Toward Automated Detection of Landfast Ice Polynyas in C-Band Synthetic Aperture Radar Imagery with Convolutional Neural Networks

by

Neil Brubacher

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Landfast ice polynyas — areas of open water surrounded by ice — are important features in many Northern coastal communities, and their automated detection from spaceborne synthetic aperture radar (SAR) imagery is positioned to support on-ice travel safety under changing Arctic sea ice and climate conditions. The characteristically small spatial scales and sparse distribution of landfast ice polynyas present key challenges to their detection, and limit the suitability of established methods developed for SAR-based sea ice and open water classification at broader spatial scales.

This thesis explores the development of deep learning-based object detection networks for landfast ice polynya detection in dual-polarized C-band SAR imagery, having three main contributions. The first is a characterization of landfast ice polynya signatures and separability in SAR imagery based on datasets of polynyas mapped over several seasons near the communities of Sanikiluaq, NU, and Nain, NL. Results from this analysis highlight the challenging and variable nature of polynya signatures in dual-polarized backscatter intensity, motivating the use of convolutional neural networks (CNNs) to capture relevant textural, geometric and contextual polynya features. The second contribution is the development and evaluation of CNN-based object detection networks for polynya detection, drawing on advancements in the natural-scene small object detection field to address the challenging size and sparsity characteristics of polynyas. A simplified detection network architecture optimized for polynya detection in terms of feature representation capacity, feature map resolution, and training loss balancing is found to reliably detect polynyas with sufficient size and local contrast, and demonstrates good generalization to regions not seen in training. The third contribution is an assessment of detection model generalizability between imagery produced by Sentinel-1 (S1) and Radarsat Constellation Mission (RCM) SAR sensors, illustrating the ability for models trained only on S1 imagery to effectively extract and classify polynya features in RCM despite differences in resolution and noise characteristics.

Across regions and sensors, missed polynyas are found to have smaller sizes and weaker signatures than detected polynyas, while false predictions are often caused by boundary areas between smooth and rough landfast ice. These represent fundamental limits to polynya / landfast ice separability in the medium-resolution, dual-polarized C-band SAR imagery used in this thesis, motivating future research into multi-temporal, multi-frequency, and/or higher-resolution SAR imagery for polynya detection. Ongoing and future progress in the development of robust landfast ice hazard detection systems is positioned to support community sea ice safety and monitoring.

Acknowledgements

The days, months, and years spent occupied with the work presented in this thesis were full of learning, growth, connection, and adventure, for which I owe thanks to many.

I first extend my gratitude to my supervisors, Drs. David Clausi and Andrea Scott, for encouraging and supporting my engagement with a diverse set of experiences well beyond those strictly necessary for degree requirements. Your patience, trust, and shoulder-taps enriched the past years tremendously, and have opened doors to subsequent chapters.

I would also like to thank Becky Segal, Joel Heath, and others from the Arctic Eider Society for their support and mentorship throughout this research. The breadth of my learning extended well beyond computer vision and remote sensing thanks to the people, perspectives, and priorities that AES represents, and I am grateful to have been involved in the SIKU project.

My appreciation extends to many in Sanikiluaq who welcomed and shared with me during my visit. In particular, thanks to Lisi Kavik-Mickiyuk for facilitating discussion, Lucassie Arragutainaq for providing valuable perspectives on the research, and Johnassie Ippak and Simeonie "Wayne Gretzky" Kavik for taking me out on the land.

Thanks to Shawna Dicker at SmartIce for producing and sharing polynya data for the Nain region, and to Dr. Katherine Wilson for facilitating the development of the datasets as well as for the valuable feedback provided at several stages of the research.

Too numerous to list are the friends and colleagues with whom engaging and broad-ranging conversations entertained, supported, and stretched me in many ways over the past years. Thank you John, Zach, Javier, Fernando, Muhammed, and many more – I appreciate each of you tremendously.

Allison; thank you for your enduring patience, for taking interest in my day-to-day, and for sharing grounding perspectives when things got stressful. I'm deeply grateful for your presence in the last years, and tremendously excited for our next adventure.

Finally, thanks to Drs. Linlin Xu and Richard Kelly for their keen review and valuable input to the content presented in this thesis; your volunteered time and expertise is greatly appreciated.

Dedication

Dedicated to Patrick Shipton, and all that you shared;
Bitter injustice is softened by glowing memories.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AP average precision 50, 56, 58, 60, 64–66, 78, 79

CNN convolutional neural network 15–17, 20, 25, 38, 83

FN false negative 22, 50

FP false positive 22, 50, 66, 79–81

FPN Feature Pyramid Network 19, 52, 53

FRCNN Faster-RCNN 18, 19, 39, 43, 51, 55, 61, 62, 69, 76

IoU intersection-over-union 21, 41, 49

IQ Inuit Qaujimajatuqangit 4, 23

NESZ noise-equivalent sigma zero 10, 34

NMS non-maximum suppression 21, 49

RCM Radarsat Constellation Mission 71, 75–79, 83

RoI Head Region-of-Interest Head 20, 41, 42, 52, 77

RPN Region Proposal Network 20, 41, 44, 52, 57, 76, 77

S1 Sentinel-1 25, 32, 44, 68, 71, 77–79, 83

S2 Sentinel-2 25, 67, 68

SAR synthetic aperture radar 5, 9, 18, 23, 38, 71, 83

TP true positive 22, 50, 81

List of Symbols

K_{neg} Ratio of negative image patches not containing polynyas to positive image patches containing polynyas used for patch sampling during model training 47, 54, 58, 60

δ^{μ}_{HH} Contrast measure between mean HH-polarized backscatter intensity from a mapped polynya area versus surrounding landfast ice 31, 32, 35, 43, 45, 46, 51, 57, 61, 62, 64–69, 76, 77, 81

δ^{μ}_{HV} Contrast measure between mean HV-polarized backscatter intensity from a mapped polynya area versus surrounding landfast ice 32, 35

σ^{0} SAR backscatter coefficient normalized to ground area 10, 25, 26, 30, 33, 37, 62, 66, 73, 81

τ_{IoU} Intersection-over-union threshold to categorize predicted and ground-truth bounding boxes as true positive, false positive, or false negative during evaluation 22, 50, 51, 57, 58, 63, 64, 69, 79

τ_{score} Threshold applied to predicted bounding box confidence scores defining the final set of model detections for inference and evaluation 22, 49, 50, 56, 58, 64, 69, 79

Chapter 1

Introduction

1.1 Motivation

In coastal communities across the circumpolar Arctic, landfast sea ice is relied upon for travel, sustenance hunting, and other activities throughout the winter season [16, 39, 63]. As ice and climate conditions change in response to warming air and ocean temperatures [100, 1], many Inuit — peoples Indigenous to Arctic regions — have expressed concerns related to the safety of on-ice travel [36, 141, 124]. In particular, some Inuit have identified the location of landfast ice polynyas — recurring areas of open water in landfast ice — as a priority piece of information that could assist in planning safe on-ice travel [16]. While experienced Inuit hunters and travellers are often familiar with recurring polynya locations, there have been concerns that the timing and locations of some polynyas are changing.

Synthetic Aperture Radar (SAR) imagery is well suited to provide information on community-scale ice conditions given its darkness and cloud penetrating capabilities, high spatial and temporal resolution at northern latitudes, and sensitivity to sea ice and water geophysical properties. Methods that automate the interpretation of SAR imagery offer the potential to extract relevant, salient and timely information about potential ice hazards such as polynyas, and could be leveraged to systematically monitor these features over time in support of geophysical and climate change-related studies. The uptake of digital platforms that connect those planning on-ice travel with satellite imagery and derived information products present opportunities for such automated hazard detection systems to deliver impact.

While community-relevant landfast ice polynyas have in some cases been documented by local ice experts [35] and studied in biological [131, 55] and socio-environmental [39, 119,

108] contexts, a previous lack of structured datasets cataloguing these small-scale features has precluded their previous study from a SAR pattern recognition perspective. Such a study is positioned to draw on known SAR sensitivities to sea ice and water, as well as recent advances in deep learning-based small object detection in other imaging contexts. The research described in this thesis represents a first exploration of deep learning-based landfast ice polynya detection from SAR imagery.

1.2 Thesis Objectives

The development and evaluation of deep learning models for landfast ice polynya detection in dual-polarized C-band SAR imagery is pursued with the following objectives:

1. Characterize landfast ice polynya signatures in SAR imagery and describe key challenges to their automated detection;
2. Develop deep learning-based object detection networks to address these challenges, and assess their suitability for automated localization of polynyas based on original datasets developed in collaboration with Inuit ice experts;
3. Evaluate the generalizability of developed polynya detection models between regions;
4. Evaluate the generalizability of developed polynya detection models between SAR sensors and image products.

1.3 Thesis Organization

The remainder of this thesis is outlined as follows:

- Chapter 2 provides background on the context and content of this research;
- Chapter 3 describes the development of original polynya datasets used as a basis for this study, and analyses the signatures and separability characteristics of polynyas from landfast ice in dual-polarized SAR imagery;
- Chapter 4 presents the key developments in adapting, optimizing, and assessing object detection networks for polynya detection in Sentinel-1 imagery;

- Chapter 5 presents an assessment of the generalization characteristics of models between Sentinel-1 and Radarsat Constellation Mission (RCM) imagery toward multi-sensor, multi-region hazard detection systems; and
- Chapter 6 concludes the thesis, reflecting on progress, limitations, and next steps.

Chapter 2

Background

2.1 Sea Ice in a Community Context

Sea ice is a fundamental part of life for Indigenous peoples across the circumpolar Arctic. Landfast sea ice — that fastened to land — has been used by Inuit as a platform for subsistence hunting, resource gathering, travel, and other cultural activities for thousands of years [63, 93]. The resulting expertise in sea ice features and conditions relevant to safe and efficient on-ice travel can be seen in the rich and detailed local terminologies for ice, snow, winds, and weather in communities across Inuit Nunaat [39, 8] — this body of knowledge, language, skills, and values passed down between generations is sometimes referred to as Inuit Qaujimajatuqangit (IQ) [134, 71]. As the Arctic warms at a rate exceeding the global average [24] and associated impacts such as thinning ice, changes in freeze- and break-up timing, and unexpected weather patterns are observed [1, 36, 107, 38], many Inuit have expressed concerns around the safety of on-ice travel that are compounded by socio-cultural changes challenging the transfer of IQ to younger generations [64, 8, 37].

Amidst these changes are opportunities to connect Inuit knowledge and technologically-derived information on ice and weather conditions in support of on-ice travel safety and climate change adaptation. Indeed, many Inuit regularly combine local knowledge and observations with information from online environmental services such as weather and tide forecasts when planning travel [25, 16], but have identified key gaps in these services related to interpretability, local relevance, and internet bandwidth requirements [124, 16]. Participatory mapping and training initiatives have been co-developed by Inuit and southern facilitators to help equip Inuit GIS specialists to regularly produce local travel safety maps highlighting ice features and hazards relevant to travel in their own communities [141, 140, 8].

Such mapping often leverages synthetic aperture radar (SAR) satellite imagery, which has been identified as a potentially useful source of information on community-scale ice conditions for its winter coverage of the Arctic, high spatial resolution, and sensitivity to sea ice and water geophysical properties [121, 77]. However, challenges in interpreting SAR images can pose a barrier to their use as a reliable information source [77, 121].

A digital platform that integrates many of these sources of information is the SIKU mobile application (siku.org) developed by the Arctic Eider Society (AES), an Inuit-driven organization based in Sanikiluaq, Nunavut. Many Inuit in communities across Canada, Alaska, and Greenland use SIKU to access satellite imagery and ice information products, share observations while out on the land, and manage community-based monitoring projects addressing local research priorities [16, 17]. SIKU's integration of observation- and knowledge-sharing between users, environmental information products, and satellite imagery provides a platform to develop and validate technologies in collaboration with communities that address local information priorities. For instance, some Inuit have identified the location of landfast ice polynyas as an important piece of information to know before travelling, due to their socio-ecological importance as well as to the hazards they pose to on-ice travel [16]

The regular availability of high-resolution satellite SAR imagery offers a means to identify sea ice hazards like polynyas when planning on-ice travel. In particular, there is a strong opportunity for algorithmic identification of polynyas in SAR imagery based on known microwave sensitivities to sea ice and open water. Outputs of such detection algorithms on a platform like SIKU could be weighed and validated against ice observations made by those on the ground, all within the same digital ecosystem. Reflecting a collaboration between AES and the University of Waterloo's Vision and Image Processing Lab, the algorithmic detection of landfast ice polynyas from SAR imagery is the focus of this thesis.

2.2 Landfast Ice Polynyas

Polynyas, broadly defined, are persistent and/or recurrent areas of open water or reduced ice concentration where thicker sea ice would typically be expected. Beyond this definition, there is significant diversity in the taxonomy of polynyas. A widespread categorization is of polynyas as either "latent-heat", wherein polynyas are maintained by the latent heat of fusion during ice production and export, or "sensible-heat", wherein ice production is prevented within the polynya due to the continual replacement of warm sea water [7]. Observing that these two mechanisms are often both present, Morales Maqueda et. al

[102] proposed a taxonomy of "shelf-water" (a.k.a. "coastal") and "deep-water" polynyas, differentiating between those occurring over or beyond continental shelves. Williams et. al [139] presented a framework for studying polynyas more directly in terms of their particular forcing mechanisms, categorized as "mechanical" factors causing ice divergence and "heating" factors preventing ice growth.

Many wind-driven, latent-heat coastal polynyas on the scale of hundreds to thousands of kilometers in extent are important sites of atmosphere-ocean heat flux, gas exchange, sea ice production and associated high-density water masses that contribute to thermohaline circulation [50, 7]. Such large-scale polynyas have been studied extensively with remote sensing data to characterize and infer polynya dynamics, causal mechanisms, and sea ice production, e.g. [50, 132, 58, 94]. In contrast, sensible-heat polynyas occurring *within* landfast sea ice of relevance to on-ice travel near coastal communities, referred to herein as "landfast ice polynyas", have spatial scales on the order of hundreds of meters and are typically created and maintained by tidal forcing of warm seawater over sills or through topographic channels [131, 139].

