, *122*(9), 1619–1634. https://doi.org/10.1002/2017JF004231

Romanovsky, V. E., Smith, S. L., Isaksen, K., Shiklomanov, N. I., Streletskiy, D. A., Kholodov, A. L., Christiansen, H. H., Drozdov, D. S., Malkova, G. V., & Marchenko, S. S. (2019). Terrestrial Permafrost [in "State of the Climate in 2018"]. In *Bulletin of the American Meteorological Society: Vol. 100(9)* (Issue 7, pp. 153–160).

https://doi.org/doi:10.1175/2019BAMSStateoftheClimate.1.

Rudy, A. C. A., Lamoureux, S. F., Kokelj, S. V., Smith, I. R., & England, J. H. (2017). Accelerating Thermokarst Transforms Ice-Cored Terrain Triggering a Downstream Cascade to the Ocean. *Geophysical Research Letters*, *44*(21), 11,080-11,087. https://doi.org/10.1002/2017GL074912

Runge, A., Nitze, I., & Grosse, G. (2022). Remote sensing annual dynamics of rapid permafrost thaw disturbances with LandTrendr. *Remote Sensing of Environment*, *268*, 112752. https://doi.org/10.1016/j.rse.2021.112752

Saarela, J. M., Sokoloff, P. C., Gillespie, L. J., Bull, R. D., Bennett, B. A., & Ponomarenko, S. (2020). Vascular plants of Victoria Island (Northwest Territories and Nunavut, Canada): a specimen-based study of an Arctic flora. *PhytoKeys*, *141*, 1–330. https://doi.org/10.3897/phytokeys.141.48810

Segal, R. A., Lantz, T. C., & Kokelj, S. V. (2016). Acceleration of thaw slump activity in glaciated landscapes of the Western Canadian Arctic. *Environmental Research Letters*, *11*(3), 034025. https://doi.org/10.1088/1748-9326/11/3/034025

van der Sluijs, J., Kokelj, S. V., Fraser, R. H., Tunnicliffe, J., & Lacelle, D. (2018). Permafrost terrain dynamics and infrastructure impacts revealed by UAV photogrammetry and thermal imaging. *Remote Sensing*, *10*(11). https://doi.org/10.3390/rs10111734

Van Everdingen, R. O. (1998). *Multi-Language Glossary of Permafrost and Related Ground-Ice Terms in Chinese, English, French, German, Icelandic, Italian, Norwegian, Polish, Romanian, Russian, Spanish, and Swedish*. International Permafrost Association, Terminology Working Group. https://globalcryospherewatch.org/reference/glossary_docs/Glossary_of_Permafrost_and_Ground-Ice_IPA_2005.pdf

Walker, D. A., Reynolds, M. K., Daniëls, F. J. A., Einarsson, E., Elvebakk, A., Gould, W. A., Katenin, A. E., Kholod, S. S., Markon, C. J., Melnikov, E. S., Moskalenko, N. G., Talbot, S. S., Yurtsev, B. A., Bliss, L. C., Edlund, S. A., Zoltai, S. C., Wilhelm, M., Bay, C., Gudjónsson, G., … Vairin, B. A. (2005). The Circumpolar Arctic vegetation map. *Journal of Vegetation Science*, *16*(3), 267–282. https://doi.org/10.1111/j.1654-1103.2005.tb02365.x

Ward Jones, M. K., Pollard, W. H., & Jones, B. M. (2019). Rapid initialization of retrogressive thaw slumps in the Canadian high Arctic and their response to climate and terrain factors.

Environmental Research Letters, *14*(5), 055006. https://doi.org/10.1088/1748-9326/ab12fd

Appendix

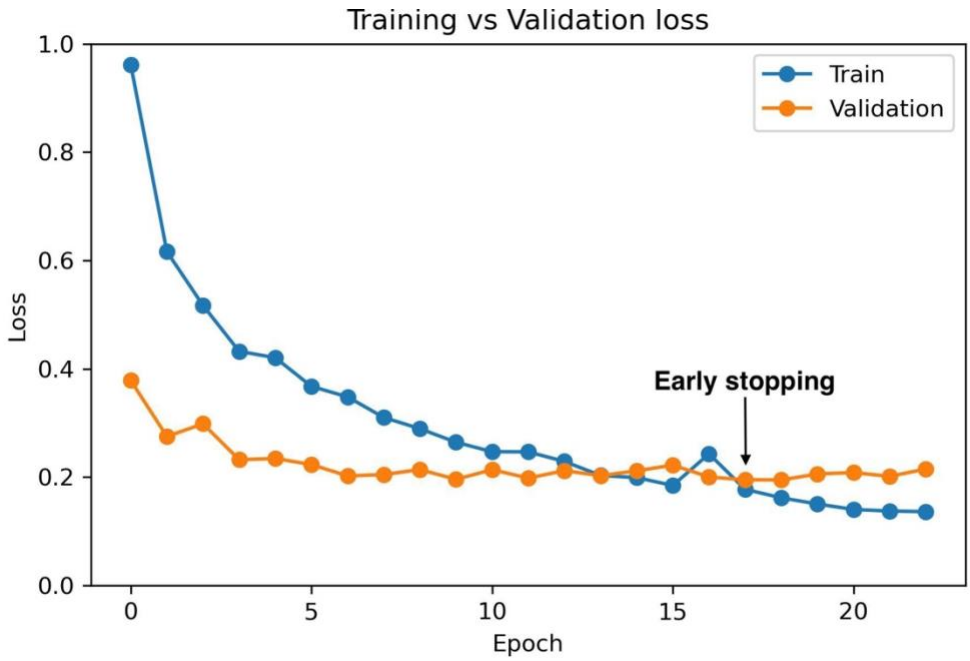

Figure A1. Loss curve of the model with the highest AP from model setting 6.