Landfast ice polynyas are ecological hotspots [131, 55], and key socio-environmental features for nearby communities [119, 108, 39]. Indeed, the breadth of terminologies used by Inuit across and within regions to describe polynyas, e.g. [10, 66, 39, 121], demonstrates their importance and prevalence in many communities. While the locations and patterns of polynyas have been documented in some cases by local ice experts, e.g. [35, 39], the detection of small-scale landfast ice polynyas has not previously been studied from a remote sensing perspective.

2.3 Remote Sensing of Polynyas with Synthetic Aperture Radar

2.3.1 Why SAR?

Polynyas, as with most things, are most easily seen and interpreted by the human eye in the visible spectrum of light. Spaceborne "optical" sensors, operating with similar theory to everyday cameras, capture reflected sunlight in the visible range and can provide high-resolution (centimeter- to meter-scale) images in which thick sea ice and open water can clearly be distinguished. However, optical sensors possesses two fundamental constraints that limit their utility to the spaceborne remote sensing of sea ice. First, the "passive" (reflectance-based) nature of these sensors means that sunlight is required to form an image.

Figure 2.1: Panorama photograph of a landfast ice polynya near Sanikiluaq, NU, captured on foot in January 2024.

Second, the visible-spectrum wavelengths captured by optical sensors do not penetrate cloud. In Arctic coastal environments, cloud cover is frequent and much of the winter season sees little to no sunlight due to Earth's axial tilt. Optical image availability is thus limited or nonexistent for monitoring sea ice in the winter. Sensors operating at neighbouring infrared wavelengths sensitive to surface temperature also suffer occlusion by cloud cover.

Passive microwave (PM) sensors overcome these limitations by capturing microwave energy emitted by the earth's surface. The brightness temperatures of sea ice and water are separable, and PM has been used as the workhorse data for monitoring regional-scale sea ice concentration (SIC) for decades [117]. PM data have also been used for large-scale polynya study based on SIC thresholds defining regions of open water (OW). However, the relatively low microwave energy intensity naturally emitted by materials limits the available signal and ultimately spatial resolution of PM sensors. For instance, the AMSR2 PM sensor has a maximum spatial resolution of 3x5 km. Identifying this resolution limitation in studying small-scale polynyas, Liu et. al used an original super-resolution network [88] to downscale AMSR2 PM images and use derived SIC estimates to study a transient polynya event off the coast of Greenland [89]. While this approach improved results for the study of coastal polynyas, even the 1.5x1.5 kilometer footprints of the fourfold super-resolved AMSR2 pixels are still too large to resolve many landfast ice polynyas.

In contrast, "active" sensors that emit their own electromagnetic energy in the mi-

crowave spectrum offer reliable, year-round coverage of polar regions at meter-scale resolutions. Synthetic Aperture Radar (SAR) sensors have been developed to provide such valuable imagery, and are a basis for the contributions developed in this thesis.

2.3.2 SAR Fundamentals

The general principle of radar (RAdio Detection And Ranging) systems is the directed emission of electromagnetic (EM) energy and subsequent reception of signals "backscattered" from objects within the radar aperture. The received signal power from a radar-illuminated target is determined by the transmitted power, P_t, the distance from radar to target, r, the antenna gain at the target, G_t, the "radar cross-section" of the target, σ, and the effective area of the receiving antenna, A_r [112, 143]:

$$P_r = \frac{P_t}{4\pi r^2} \times G_t \times \frac{\sigma}{4\pi r^2} \times A_r \tag{2.1}$$

Mathematically, σ represents the cross-sectional area of a sphere having a perfectly-conducting surface that re-radiates incident EM waves completely and uniformly in all directions. In practice, particularly in the context of radar-based geophysical measurement, target objects or surfaces generally absorb some proportion of incident energy, and the energy they do return is rarely uniformly scattered. Broadly, the proportion and direction of incident radar energy scattered by a surface is a function of its dielectric permittivity (electrical properties influenced by, e.g., salinity and moisture), its roughness at scales similar to the radar wavelength, λ, and its geometry (at scales larger than λ) relative to the radar [80, 51, 143]. Differences in these characteristics between materials and surfaces at various EM wavelengths and polarizations is good news for radar remote sensing, as it allows for distinction between targets. Given measurements taken with a fixed and known set of radar parameters, equation 2.1 can be inverted to find σ and thus infer target properties.

To exploit the sensitivity of radar backscatter to surface geophysical properties and map features of interest from space, satellite sensors must produce two-dimensional images with sufficiently high spatial resolution. While traditional "real-aperture" radars can be flown in a side-looking configuration to generate such two-dimensional (range and azimuth) images, their resolution is limited by the dimensions of the transmitting antenna. For instance, a radar flown on a satellite orbiting 693km above the Earth, slanted from nadir by 20°, transmitting at a wavelength of 5cm with a 12.3-meter antenna could, at the theoretical limit described by Moreira et. al [103], achieve a spatial resolution of 3 kilometers in

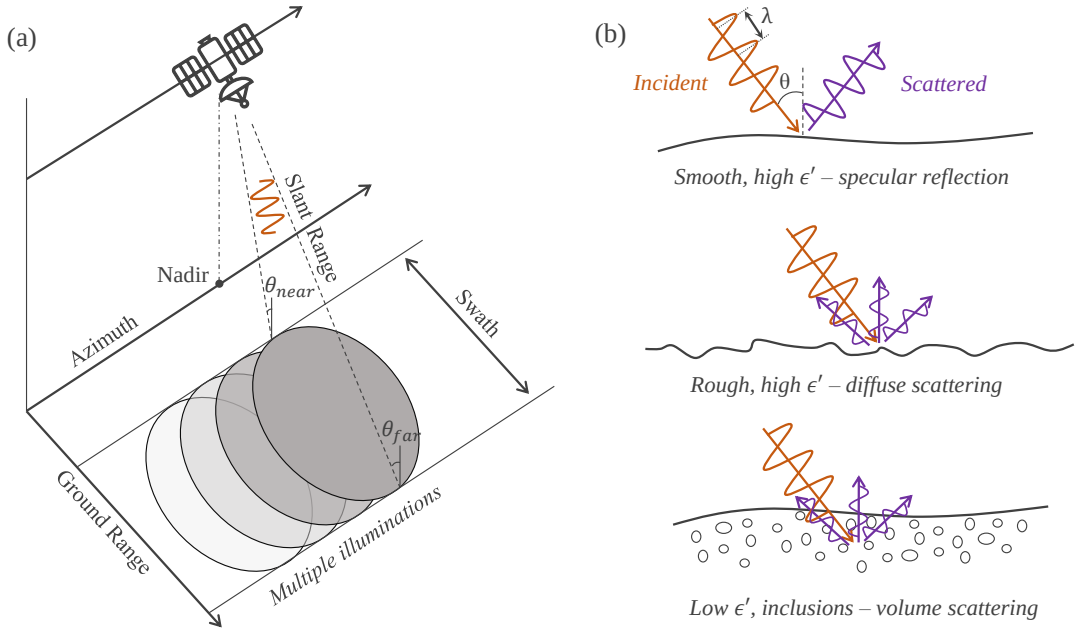

Figure 2.2: (a) Basic geometry of a spaceborne SAR system (not drawn to scale) and (b) different scattering scenarios based on surface properties. θ refers to incidence angle, λ to radar wavelength, and ϵ' to the real component of the relative complex permittivity of the surface, e.g. described in [52].

azimuth — clearly insufficient to delimit targets on the scale of hundreds of meters. By coherently summing multiple illuminations of a surface during an overpass, a radar system can simulate a larger physical antenna via precise signal processing and thereby drastically increase spatial resolution. This technique is called "synthetic aperture" radar (SAR) and has been widely used for aerial and spaceborne radar remote sensing since the 1980s [56]. Using the same radar parameters described above, the European Space Agency's Sentinel-1 SAR satellite produces images with spatial resolutions finer than 5 meters [32].

SAR systems have several parameters important to geophysical remote sensing applications including polynya detection. These properties are briefly summarized here.

Wavelength, λ, describes the EM wavelength transmitted and received by the SAR system. Longer wavelengths have lower spatial resolution for the same antenna size, but penetrate farther into materials [52] yielding higher potential for volume scattering mechanisms (Fig 2.2b).

Polarization describes the orientation of EM waves transmitted and received by the SAR sensor. For instance, "HV" indicates EM waves transmitted at horizontal polarization and corresponding backscattered signals measured at vertical polarization. SAR systems may transmit and/or measure near-simultaneously at different polarizations, e.g. providing coincident "dual-polarized" HH and HV images. The known polarization of outgoing and incoming energy lends another window through which to infer target properties.

Incidence Angle (IA), θ, describes the angle of incident EM energy transmitted by a SAR system from vertical relative to Earth's surface, or a model thereof. Higher IA corresponds to a more oblique viewing geometry; distance between sensor and target increases with IA, but so too does roughness in terms of perceived vertical surface fluctuations.

Spatial resolution, often given in ground range and azimuth coordinates, defines the smallest distance between two targets that can be delimited unambiguously by a SAR system. SAR satellites often produce images with different spatial resolutions by using different imaging modes. The primary tradeoff to producing higher-resolution SAR images is generally an associated reduction in swath width (i.e., spatial coverage), which ultimately decreases the available temporal resolution for a given region. To account for varying spatial resolutions across SAR image products, the measured backscatter intensity σ must be normalized to the area over which the measurement was made [143]. For sufficiently level surfaces the *ground area* defined by an ellipsoidal Earth model is typically used [126], producing the "backscatter coefficient" $\sigma^0 = \sigma/A$. This coefficient is then usually converted to decibels (dB) [122]: $\sigma^0 = 10\log(\sigma^0_{linear})$.

Noise-equivalent sigma zero (NESZ) is an estimate of the thermal noise floor of the SAR receiver, expressed in σ^0. Measurements at or near this threshold provide little to no information regarding the imaged surface or target, being instead dominated by noise.

The research described in this thesis leverages C-band ($\lambda \approx 5.5cm$), dual-polarized HH-HV SAR imagery based on its operational availability over Arctic coastal regions. The dataset is described in chapter 3.

2.3.3 SAR Sensitivities to Sea Ice and Open Water

The sensitivities of HH and HV C-band SAR backscatter to snow, sea ice, and seawater properties are well understood, having benefited from several decades of polar microwave remote sensing research. The key sensitivities relevant to landfast ice polynya detection are summarized here.

Seawater has high a high dielectric permittivity due to its polarity and salinity content, and thus causes surface scattering [122]. Under calm conditions, specular reflection leads to

backscatter intensity that decreases rapidly, and approximately linearly in the logarithmic domain, with incidence angle [92]. Wind roughening of open water creates surface waves to which SAR backscatter is sensitive as a function of wavelength, polarization, and imaging geometry; in general, stronger winds aligned with the SAR range axis lead to increased backscatter intensity in HH, while HV is less affected [27].

Microwave backscatter from seasonal ("first-year") sea ice, including landfast ice, is also dominated by surface scattering mechanisms due to persistent salinity within the ice [122]. Surface roughness and geometry relative to the SAR sensor are thus the primary controls on backscatter intensity. Level first-year sea ice tends to produce low backscatter intensity at both HH and HV polarizations, with HV backscatter often being below the noise floor of the SAR system [27]. Deformed ice features such as pressure ridges and piled ice around polynyas or near the floe edge increase backscatter intensity, particularly in HV due to strong depolarization effects [27]. HH backscatter intensity has been shown to decrease linearly with incidence angle over first-year ice [96, 99], with HV showing a similar but less pronounced trend [98, 91]. Thin ice, such as that which may form within polynyas, has been observed to produce backscatter intensities in HH polarization that increase rapidly with ice growth due in part to the development of rough crystalline structures on the otherwise smooth ice surface, e.g. [97].

With the possible exception of newly formed ice in cracks, leads, and polynyas, sea ice is almost always covered by snow. At the low temperatures typical of winter-period Arctic environments, snow tends to be sufficiently "dry" for C-band microwaves to transmit through and scatter at the ice/snow interface. As air and snow temperatures increase, saline liquid water content (brine) develops and contributes to microwave scattering within the snowpack [43]. Increased brine volume fraction has been shown to increase C-band backscatter over snow-covered first-year ice during transitions from cold to warmer air temperatures, e.g. [6, 104].

Together, these established sensitivities yield an expectation for the dual-polarized C-band signatures of landfast ice polynyas. At sufficiently high incidence angles and low wind roughening conditions, open water polynyas are expected to yield lower backscatter intensities than landfast ice, although these signatures may overlap with low backscatter intensities over smooth landfast ice. Low incidence angles and/or high wind roughening may produce higher backscatter over open water than over landfast ice. Thin ice development within polynyas may cause low to high backscatter (relative to surrounding landfast ice) depending on the stage of growth. HV-polarized backscatter may offer some utility in identifying areas of deformed landfast ice, however this utility may be limited by proximity to the sensor noise floor. Backscatter over landfast ice may increase as melt onset approaches due to increased brine volume fraction in the overlying snowpack. In summary,

Figure 2.3: Sentinel-1 SAR image (σ^0 dB) acquired May 5, 2016, over the Belcher Islands in Hudson Bay. Composited as (R=HV, G=HH, B=HH), with each band clipped to (0.5, 99.5) percentiles. Annotations indicate different sensitivities of HH and HV backscatter intensity to ice types and open water. Relatively high HH backscatter over smooth landfast ice in this image may reflect a brine-wetted snowpack as spring air temperatures increase.

polynya signatures are expected to vary with a number of imaging and environmental factors. This is explored empirically in chapter 3.

2.3.4 Methods for Sea Ice / Water Classification with SAR

Classification, in the broadest terms, refers to the process of assigning a name or label, $\hat{y}$, to a set of features derived from measurements or observations, $\underline{x}$, of some configuration of reality, y. In a SAR remote sensing context, an image of backscatter values represents a set of measurements, $\underline{x}$, and the remote sensing scientist's definition of what was imaged (type of land cover [110], tonnes of biomass per hectare [125], polynya or landfast ice, etc.) constitutes the label, y. In general there exists some underlying "forward model", f, that completely describes the observation process:

$$\underline{x} = f(y); \tag{2.2}$$

The objective of classification is to solve the "inverse problem", i.e., invert f to infer y from $\underline{x}$:

$$\hat{y} = g(\underline{x}),\ where\ g \approx f^{-1}. \tag{2.3}$$

In practice f is often unknown due to the complex physical processes and sources of noise that make up real-world measurement, and is rarely invertible via analytical means. f^{-1} is thus often estimated empirically from a set of previously obtained $\{\underline{x},\ y\}$ pairs ("labelled" or "training" data) via an optimization scheme that seeks to minimize the difference between model prediction, $\hat{y}$, and label, y. This approach to classification model development is called *supervised learning*.

The polynya detection research described in this thesis follows a supervised learning approach, drawing inspiration from a large body of previous research on SAR-based sea ice and open water classification. This body of research includes investigations into different types of SAR imagery and derived features ($\underline{x}$)[1], classification models (g), and target labels (y). A brief summary highlighting methods and findings relevant to landfast ice polynya detection in dual-polarized C-band SAR imagery is given subsequently.

[1] A more rigorous notation might distinguish between *measurements* and *features*, which have been lumped together here for brevity. Indeed, as will be seen, some forms of g operate on derived features while others operate directly on measurements.

Pixel-Based Classifiers

The most straightforward classifiers use a SAR image pixel's backscatter intensity directly to infer class membership (e.g. sea ice type or open water) based on thresholds chosen to minimize expected error based on per-class backscatter distributions in a training/analysis dataset. This has been implemented for first-year ice and open water classification from single- and dual-polarized SAR imagery with, e.g., Bayesian [91], decision tree [40], and neural network [73] classifiers. While these approaches yield some success, the overlap in dual co-/cross-polarized backscatter distributions between classes under different environmental and imaging conditions has been found to limit classifier performance. Furthermore, the pixel-based nature of such classifiers makes them sensitive to multiplicative speckle noise present in SAR imagery. It is widely recognized that the incorporation of spatial features is positioned to better discriminate sea ice and water classes from SAR images [122].

Textural Features

The utility of SAR image texture features was identified early in the exploration of automated sea ice classification models, guided by the hypothesis that sea ice and open water surfaces would present distinct spatial patterns and correlation scales based on their different physical properties.

Similä [123] investigated the use of the autocorrelation feature on two VV-polarized ERS-1 SAR images for sea ice and water classification, and indeed observed higher correlation for sea ice than water at spatial windows ranging from 700m^2 to 1.9km^2. Karvonen et. al [72] later used autocorrelation implemented on a segment-wise basis to classify open water and first-year sea ice in HH-polarized Radarsat-1 SAR imagery. This method achieved 90% overall accuracy for open water classification, but the authors note that the algorithm tended to misclassify small areas of water as ice in part due to signature ambiguities with level landfast ice areas.

Grey level co-occurrence matrix (GLCM) features are another texture measure that have been investigated [5, 22, 92, 118] and applied operationally [79, 145] for sea ice type and ice / open water classification from dual-polarized SAR imagery. GLCM texture features have also been explored in studies investigating the detection of small-scale open water features within first-year sea ice. For instance, Zitman [148] studied the detection of hundred meter-scale plume-driven polynyas from dual-polarized Sentinel-1 SAR imagery using GLCM features. A limitation to small-target detection consistently encountered by these

studies is the positive relationship between window size and classification performance. Indeed, Zitman found that window sizes less than 440m^2 did not produce satisfactory classification between polynya open water and surrounding sea ice. Lohse [92] similarly observed that classification accuracy between open water and first-year ice increased steadily with window size from 200m^2 (worst performance) to 2.04km^2 (best performance). This pattern suggests a limitation in the ability of GLCM and other window-based texture measures to effectively detect landfast ice polynyas on the scale of hundreds of meters.

Deep Learning Approaches

Rather than relying on task-agnostic image features that require manual design and selection, deep learning classifiers leverage large amounts of data to learn task-specific image features that optimize the classification objective. Convolutional neural networks (CNNs) are a type of deep learning model originally developed for natural-scene image analysis, and have been applied widely across many domains. Indeed, many recent approaches to SAR-based sea ice classification have demonstrated the efficacy of CNNs in learning discriminative features to accurately classify ice types and open water under a range of conditions [61, 82]. CNN-based approaches have been formulated as either image patch classification tasks, whereby the network makes one prediction per image, e.g. [12, 76, 54, 74, 101] or semantic segmentation tasks, whereby the network makes a prediction per image pixel, e.g. [137, 114, 83, 28, 20].

While CNN-based segmentation networks have been found to work well for relatively balanced ice/water classes, the comparative sparsity and small size of landfast-ice polynya targets challenge the suitability these models, which traditionally struggle to capture fine-grained spatial detail [19, 69]. On the other hand, patch classification methods do not yield sufficient localisation precision for small polynya targets, and are at greater risk of overfitting to land formations correlated with polynya occurrence in repeat-observation training datasets. These characteristics, along with the compact geometries of polynyas, motivate an object detection formulation for landfast ice polynya detection.

2.4 Convolutional Object Detection Networks

Object detection networks refer generally to deep learning-based classifiers that seek to predict a set of bounding boxes and associated class scores that delimit targets of interest in an input image. Convolutional object detection networks are those that leverage convolutional neural networks to extract learned features from input images, from which

bounding box predictions are made. The fundamental building blocks of convolutional object detection networks are summarized in this section, laying groundwork for experimental investigations conducted in chapter 4.

2.4.1 Convolutional Neural Networks

The theory and mechanisms behind convolutional neural networks (CNNs) stem from fully-connected artificial neural networks (ANNs). ANNs are directed graphs composed of sequentially-connected layers of simple computational units (biologically-inspired "neurons") parameterized by learnable connection weights. Equation 2.4 describes the output o of a single neuron i of layer l in response to the outputs of all N neurons in the previous layer $l-1$. This response is governed by connection weights w_{ij}, a bias term b_i, and a nonlinear activation function f_a. By composing a sufficient number of neurons and layers in this way, ANNs can theoretically approximate any continuous function [59].

$$o_i^l = f_a(\sum_{j=1}^{N} w_{ij} o_j^{l-1} + b_i) \tag{2.4}$$

While ANNs can be applied directly to image analysis by assigning an input neuron to each image pixel [78], the fully-connected layers are extremely parameter-dense and thus produce models that are computationally inefficient and difficult to optimize with limited training data. Rather than explicitly modelling the interactions of *all* image pixels at once, CNNs model only *local* interactions of pixels via sliding windows with translation-invariant parameters. This is implemented with the 2D convolution operation [46]:

$$(g * f)[x, y] = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f[x-i, y-j] g[i, j] \tag{2.5}$$

Here, f represents the image and g a convolutional kernel (a relatively small matrix with translation-invariant weight values) with a center at $[i, j] = [0, 0]$. f and g must have the same number of dimensions — if f is an image with 3 channels, g should be HxWx3 (e.g., a commonly used CNN kernel size is 3x3). At each location $[x, y]$, the convolution operation yields the sum of the Hadamard (element-wise) product between f and g centered at $[x, y]$. Over all $[x, y]$ this can be understood as "sliding" the kernel across the image and producing a 2D "feature map" identical in size to the input image[2]. CNNs treat convolutional kernel

[2]This assumes the use of image-boundary padding, which is typically the case in CNNs

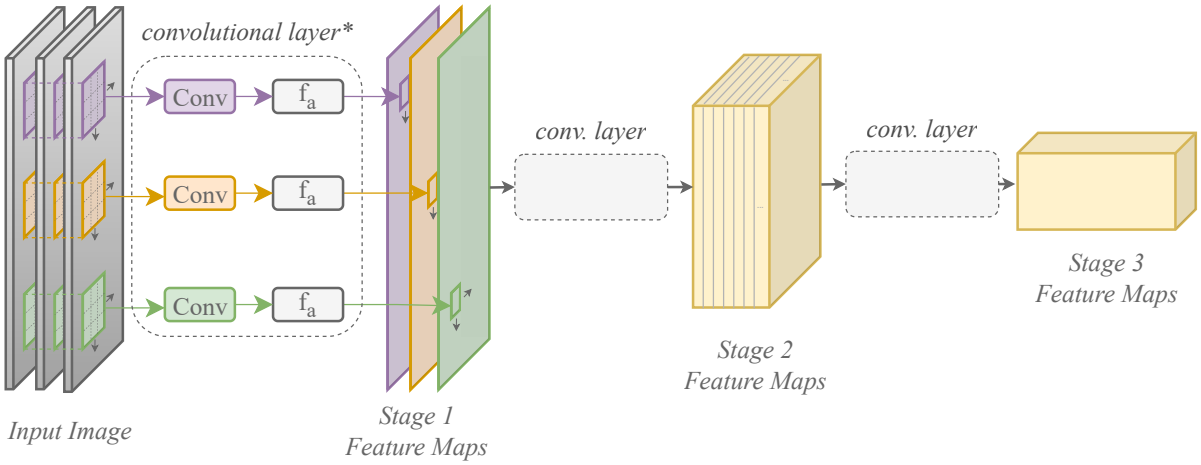

Figure 2.4: Illustration of a standard CNN architecture. f_a represents a nonlinear activation function. A first convolutional layer with three filters is depicted; deeper layers typically use more filters and thus produce more feature maps.
*Conv. layers often include other operations not illustrated, e.g. batch normalization and downsampling.

weights as learnable parameters analogous to $\underline{w}$ of ANNs, creating learnable translation-invariant feature extractors that possess relatively few parameters independent of input image size.

The architecture of a CNN subsequently follows similar compositional principles to that of an ANN: each convolutional layer is composed of multiple kernels and a nonlinear activation function, and several sequential layers, each operating on the outputs of the previous layer, are used. This repeated nonlinear composition of spatial features progressively expands the field of view, or "receptive field", of CNNs, allowing the networks to capture both small-scale intensity patterns and broader-scale contextual information within images at various levels of abstraction. For instance, it has been observed that many CNNs trained on natural images learn features similar to Gabor filters and colour blobs in early layers, while deeper layers learn higher-level task-specific features [144].

Most CNN architectures increase the number of kernels, or feature "channels", with network depth to increase capacity for learning complex relationships between lower-level features needed for accurate task-specific pattern recognition. A practical challenge in expanding feature representation capacity in this way is the memory requirements associated with storing increasing numbers of full-resolution feature maps. For this reason, and to increase the receptive fields of deeper layers, it is common practice to progressively downsample the spatial resolution of feature maps concurrent with the expansion of feature

channels. This is often implemented with pooling operations or strided convolutions.

CNNs are typically optimized iteratively using the backpropagation algorithm [116], adjusting network weights based on their contributions to prediction error as measured by an objective function. This is done by calculating the gradient of a differentiable loss function with respect to individual weights throughout the network, an optimization scheme known as gradient descent [3]. In a supervised training context this loss function is often a measure of dissimilarity between the model prediction, $\hat{y}$, and the training label, y. Loss function selection is an important design element in the development of a CNN as it ultimately guides what the network learns.

2.4.2 Developments in Small Object Detection

Building on and contributing to the remarkable success of CNNs for image analysis, the natural-scene computer vision field has driven the rapid development of convolutional object detection networks in the last decade [87, 149]. Detection networks have broadly developed along two phylogenies: "two-stage" detectors, which include a region proposal module followed by a more fine-grained classifier, and "single-stage" detectors, which discard the extra step and predict final detections directly. Two-stage detectors generally achieve higher detection accuracy at the cost of increased computation, while single-stage detectors are faster but tend to perform worst on dense and/or small objects [149]. "YOLO" [111] and "Faster-RCNN" [113] are foundational single- and two-stage detection models, respectively. A persistent challenge in both natural-scene and optical remote sensing image analysis contexts is that of detecting small targets with characteristically low spatial extents, weak features, and, frequently, sparse occurrence patterns that result in highly imbalanced training datasets [90, 53]. These challenges of target size and sparsity are clearly present in the SAR-based landfast ice polynya detection problem, motivating the examination of methods developed within the "small object detection" domain to address these limitations.

Two broad approaches have been proposed regarding the detection of small targets: multi-scale feature fusion and input super-resolution. In traditional object detection network configurations, only the last, deepest feature layer produced by a CNN is used for bounding box prediction. Spatial downsampling operations within the CNN run the risk of discarding important discriminative and/or localization information on small targets available at higher resolutions in earlier layers. However, the more complex features of the deepest layer may still be needed for accurate target classification. The fusion of high spatial resolution, shallow features with low-resolution, deep features was proposed by Lin et.

al [84] for object detection. This method, developed on top of the Faster-RCNN detector and called "Feature Pyramid Network", achieved state-of-the-art performance on small object detection and continues to be referenced widely as a baseline [149]. Farther upstream, input super-resolution approaches seek to increase the pixel extent of small objects in the input image while enhancing their discriminative feature information. This approach, often realized with generative adversarial networks (GANs), has been found to improve results on natural scene [105] and optical remote sensing [136] small object detection tasks.