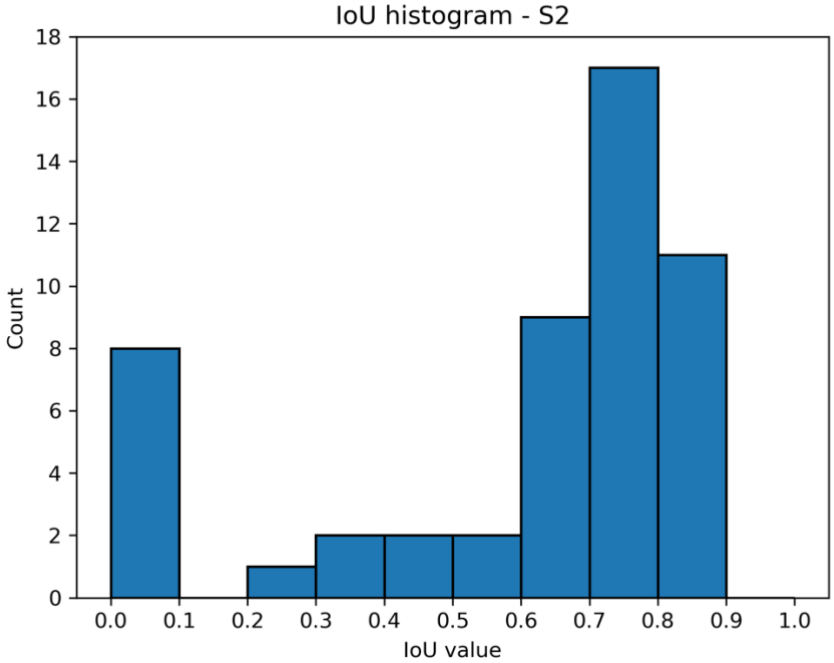

Figure A2. IoU histogram of digitized RTSs on S2.

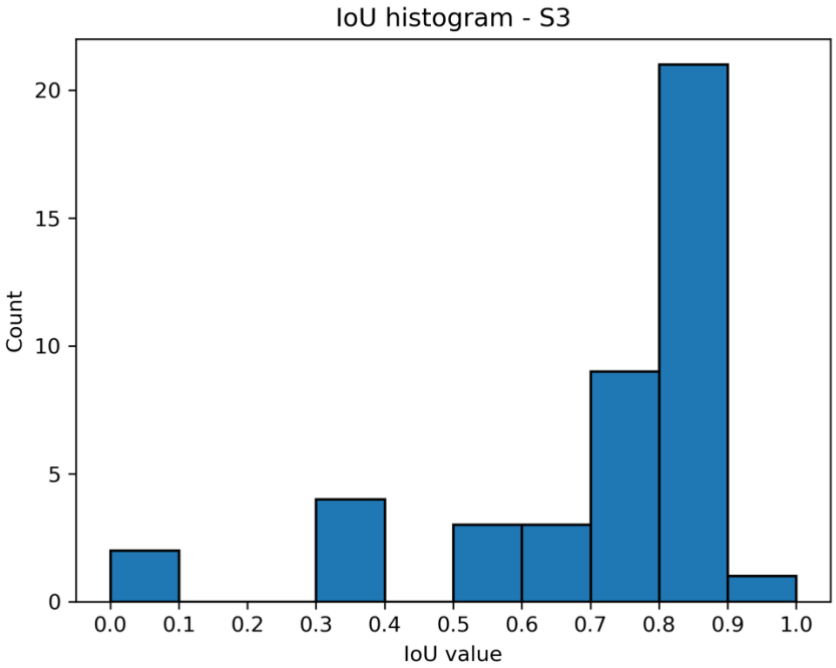

Figure A3 IoU histogram of digitized RTSs on S3.

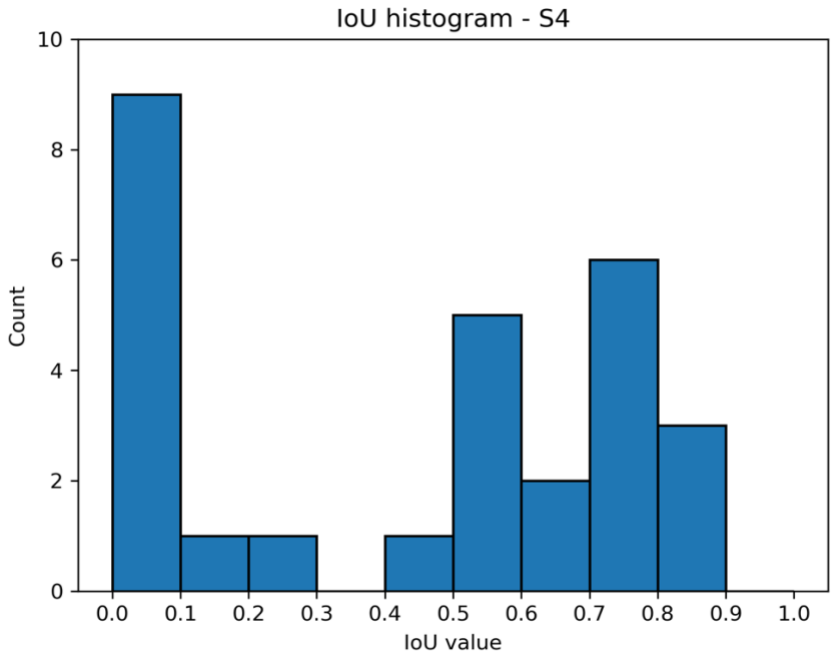

Figure A4. IoU histogram of digitized RTSs on S4.