In settings of target sparsity relative to the "background" class, the supervised training process is at risk of being guided too heavily by a vast number of background examples and not sufficiently learning discriminative features or decision boundaries for the rare target class. Two key approaches proposed for handling these difficulties are sampling positive and negative examples during training, and the use of modified loss functions to better focus on positive examples. Positive/negative sampling strategies seek to constrain the training loss signal to focus on extracting and discriminating representative target and background features. This has been realized both with negative example sub-sampling, e.g. as implemented at the bounding box level in Faster-RCNN [113], and positive example over-sampling, e.g. [75]. A risk of negative sub-sampling is a loss of representative negative samples, while positive over-sampling can lead the model to overfit to specific targets; data augmentation strategies can be leveraged to decrease the risk of the latter. Rather than discarding negative samples, Lin et. al [85] proposed an alternative loss function, "focal" loss, that down-weighs loss for more confident predictions while increasing loss for less confident predictions, prompting the network to "focus" more on rare, poorly-classified targets during training. This loss function significantly improved performance over the more standard cross-entropy loss in the extremely imbalanced single-stage detection setting [85].

While hundreds of papers have been published on the topic of small object detection in natural scene and optical remote sensing images, most seek to address the same key challenges related to image and feature resolution and positive/negative example representation during training [90, 53]. The methods discussed represent foundational approaches to addressing these challenges, warranting investigation for the development of baseline landfast ice polynya detection networks.

2.4.3 Faster-RCNN and Feature Pyramid Network

The two-stage Faster-RCNN (FRCNN) object detection network [44] with Lin et. al's Feature Pyramid Network (FPN) [84] was chosen as a foundational architecture for this

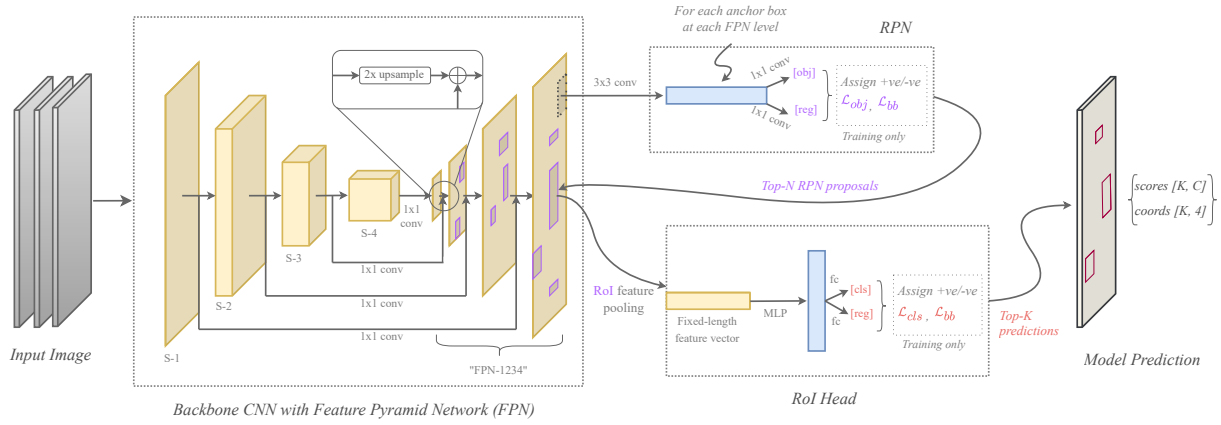

Figure 2.5: Architectural diagram of the Faster-RCNN object detection network with Feature Pyramid Network, as interpreted from the original papers [113, 84]. The backbone CNN is abbreviated (fewer stages than in [84]) for clarity.

research based on its widespread use as a baseline detector and demonstrated advantages over one-stage detection networks on small-object detection tasks in natural-scene and remote sensing contexts [53, 90, 87]. FRCNN comprises three sub-networks, as illustrated in Fig. 2.5:

1. A "backbone" CNN, chosen here as FPN, that extracts learned features from input images;
2. A lightweight CNN-based Region Proposal Network (RPN) that, operating on feature maps produced by the backbone network in a sliding-window fashion, predicts binary "objectness" scores and localization offsets for a set of prior "anchor boxes" of various sizes and aspect ratios sampled densely across the feature map, yielding a set of score-ranked object proposals;
3. A Region-of-Interest Head (RoI Head) network comprising fully-connected neural network layers that predict class probability scores and refined bounding box positions for the top-N scoring RPN object proposals from their corresponding backbone-produced features.

In the backbone network, FPN attempts to synthesize semantically strong, high-resolution feature maps by sequentially merging low-resolution feature maps produced by deeper network stages with higher-resolution feature maps produced by shallower network stages. As

described in [84], this is done on a pairwise basis by first subjecting both feature maps to a learnable 1x1 convolution to expand or compress the number of feature channels to a common value, then upsampling the deeper features to the dimension of the shallower features before combining the feature maps via element-wise addition. A learnable 3x3 convolution is applied following feature addition to reduce potential aliasing caused by upsampling. The backbone network stages included in the FPN is an important design element that influences the complexity and resolution of features used for downstream detection. Several configurations are explored for polynya detection in section 4.3.

FRCNN is typically trained in an end-to-end fashion (all three sub-networks jointly) based on four loss functions, two in the RPN and two in the RoI Head sub-networks. By the standard implementation, both sub-networks apply cross-entropy loss to objectness and class probability scores, respectively, and a smoothed L1 loss [44] to the x- and y-coordinate offsets between predicted boxes classified as "positive" and the ground-truth boxes they intersect. Of these four loss functions, that applied to objectness scores in the RPN, $\mathcal{L}_{obj}$, is of particular importance to small, sparse object detection problems. This scoring is responsible for selecting a small proportion of candidate boxes from thousands of densely-sampled anchors, and thus faces the greatest positive/negative class imbalance in the model. It is typically implemented as binary cross-entropy (CE) loss as in equation 2.6, where score $p_{obj} \in [0, 1]$ represents the predicted objectness probability of an anchor box with label $y \in \{0, 1\}$. This and alternative RPN objectness loss functions are also investigated in section 4.3.

$$\mathcal{L}_{obj}^{CE} = \left\{ \begin{array}{ll} -log(p_{obj}) & if\ y = 1 \\ -log(1 - p_{obj}) & otherwise \end{array} \right\} \tag{2.6}$$

During FRCNN model training and inference there are several operations that are not learned, but rather implemented with user-defined hyperparameters. The categorization of "positive" and "negative" predicted boxes, for instance, is implemented with user-defined thresholds on the intersection-over-union (IoU) metric, a measure of overlap between a predicted box, A, and ground-truth box, B, described by equation 2.7. Such definitions are needed for class-balanced loss sampling during training and predicted box categorization during evaluation. The greedy non-maximum suppression (NMS) algorithm described in [45] is applied in the RPN and RoI Head sub-networks to eliminate redundant or conflicting predictions with high overlap, also based on user-defined IoU thresholds. The RPN and RoI Head sub-networks only forward a limited number of top-scoring boxes based on user-defined maximums that reflect the expected number of targets in each image. These and other hyperparameters are discussed further in section 4.1.2.

$$IoU = \frac{A \cap B}{A \cup B} \tag{2.7}$$

Finally, during model inference, predicted boxes from the RoI Head sub-network having maximum target-class confidence scores below a user-defined threshold, τ_{score}, are discarded. This threshold affords control over the degree of confidence required for a predicted box to be considered a model "detection". During evaluation, detections scoring above τ_{score} are categorized against ground-truth labels using an IoU threshold, τ_{IoU}, which describes the overlap necessary to categorize each detection as a true positive (TP) or false positive (FP), and each target as true positive or false negative (FN). τ_{score} and τ_{IoU} differ from model hyperparameters as they can be modified *after* model training, e.g. to evaluate a model in different ways. Evaluation methods and metrics are described in more detail in section 4.2.8.

Chapter 3

Dataset Synthesis & Characterization

Given a lack of previous studies on the remote sensing of community-relevant landfast ice polynyas, it was necessary to first build and characterize datasets of polynya observations in SAR imagery to enable the development and evaluation of detection models. This chapter describes the synthesis and characterization of two such datasets. Section 3.1 describes the regions of interest (RoIs), imagery, and labels that comprise each dataset; section 3.2 investigates the separability characteristics of polynyas from landfast ice in the datasets' dual-polarized SAR imagery; and section 3.3 explores the influences of imaging and environmental factors on polynya signature variability.

3.1 Dataset Synthesis

The development of landfast ice polynya datasets described subsequently was carried out in close collaboration with the Arctic Eider Society (AES, https://arcticeider.com/) and SmartICE (https://smartice.org/), organizations with established connections to the communities of Sanikiluaq, Nunavut, and Nain, Nunatsiavut (NL). A methodology for mapping sea ice hazards, part of a broader initiative of Inuit-driven IQ mobilization, e.g. [141, 142], was developed by SmartICE and AES in consultation with the author regarding requirements for machine learning. Mapping of Nain polynyas was carried out by Inuit SmartICE operators in Nain, while mapping of Sanikiluaq polynyas was carried out by the author. Polynyas mapped around Sanikiluaq were validated with local ice experts during a visit to Sanikiluaq in January 2024.

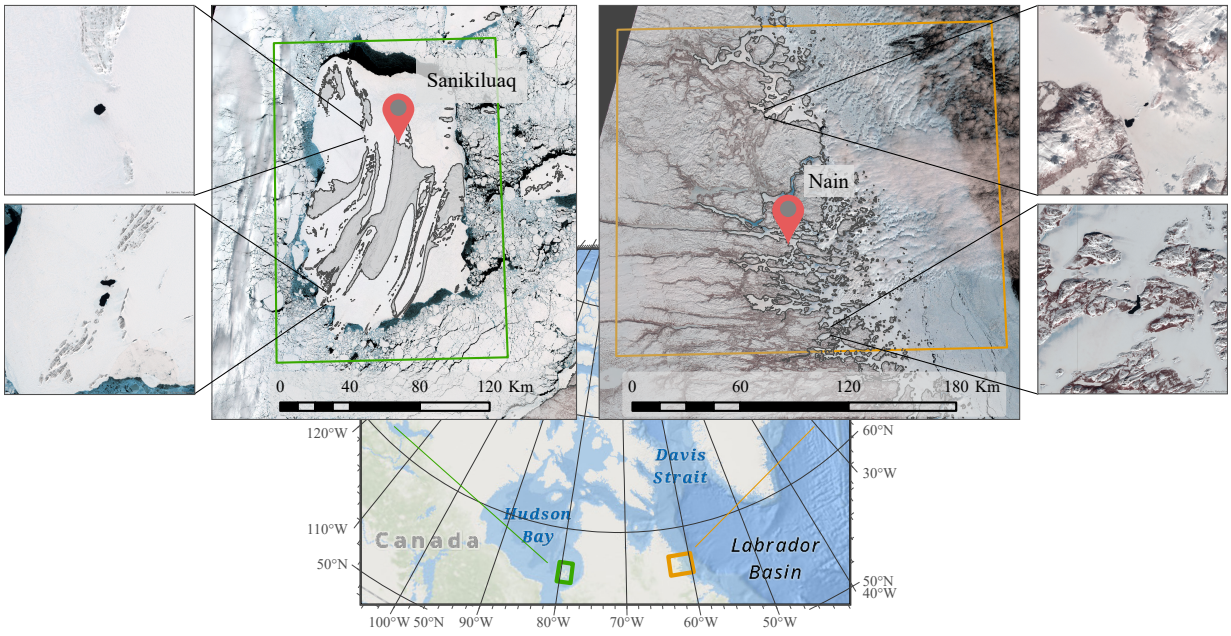

Figure 3.1: Map showing study areas encompassing landfast ice around the communities of Sanikiluaq, NU, and Nain, NL. Near-infrared, green and blue bands of Sentinel-2 imagery captured April 2, 2021 and April 16, 2021, respectively, are shown (Copernicus Sentinel data 2021). These areas capture common travel routes used in each community.

3.1.1 Regions of Interest

Sanikiluaq, NU, is located on the Belcher Islands in Hudson Bay, while Nain is located in Nunatsiavut (NL) on the east coast of Canada, as illustrated in Fig. 3.1. Both communities are primarily Inuit by population, and travel over landfast ice by snow machine for hunting, leisure, and environmental monitoring is common [15, 9, 63]. For both communities, landfast ice typically forms in December/January and breaks up in May/June, and contains sensible-heat polynyas due to the presence of many small islands which cause localized areas of high current. These polynyas typically persist throughout the winter season and recur in similar locations from year to year, but may open or close irregularly when the thermodynamic balance between warm, tidally-forced seawater and sub-freezing air temperatures is perturbed. Both regions are at a low latitude relative to other Northern communities and thus have more winter sunlight, providing greater optical satellite data coverage that can be leveraged for polynya mapping.

3.1.2 Imagery

Dual-polarized Sentinel-1 (S1) SAR imagery was obtained via Google Earth Engine (GEE) [48] for the 2016 through 2022 winter and spring seasons over Sanikiluaq, and the 2020-2022 seasons over Nain. For consistency and based on data availability defined by the S1 observation scenario over the RoIs [30], EW-mode GRDM images having HH+HV polarization were downloaded. This imagery has a nominal spatial resolution of 93 x 87 meters (range x azimuth) and a pixel spacing of 40 meters [32]. Processing algorithms applied to S1 backscatter images in the GEE catalogue include additive noise removal, radiometric calibration to σ^0, orthorectification using the SRTM 30m DEM, and decibel scaling [47]. Incidence angle values for each pixel are included alongside backscatter channels in the imagery. No speckle filtering algorithms were applied, as CNNs have proven capable of isolating speckle in SAR imagery [21, 120] and are thus positioned to learn the most effective speckle reduction strategies to perform well on the detection task represented by the training labels. All S1 data was stored in 32-bit floating point format.

Sentinel-2 (S2) multispectral imagery was also obtained from GEE. B4, B3, B2 (RGB, respectively) and B8 (near-infrared) bands from level-2A S2 image tiles with less than 60% cloud mask coverage were downloaded. Each of these bands has 10m resolution and pixel spacing. All S1 and S2 images were reprojected to UTM zones 17N and 21N for Sanikiluaq and Nain datasets, respectively.

3.1.3 Land and Landfast Ice Delimitation

Weekly regional ice charts from the Canadian Ice Service (CIS) Ice Archive (https://iceweb1.cis.ec.gc.ca) were used to delimit the approximate landfast ice extent within the RoI on a weekly basis. To more precisely delimit land areas, a landmask was derived from the GSHHG version 2.3.7 coastlines dataset [138]. Morphological closing [127] was applied to fill small gaps (up to 400m) between the landmask and landfast ice areas delimited by the coarser-resolution ice charts.

3.1.4 Polynya Mapping

In both datasets, polynya segmentation labels were generated based on visual inspection of S1 and S2 imagery. Sanikiluaq polynyas were mapped from January 22 to May 14 across seven consecutive seasons (2016-2022), and Nain polynyas were mapped from March 19 to May 14 across three consecutive seasons (2020-2022).

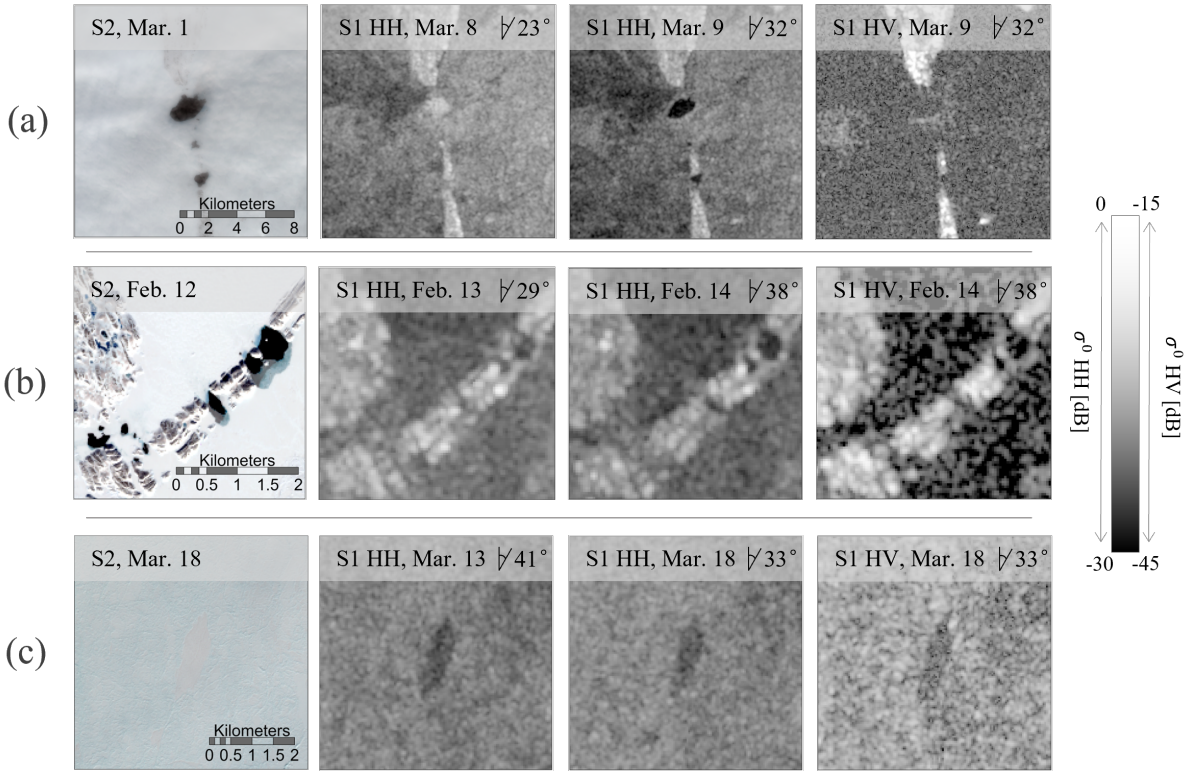

Figure 3.2: Examples of polynya and landfast ice appearance in Sentinel-2 optical and Sentinel-1 dual-polarized SAR imagery (σ^0). (a) & (b) indicate dependence of polynya backscatter intensity on incidence angle in HH polarization, and low backscatter intensity in HV polarization over landfast ice and open water regions largely dominated by thermal noise. (c) demonstrates an area of smooth ice of similar scale and appearance in S1 backscatter to polynya signatures.

Given the high volume of imagery, persistent nature of polynyas, and inconsistent availability of low-cloud optical imagery, mapping was carried out on a per-week, rather than per-image, basis. In other words, one set of polynya labels was generated per week based on all available imagery for that week. When available, low-cloud S2 imagery provided unambiguous polynya identification and delimitation (illustrated in Fig. 3.2). It is noted that thin new ice can appear transparent and thus be difficult to identify in spaceborne optical imagery such as S2. While a band combination using B8 (near-infrared, 842nm) rather than B4 (red, 665nm) was found to better highlight thin ice in some cases, the presence of thin ice cover within polynyas was not catalogued during mapping. When no low-cloud S2

imagery was available for a given week and location, polynya presence was inferred from a combination of known polynya locations, signature in S1 images for that week, and any available low-cloud S2 imagery from the previous week. To reflect potential ambiguities in S1 backscatter around suspected polynya locations, a discrete confidence rating from 1 to 3 was assigned to each polynya label. Only those polynyas mapped with a confidence score of 1, indicating the observation of a clear polynya signature in S2 imagery for that week or a probable signature in S1 imagery at a recurring polynya location, were considered as targets for training and evaluation of detection models (data volume implications are given in table 3.1).

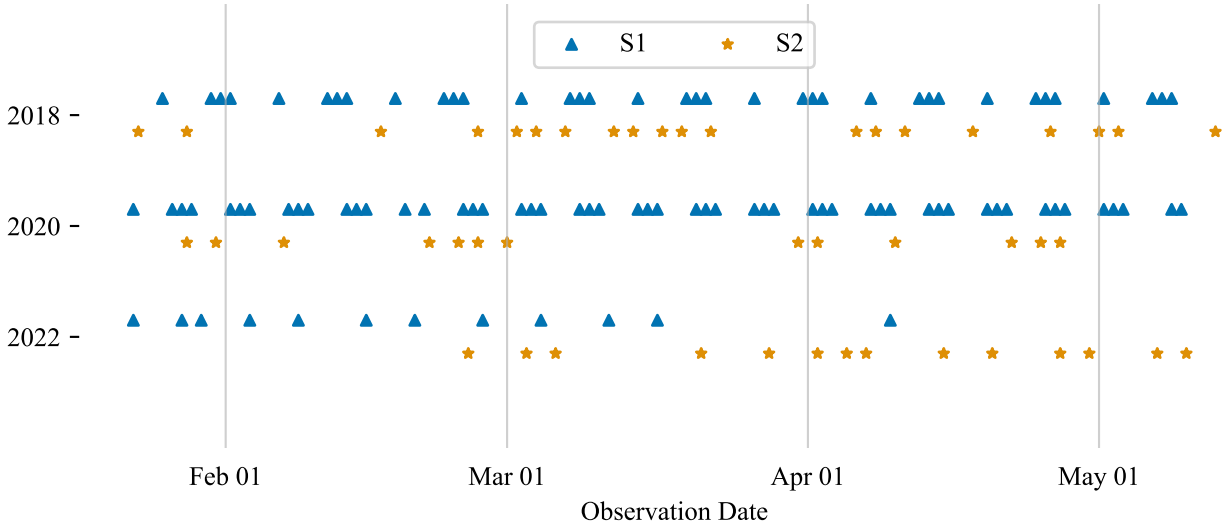

Figure 3.3: Example of Sentinel-1 (S1) and low-cloud Sentinel-2 (S2) image availability for a single recurring polynya area near Sanikiluaq (55°55'15"N, 80°03'30"W) over three years. S2 coverage frequency varies between years, and is typically much lower than S1. A reduction in S1 volume in 2022 reflects the loss of the Sentinel-1B satellite in Dec. 2021.

Fig. 3.3 demonstrates the relative frequencies of S1 and low-cloud S2 observations for a single recurring polynya location within the Sanikiluaq RoI over the course of three different seasons. Given observed and previously studied challenges in spectral reflectance-based cloud masking over snow-covered regions [130, 65], an approximate definition for "low-cloud" in this analysis was taken as the total coverage of S2 tiles having less than 60% cloudy-pixel percentage based on the QA60 cloud mask [31].

All mapping was carried out in ArcGIS Pro [29], and mapped polygons were exported

Table 3.1: Number of S1 polynya observations in Sanikiluaq and Nain datasets before and after confidence and size filtering

Target criteria	**Sanikiluaq** (01/22 - 05/14, 2016 - 2022)	**Nain** (03/19 - 05/14, 2020 - 2022)
All	8505	438
Confidence-1	7741	339
Area $\geq$ 100m radius-equivalent	5089	306

and saved in shapefile format. Each polygon includes metadata on confidence score, the specific imagery used to delimit the polynya, and the spatial area of the polygon.

3.1.5 Dataset Characteristics

Data Volume and Polynya Target Filters

For the Sanikiluaq dataset, 305 S1 EW SAR images were downloaded and polynya mapping produced 3,484 per-week polynya labels with 8,505 corresponding S1 observations. For the Nain dataset, 18 S1 EW images were downloaded and mapping produced 391 polynya labels with 438 S1 observations[1].

As shown in Fig. 3.4, many mapped polynyas were extremely small and infeasible to detect in 40m S1 EW pixels. An areal extent of 31,415m^2, corresponding to a circular target with a 100m radius, i.e., a 5x5-pixel bounding box in 40m imagery, was chosen as a minimum target size threshold. This is referred to hereafter as the "100m radius-equivalent" area. In addition to polynyas mapped with low confidence, polynyas below the 100m radius-equivalent size threshold were flagged to be excluded from model training and evaluation. After confidence and size filtering, the Sanikiluaq dataset contained 5,089 S1 polynya observations and the Nain dataset contained 306 S1 polynya observations.

Size Distributions of Mapped Polynyas

The distributions of mapped confidence-1 polynya areas derived from segmentation labels and corresponding bounding box long-edge lengths in 40m S1 imagery are shown in Fig.

[1]The notably lower imagery volume in the Nain dataset compared to Sanikiluaq is due to the use of fewer years, fewer weeks per year, and different S1 acquisition scenarios between regions

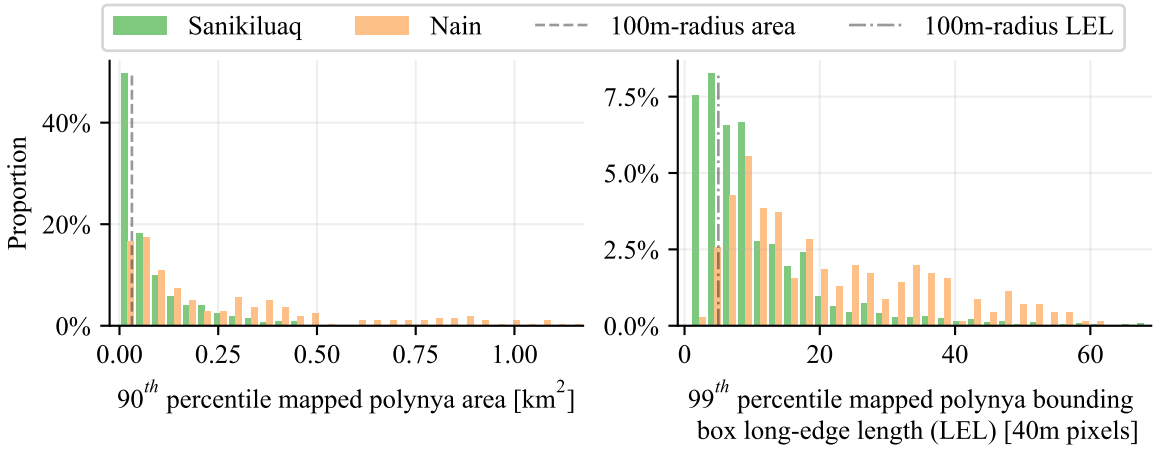

Figure 3.4: Distribution of mapped polynya sizes for Sanikiluaq and Nain datasets. Only polynyas with a mapped confidence score of 1 (high confidence) are included.

3.4. The Sanikiluaq region is seen to contain smaller polynyas on average than the Nain region, however polynyas in both regions are clearly on the scale of tens to hundreds of meters in extent. For the object detection task in 40m SAR imagery, most targets are well below the commonly used 30x30-pixel definition of "small objects". Indeed, the median bounding box long-edge length for targets *above* the 5-pixel size threshold is 11 pixels for the Sanikiluaq dataset and 18 pixels for the Nain dataset.

3.2 Polynya Separability in Dual-Polarized SAR

3.2.1 Methods

The backscatter signatures of polynya areas, as mapped on a weekly basis, versus landfast ice, as delimited by CIS charts and a higher-resolution landmask, were characterized on both full-scene and polynya-local scales in each S1 SAR image. These different scales of analysis were carried out to illustrate potential differences in the difficulties of isolating polynyas from all landfast ice in a large region versus delimiting polynyas from immediately surrounding landfast ice. The full-scene analysis included a large (n=50M) sample of all

landfast ice pixels in all S1 images across seasons[2], while the local-scale analysis pulled landfast ice pixels only from dynamically-sized square windows centered on each mapped polynya. For a polynya having a bounding box long-edge length L, a square window was drawn with dimensions $2L$. Box-plot distributions and the Jeffries-Matusita distance were used to illustrate the degree of separability between polynya and landfast ice classes. Given the known dependence of dual-, and in particular, HH-polarized backscattering intensity on incidence angle for sea ice and water classes [91], separability analyses were carried out by incidence angle in bins of 4°.

Jeffries-Matusita Distance

The Jeffries-Matusita (JM) distance is a widely-used index for quantifying separability between distributions, e.g. [23], yielding values from 0 (highly non-separable) to 2 (perfectly separable). It is defined as

$$J_{ij} = 2(1 - e^{-d_{ij}}), \tag{3.1}$$

where d_{ij} is the Bhattacharyya distance between distributions of, in this case, the measured backscatter intensity values (σ^0 image pixels), x, of the polynya and landfast ice classes. The non-parametric Bhattacharyya measure as described in [70] was used, implemented as in equation 3.2 where $p_c(x)$ represents per-class distributions of backscatter values. This approach was motivated by the observed occurrence of bimodal σ^0 distributions within some incidence angle bins, which are not well represented by parameterized unimodal distributions (e.g. normal [42] or Wishart [23]) as have been used in other studies.

$$d_{ij} = -\ln\left(\sum_x \sqrt{p_{ice}(x)p_{polynya}(x)}\right) \tag{3.2}$$

Contrast-Based Polynya Separability Metric

Given the observed variability of polynya backscatter signatures, a *contrast-of-backscatter-means* metric (equations 3.3 and 3.4) was developed as a simple representation of local polynya separability in backscatter intensity. This approach was used in place of a distribution-based separability measure such as JM distance due to the limited number of

[2]This sampling was motivated by computational challenges in simultaneously processing 305 full-scene SAR images, i.e. $\sim$ 2 bn landfast ice pixels

Table 3.2: Number of S1 polynya observations stratified by varying thresholds on local contrast-of-backscatter-means metric, δ^{μ}_{HH} (eq. 3.3).

Target criteria	**Sanikiluaq**	**Nain**
$\delta^{\mu}_{HH} \geq$ +5dB	81	4
$\delta^{\mu}_{HH} \geq$ +3dB	322	10
$\delta^{\mu}_{HH} \geq$ +1dB	1066	37
No contrast criteria	5089	306
$\delta^{\mu}_{HH} \leq$ -1dB	2723	216
$\delta^{\mu}_{HH} \leq$ -3dB	1704	151
$\delta^{\mu}_{HH} \leq$ -5dB	949	112

pixels available on a per-polynya basis and the associated sensitivity of variance estimates. The metric is computed on a per-window basis, with each window centered on a mapped polynya. A dynamic window size of twice the polynya's bounding box long-edge length was used, consistent with the distribution plotted in Fig. 3.4. It was observed that contrast decreased with larger relative window sizes (up to 4x the polynya bounding box long-edge length was tested), possibly due to rough ice formations causing higher backscatter intensities near polynya edges and thereby increasing contrast with low-backscatter polynya open water areas. Mapped polynya polygons were used to delimit pixels contributing to $\mu_{\sigma^0\ polynya}$, while land and landfast ice masks were used to isolate landfast ice pixels contributing to $\mu_{\sigma^0\ ice}$. This metric offers a proxy for target difficulty (i.e., polynyas with higher absolute contrast are more visually salient and statistically separable than those with low absolute contrast) of use in subsequent detection experiments. Table 3.2 illustrates the distribution of S1 polynya observations in both datasets by measured backscatter contrast in HH, while Fig. 3.5 demonstrates the appearance of polynyas with different local contrast levels.

$$\delta^{\mu}_{HH} = \mu_{\sigma^0_{HH}\ polynya} - \mu_{\sigma^0_{HH}\ local\ ice} \tag{3.3}$$

$$\delta^{\mu}_{HV} = \mu_{\sigma^0_{HV}\ polynya} - \mu_{\sigma^0_{HV}\ local\ ice} \tag{3.4}$$

3.2.2 Results

In aggregate across all polynyas mapped with high confidence and having spatial areas greater than the 100m radius-equivalent, low overall separability from landfast ice in HH-

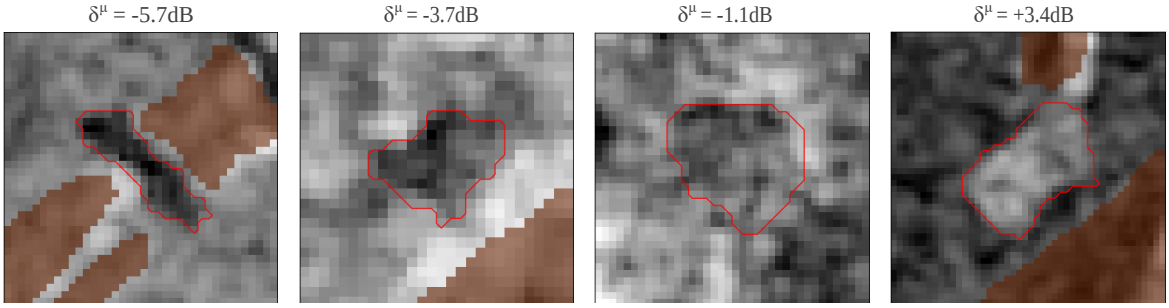

Figure 3.5: Examples of HH-polarized S1 observations with different δ^{μ}_{HH} values (eq. 3.3). Viewing window dimensions correspond to those used for calculation of δ^{μ}, i.e., twice the mapped polynya's bounding box long-edge length. Land areas are masked in brown.

Table 3.3: Number of observations and mean contrast metrics (eqs. 3.3 & 3.4) by incidence angle (IA). Standard deviations of contrast metrics are given in parentheses.

	IA Threshold	**N. Observations**	**Mean δ^{μ}_{HH} [dB]**	**Mean δ^{μ}_{HV} [dB]**
Sanikiluaq	None	5089 (100%)	-1.75 (3.39)	-1.87 (2.84)
	$\geq 25°$	3670 (72.1%)	-2.53 (3.10)	-2.28 (3.12)
	$\geq 35°$	1965 (38.6%)	-3.05 (3.22)	-2.41 (3.12)
Nain	None	306 (100%)	-3.21 (3.37)	-2.78 (2.88)
	$\geq 25°$	306 (100%)	-3.21 (3.37)	-2.78 (2.88)
	$\geq 35°$	168 (54.9%)	-3.77 (3.30)	-2.79 (3.00)

and HV-polarized backscatter intensity was found for both Sanikiluaq (n=5089 polynya observations) and Nain (n=306 polynya observations) datasets. JM distances computed by incidence angle bins of 4° were consistently below 0.5, while the mean HH-polarized local contrast metric (eq. 3.3) was -1.75 dB for Sanikiluaq and -3.21 dB for Nain (table 3.3). Separability patterns related to incidence angle and landfast ice pixel population were nonetheless observed.

Considering first the intensity-based separability of polynyas in the Sanikiluaq dataset (Fig. 3.6a), separability in HH increased with incidence angle over the medium range ($\sim$ 29° to 37°), while observations taken at the lowest incidence angles ($\sim$ 19° to 23°) showed slightly higher separability than those at the low-to-medium range ($\sim$23° to 29°). Simultaneously, separability from *nearby* landfast ice was greater than that from *all* landfast ice in the RoI for incidence angles above 25°, following observations made during polynya mapping that landfast ice backscatter tends to increase near polynyas. This may be due to

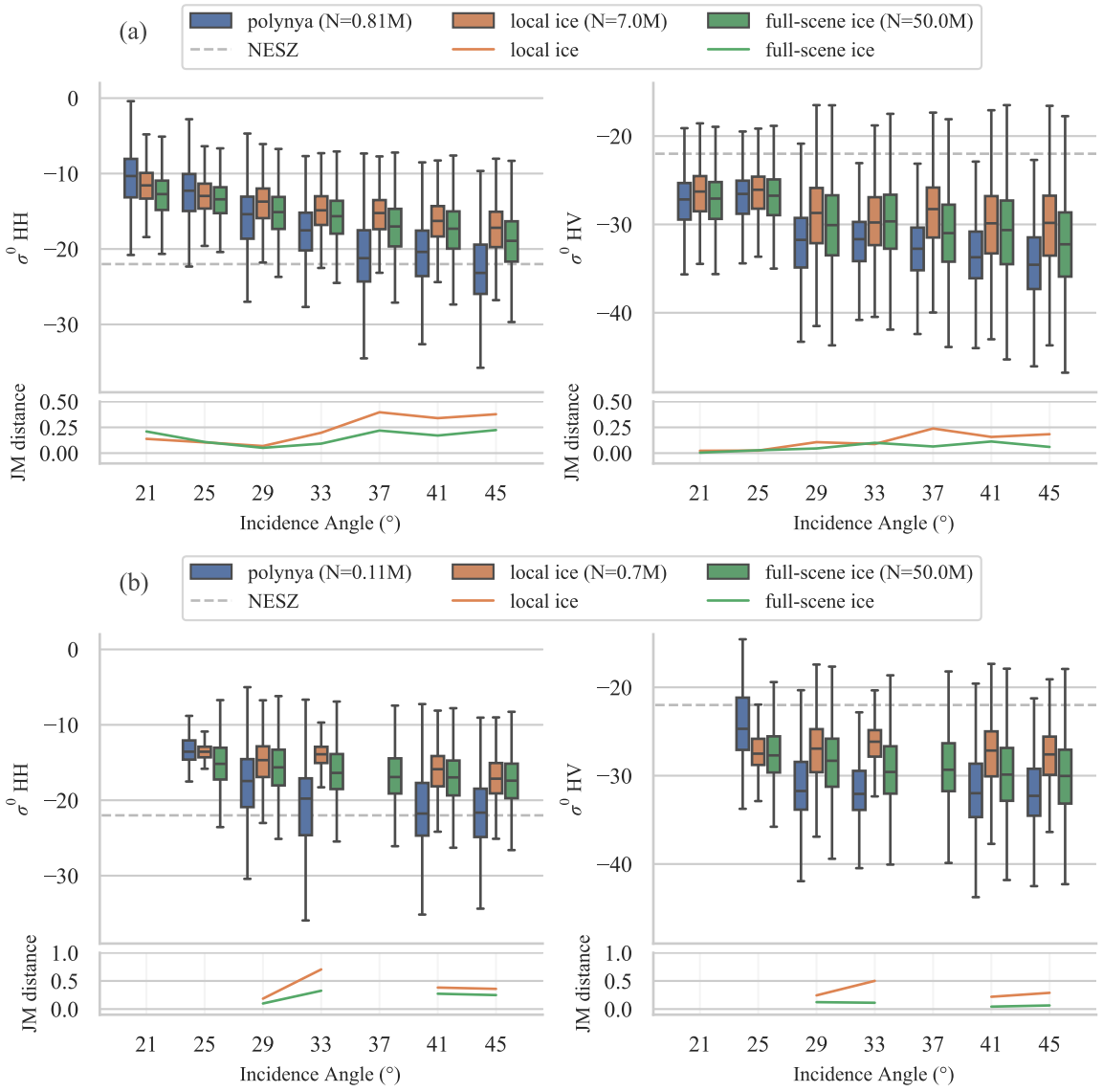

Figure 3.6: Backscatter intensity distributions in HH- and HV- polarized σ^0 of polynya pixels (left group) vs. polynya-local (center group) and large samples of all (right group) landfast ice pixels in the Sanikiluaq (a) and Nain (b) datasets. Box-plot whiskers indicate 1.5x the interquartile range. Jeffries-Matusita (JM) distances between polynyas and the two ice populations are calculated and plotted by the same incidence angle bins as those used for box-plotting.

deformed ice features associated with polynya forcing mechanisms, such as current-driven piling or fracture. This "contextual" pattern, i.e., the presence of a polynya conditionally influencing ice class backscatter distributions, highlights the advantage of spatial classifiers over pixel-based classifiers for polynya detection. Separability in HV was very low (JM distances below 0.25) for all incidence angles and ice pixel populations, commensurate with the observation that backscattering intensities were below the NESZ threshold of -22dB.

Polynya separability in backscatter intensity was observed to be somewhat higher for the Nain dataset than that for the Sanikiluaq dataset. This higher separability for Nain polynyas is possibly related to their larger size relative to Sanikiluaq polynyas, allowing them to be better resolved by the medium-resolution S1 imagery. Nonetheless, the separability of Nain polynyas was also found to be low, with JM distances consistently below 1. It is also noted that the significantly smaller number of Nain polynya observations introduce more uncertainty into the analysis of this region than Sanikiluaq.

Overall, these results emphasize the challenging signatures of landfast ice polynyas in dual-polarized backscatter intensity, limiting the potential of detection approaches using only intensity threshold-based classifiers previously applied in other sea ice / water classification contexts (e.g. [13, 40]).

3.3 Imaging & Environmental Controls on Polynya Separability

As discussed in section 2.3.3, backscattering intensity from sea ice and water is known to be sensitive to several imaging and geophysical parameters relevant to the synthesized polynya datasets. Key sensitivities include those to 1) incidence angle, 2) wind roughening of open water, 3) potential thin ice growth within polynyas, and 4) snow cover wettening in response to warming (e.g. towards the onset of spring melt). Motivated by these known sensitivities and the observed variability of polynya S1 signatures in the dataset, an exploratory analysis was carried out to investigate the degree to which signature variability could be explained by these imaging and environmental parameters.

3.3.1 Methods

An investigation into the influence of the aforementioned imaging and environmental parameters on the local backscatter contrast measures was carried out on the Sanikiluaq

dataset, given its higher volume of imagery and diversity of unique polynyas relative to the Nain dataset. A significant challenge in this endeavour, however, is the shortage of operational weather stations around the Belcher islands and resultant data sparsity with which to conduct rigorous correlative analyses. An exploratory foray was nonetheless made with regional-scale ERA5 reanalysis data provided by ECMWF [57] for the possibility that coarse patterns in some of these regional data might provide an initial outlook on controls for the observed variability in polynya signatures in the developed dataset.

Regression analyses of backscatter contrast versus incidence angle, instantaneous windspeed, and average temperature over the 12 hours preceding imaging time were performed, alongside a visual analysis of backscatter contrast by time of year (a nonlinear relationship was expected). Hourly 10-meter horizontal wind and 2-meter temperature variables from ERA5, distributed on a 0.25° latitude-longitude grid ($\sim$ 28km lat. by 15km lon. over the Belcher Islands), were used for respective analyses. In each case, the model variable was interpolated to each SAR polynya observation in space and time, and matched with the polynya's contrast-of-means metrics as a measure of polynya separability from surrounding landfast ice.

3.3.2 Results

Table 3.4: Coefficients of determination (R^2) for imaging and ERA5 environmental variables on backscatter contrasts δ^{μ}_{HH} and δ^{μ}_{HV} (eqs. 3.3 and 3.4) of mapped polynyas areas vs. surrounding landfast ice.

Parameter	R^2 **for HH Contrast** (δ^{μ}_{HH})	R^2 **for HV Contrast** (δ^{μ}_{HV})
Incidence Angle	0.183	0.0708
Instantaneous Windspeed	0.117	0.0570
Mean 12h 2m Temperature	0.0176	0.0611
IA + Windspeed	0.297	0.122
IA + Windspeed + Temp	0.324	0.199

Fig. 3.7 and table 3.4 demonstrate the sensitivity of HH backscatter intensity within mapped polynya areas to incidence angle and windspeed, while HV is less affected. Consistent with expectations, polynyas show greater contrast with respect to surrounding first-year landfast at high incidence angles and low windspeeds. The step-change in local contrast occurring in May reflects an increase in backscattering intensity over landfast ice

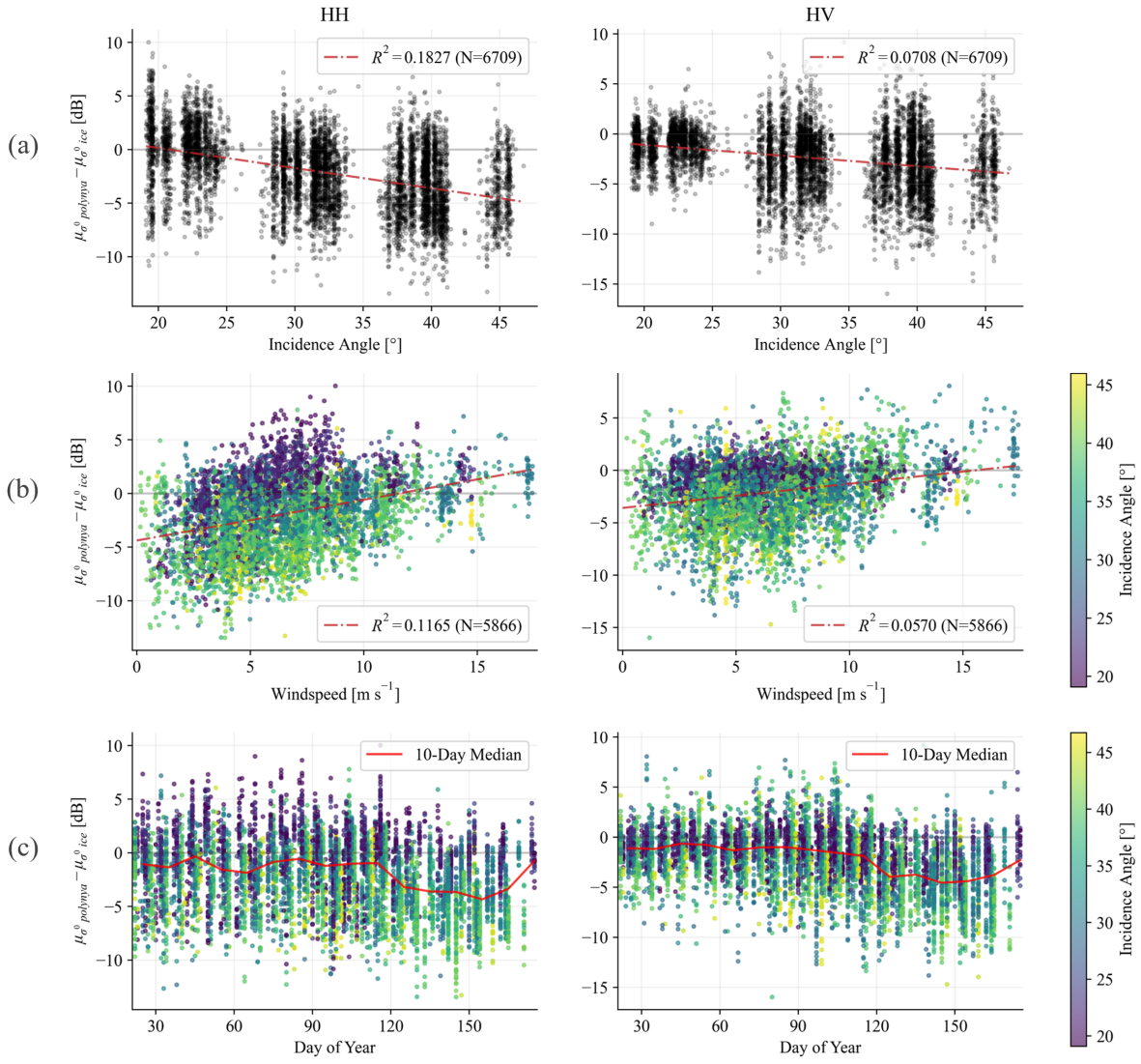

Figure 3.7: Investigating controls of imaging and environmental parameters on contrast between polynyas and surrounding landfast ice backscatter intensity. Given the strong variance with incidence angle (a), contrast observations by windspeed (b) and day-of-year (c) are coloured by IA to illustrate the combined controls on contrast.

areas, generally corresponding with early melt onset timing for the region. This pattern is consistent with previous findings related to the sensitivity of C-band σ^0 to warming snow cover parameters such as brine volume, e.g. [104].

These findings, while based on coarse regional-scale data and clearly not perfect predictors of polynya backscattering contrast, do in combination explain 30% of the variability in local polynya contrast. They provide evidence supporting the expectation that incidence angle, wind roughening, and seasonality are indeed controls on the separability of polynyas from surrounding landfast ice particularly in HH-polarized SAR. A more detailed characterization of these factors would require higher-resolution measurements of polynya-local environmental parameters coincident with SAR image acquisitions, which was out of the scope of this thesis research.

3.4 Chapter Conclusions

The datasets of mapped landfast-ice polynyas presented in this chapter have been shown to contain small, sparse targets with variable separability from landfast ice in dual-polarized C-band SAR backscatter intensity. This separability was shown to be too low for the effective use of pixel-based classifiers, with incidence angle-binned Jeffries-Matusita distances between polynya and landfast ice classes consistently below 0.5. An exploratory analysis of controls on polynya signature variability found that HH backscatter intensity contrast between polynyas and landfast ice increased with incidence angle ($R^2 = 0.18$), decreased with model windspeeds ($R^2 = 0.12$), and underwent a rapid increase near early melt onset. The HV-polarized backscatter intensity over landfast ice and open water areas was found to lie below the sensor noise floor. Polynya datasets were stratified by a separability heuristic based on local backscatter intensity, yielding groups of labeled targets with varying difficulties for retrieval method development and assessment.

Chapter 4

Object Detection Networks for Polynya Detection in Sentinel-1 SAR

In chapter 3, the problem of detecting landfast ice polynyas in dual-polarized SAR imagery is shown to be one of identifying small, sparse targets with varying and frequently low separability in backscatter intensity. This frequently low separability limits the expected performance of pixel-based classifiers that have been used in other SAR-based sea ice classification contexts [146, 91]. As discussed in chapter 2, textural features of SAR images such as spatial autocorrelation or GLCM have been used to differentiate between sea ice and open water, however their efficacy in doing so generally depends on the use of spatial windows larger than the extent of most landfast ice polynyas considered in this research (Fig. 3.4). CNNs, having previously shown efficacy for SAR-based sea ice classification, are well positioned to exploit the large volume of training data generated for this study to capture heterogeneous but distinct intensity-level, textural, geometric and/or contextual features of polynyas through supervised learning. In particular, an object detection formulation presents a desirable set of mechanisms to handle target sparsity while capturing the inherent geometric compactness of polynyas.

The overall objective of research described in this chapter is the development and evaluation of a convolutional object detection network for landfast ice polynya detection in Sentinel-1 SAR imagery using the dataset described in chapter 3. This is broken down into three sub-objectives:

1. Guided by advancements in the natural-scene small object detection field, optimize polynya detection model performance by investigating key design elements related to feature representation capacity, input spatial resolution, and training loss balancing;

2. Evaluate the developed model under a range of data scenarios to characterize input feature importance, the influence of data augmentation for regional generalizability, and the capacity of the model to detect targets with varying levels of backscatter separability;

3. Identify model capabilities and limitations from predictions on held-out evaluation datasets.

These investigations are presented in sections 4.3, 4.4, and 4.5, respectively, for which sections 4.1 and 4.2 lay important methodological groundwork.

4.1 Adapting Faster-RCNN to a Polynya Detection Context

Chapter 2 motivated the use of the Faster-RCNN (FRCNN) [113] object detection network for landfast ice polynya detection based on its demonstrated success and continued use as a baseline model in small object detection for natural-scene and optical remote sensing contexts. Nonetheless, several characteristics of the landfast ice polynya detection problem diverge from the natural-scene domains for which object detectors like FRCNN were developed, limiting the "off-the-shelf" applicability of such models to this task. These characteristics include the following:

Imaging modality - While model parameters for detection networks pre-trained on natural-scene benchmark datasets are widely available in the present deep learning ecosystem from platforms like pytorch [106] or MMDetection [18], the use of such pre-trained networks for SAR image analysis has met limited success in previous studies due to the significantly different viewing geometries and imaging modalities between natural-scene RGB and spaceborne SAR images [62].

Target characteristics - Polynya signatures in dual-polarized SAR imagery are, as demonstrated in chapter 3, heterogeneous in backscatter intensity, extremely small, and sparsely distributed within landfast ice. Despite these challenges, the set of features involved in detecting polynyas in SAR imagery is likely to be less complex than that needed to detect densely distributed, frequently occluded targets with tens to hundreds of unique class labels in a multitude of settings, as is often the challenge in natural-scene contexts [86]. Specifically, detection models with fewer parameters than widely-available natural-scene benchmark models are likely appropriate for polynya detection.

Areas and targets of interest - Landfast ice polynya detection models should ignore areas of open water beyond the landfast ice edge, areas of continuous land, and mapped polynyas that do not meet the confidence, size, or target separability criteria described in chapter 3. Such selective filtering should be handled within the model, rather than solely as a post-processing step, to ensure all boxes contributing to training loss correspond either to polynyas meeting all target criteria or landfast ice.

Preliminary modifications to adapt FRCNN to the preceding characteristics of the polynya detection task are described in this section, representing a starting point used for the experiments described in sections 4.3 and 4.4. An implementation of FRCNN in the pytorch v2.1 deep learning framework [106] (pytorch.org) was modified following the design elements described subsequently.

4.1.1 Backbone Network

The EfficientNet-B0 CNN was used as a backbone feature extraction network in the Faster-RCNN detector implemented in this study. The EfficientNet family of networks [133] is based on a Neural Architecture Search jointly optimizing model accuracy and floating point operations per second, leveraging inverted bottleneck [109] and squeeze-excitation [60] mechanisms for their representational efficiency and ability to model inter-channel dependencies, respectively. The EfficientNet models have demonstrated parameter efficiency in learning representative features on benchmark computer vision datasets, e.g. achieving better top-1 accuracy on the ImageNet dataset than the commonly used ResNet-50 CNN with 4.9x fewer learnable parameters [133]. Following the hypothesis that binary polynya / ice classification from dual-polarized SAR can be achieved from a relatively lightweight set of features, the smallest version of the EfficientNet networks — EfficientNet-B0, detailed in table 4.1 — was selected for this study.

4.1.2 Faster-RCNN Hyperparameterization

The hyperparameters of FRCNN detection models designed to handle tens or hundreds of densely-distributed object classes in natural-scene images typically employ classification sub-networks with thousands of neurons predicting hundreds or thousands of proposal boxes per image [44]. Given the sparse, two-class nature of the polynya detection problem, several hyperparameters were modified to simplify the network and better adapt it to this context. A complete list of FRCNN hyperparameters and the values used in the adapted model is given in appendix A, reflecting the following modifications in each sub-network.

Table 4.1: Summary of the full EfficientNet-B0 feature extraction ("backbone") network used in this research, as designed in [133]. MBConv_X units are the building blocks of the network, consisting of an inverted bottleneck with channel expansion ratio X and a squeeze-excitation operation. Each stage may repeat the same block a number of times. S_0 represents the original dimension of a square input image, e.g. 128x128 pixels.

Stage index	**Output feature size**	**# Output channels**	**# Parameters**
S-1	$S_0/2$	16	0.15M
S-2	$S_0/4$	24	0.17M
S-3	$S_0/8$	40	0.22M
S-5	$S_0/16$	112	1.0M
S-7	$S_0/32$	320	3.8M

Region Proposal Network (RPN) - Modifications to four aspects of the RPN sub-network were made. First, anchor box sizes were reduced based on the distribution of target polynya sizes in the synthesized datasets (Fig. 3.4). Second, the IoU thresholds defining positive and negative anchor boxes were both reduced. A reduction in the positive threshold, defining the minimum IoU between an anchor box and a ground-truth box for the anchor to be considered a positive example, allows more anchors to be classified as positive even if they do not perfectly match the labelled polynya's bounding box. A reduction in the negative threshold, defining the maximum IoU between an anchor box and a ground-truth box for the anchor to be considered a negative example (i.e., the landfast ice class), ensures that boxes partially overlapping a polynya are *not* considered negative. These modifications were implemented to place more emphasis on the *detection* rather than the *precise localization* of polynyas and to reduce ambiguity in the loss signal for partially-covered targets. Third, the number of top-scoring boxes passed from the RPN to the RoI Head was reduced from 2000 during training and 1000 during inference to 100 for both training and inference. This reduction reflects the prior knowledge that, given their sparse distribution, only a handful of polynyas will likely ever exist in a single SAR image patch (discussed in section 4.2.4). Finally, the number of boxes sampled for objectness loss computation was reduced from 256 to 64 to reduce the expected imbalance of positive versus negative anchor boxes during training — this hyperparameter is explored further in section 4.3.1.

RoI Pooling - The RoI pooling layer creates fixed-size feature vectors for each RPN object proposal by sampling the backbone-produced feature maps from a grid of fixed dimensions placed within the bounds of the variable-sized object proposals. Given the

smaller polynya target and corresponding RPN anchor box sizes relative to typical natural-scene targets, the pooling grid dimensions were reduced from 7x7 to 3x3. In addition to reducing probable redundancy from over-sampling the low-resolution features at deeper backbone layers, this modification drastically reduces the number of parameters in the fully-connected layers of the RoI Head network.

RoI Head - Modifications to three aspects of the RoI Head sub-network were made. First, the number of neurons in each fully-connected layer was reduced from 1024 to 128. This reflects the lower complexity of binary polynya/ice classification versus multi-class natural-scene targets, and also greatly reduces the number of parameters in the RoI Head sub-network. Second, the IoU threshold defining negative boxes was reduced from 0.5 to 0.3 with similar rationale to the RPN, however was kept higher than 0.1 to penalize low-precision detections more heavily than in the RPN. Finally, the number of boxes sampled for classification loss computation was reduced from 512 to 32 to again reduce the expected imbalance of positive versus negative examples. This was informed by the expected number of polynya targets within each polynya-containing image patch (on the order of 1-10) and the role of the RoI Head network in *refining* top proposals from the RPN rather than filtering out thousands of easy negatives, motivating the use of a more balanced training set. It is noted that for the standard FRCNN geared towards natural-scene targets, a larger number of loss samples is taken for classification in the RoI Head (512) than in the RPN (256). This is because the former makes multi-class predictions while the latter makes binary ("objectness") predictions. In the polynya detection context both networks make binary (polynya/ice) predictions, so a smaller loss sample is appropriate.

In summary, these hyperparameter modifications loosen the overlap-based definition of positive boxes, tighten the definition of negative boxes, and reduce the number of boxes forwarded by the RPN, the number of boxes sampled for loss computation, and the number of learnable parameters in the RoI Head sub-network. The resultant balancing of learnable parameter densities throughout the network is illustrated by Fig. 4.1. The distribution of parameters in the adapted network appears much better suited to the polynya detection context: the backbone network has the most parameters to learn discriminative polynya features, the RPN has the second most to determine effective decision boundaries between rare polynya targets and heterogeneous landfast ice, and the RoI Head has the fewest to refine RPN proposals. Indeed, evaluation results increased considerably with the adapted hyperparameterization over the default hyperparameterization, as reported in table 4.6.

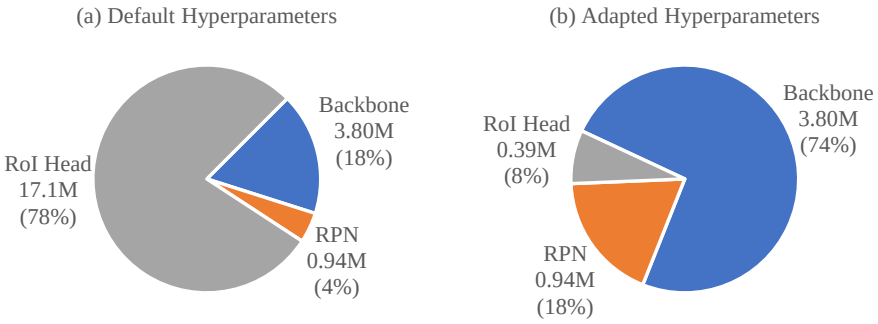

Figure 4.1: Comparison of FRCNN learnable parameter densities with (a) the default FRCNN hyperparameters [113] and (b) the adapted hyperparameters described in section 4.1.2. The number of learnable parameters for each sub-network are given in millions and their relative proportion, based on a 7-stage backbone EfficientNet-B0 network with 320 feature channels in the final layer (S-7 of table 4.1).

4.1.3 Modifications to Handle Landfast Ice Areas & Selective Targets

The pytorch v2.1 [106] implementations of RPN and RoI Head sub-networks were extended to handle two requirements particular to the polynya detection context. Firstly, anchor and proposal boxes that occurred beyond the landfast ice floe edge (as delimited by regional ice charts) or that were centered on land areas (as delimited by the GSHHG [138]-derived landmask) were discarded during model inference and training. Secondly, during training, negative boxes that covered "non-target" polynyas (mapped polynyas not meeting one or more target criteria) were discarded and excluded from loss computation.

These modifications allowed for selective control of the negative and positive boxes that contribute to the optimization objective (i.e., loss) during model training. Such "selective negatives" include only those boxes containing landfast ice, excluding water beyond the floe edge, land, or polynyas not meeting certain target criteria. Non-target polynyas were defined as those mapped with low confidence or not meeting the size threshold as described in section 3.1.5. Additionally, the local separability measure δ^{μ}_{HH} was used to categorize some polynyas as non-targets via minimum contrast thresholds used as a proxy for target difficulty. This is discussed further in section 4.4.1. Fig. 4.2 illustrates the selective sampling of negative and positive boxes from an image patch during training.

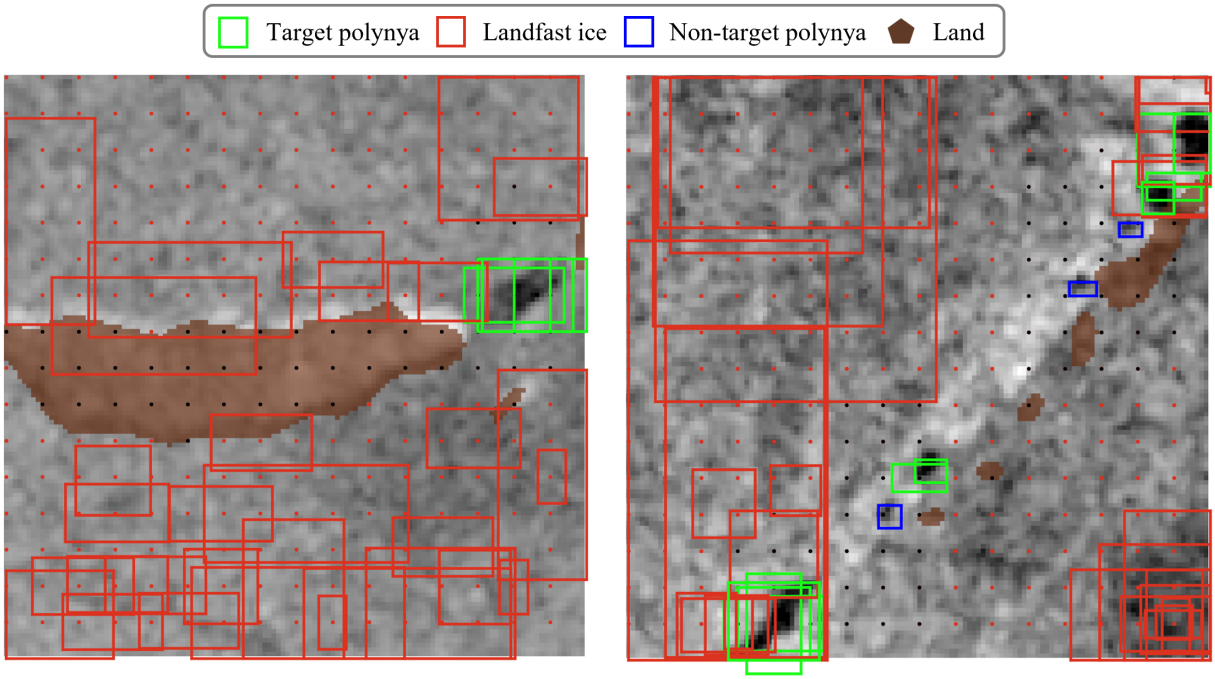

Figure 4.2: Illustration of selective training samples in S1 images taken Mar. 9, 2021 (left) and April 9, 2018 (right). Each red dot indicates the centrepoint of 9 negative RPN anchor boxes, while red boxes represent those with the highest RPN objectness scores after predicted offsets are applied. These top-N scoring boxes (for this illustration, N is set to 32) are used for RPN objectness loss computation, $\mathcal{L}_{obj}$. All predictions classified as positive are shown as green boxes. Blue boxes represent "non-target" labels ignored during training, e.g. polynyas not meeting the size threshold. Boxes centered over land, with centrepoints indicated by black dots, are discarded during training and inference.

4.2 Training and Evaluation Methods

4.2.1 Addressing Correlation between Polynyas and Land

An important characteristic of the synthesized landfast ice polynya dataset is the presence of multiple observations of the same recurring polynyas throughout and across seasons. This introduces a strong correlation between polynya occurrence and local topography, i.e., unique land formations that show up in SAR backscatter every time a recurring polynya is imaged. A risk in training spatial models on such a dataset is an overfit to specific land

Region	Use	Seasons	N. Scenes
Sanikiluaq	Training	2016-2022	278 (90%)
	Evaluation	2016-2022	27 (10%)
Nain	Evaluation	2020-2022	18 (100%)

Table 4.2: Training and evaluation dataset partitioning.

formations correlated with polynya occurrence, which may go unnoticed if the models are evaluated against these same land-correlated targets. It should be noted that some degree of land correlation with polynya occurrence can reasonably be interpreted as a useful *feature* for polynya detection, as topographic narrows often contribute to the forced-convection mechanisms that *create* polynyas. It is an overfit to *specific* land geometries that do not generalize well to identifying other causal topographic patterns that is of concern. While section 4.2.6 proposes mitigation strategies leveraging data augmentation in the training set, section 4.2.2 describes the selection of evaluation datasets that in combination are resilient to and indicative of potential land-overfit behaviour.

4.2.2 Training & Evaluation Datasets

To capture the true generalization performance of polynya detection models between regions with different topographies, training was carried out *only on polynya observations from the Sanikiluaq dataset*, while a subset of Sanikiluaq observations and all Nain observations were held out and used for evaluation. As summarized in table 4.2, this yielded 278 training and 27 evaluation scenes for the Sanikiluaq dataset (Jan. 22 to May 14, 2016-2022) and 18 evaluation scenes for the Nain dataset (Mar. 19 to May 14, 2020-2022). The 27 Sentinel-1 scenes from the Sanikiluaq dataset held out for evaluation were selected to provide representation across seasons, time of year, and target separability characteristics, i.e., a diversity in the number of polynya targets in each scene based on the local contrast measure δ^{μ}_{HH} (equation 3.3). These characteristics for each Sanikiluaq and Nain evaluation scene are detailed in Appendix B.

4.2.3 Contrast-Based Polynya Target Definitions

Analyses of polynya and landfast ice signatures described in section 3.2 illustrated variable and frequently low polynya separability from surrounding landfast ice. Rather than

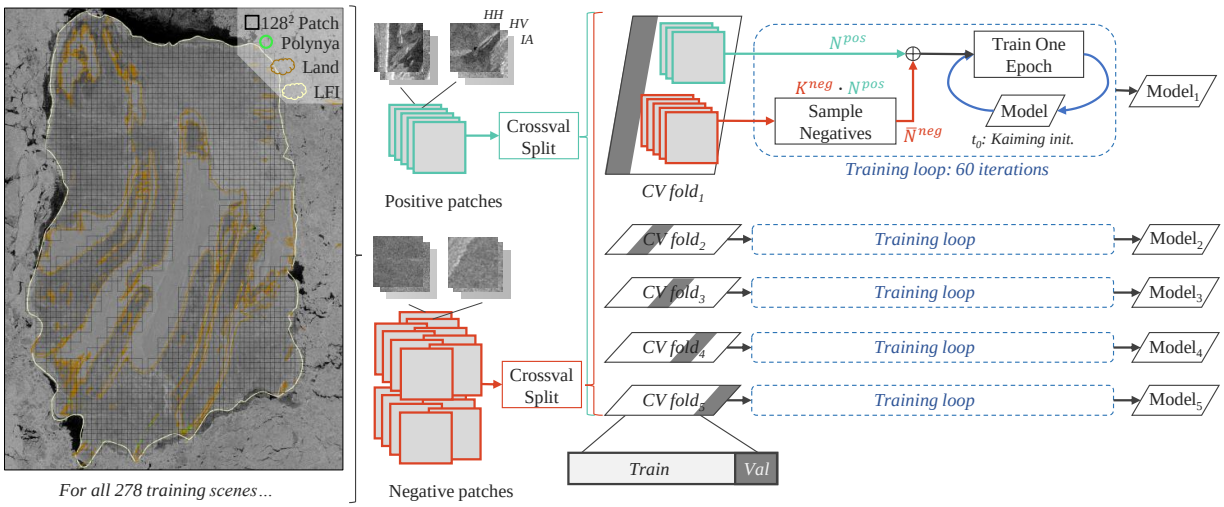

Figure 4.3: Model training workflow. LFI stands for landfast ice.

attempting to train models to detect polynyas in all conditions inclusive of those that produce extremely low and, based on observation during mapping, infeasible, separability, an approach to training models to detect "feasible" polynya signatures was desired.

Knowledge of imaging and environmental parameters that influence dual-polarized backscatter intensity from sea ice and water, outlined in section 2.3.3, could in principle be used to predict polynya separability and thus stratify a training and evaluation dataset. For instance, lower separability would be expected under conditions of strong wind roughening or low imaging incidence angle. However, the attempt to predict variations in polynya separability from available imaging and (notably limited) environmental data described in section 3.3 left a significant proportion ($\sim 70\%$) of this variability unexplained. Therefore, rather than attempting to stratify the dataset based on imaging conditions, an empirical approach to target selection for model training and evaluation was taken. The *contrast-of-backscatter-means* measure introduced in section 3.2.1 (eq. 3.3) was used to describe the separability of each SAR polynya observation, enabling a stratification of the dataset into groups representing varying levels of target difficulty. Due to the higher separability characteristics illustrated in section 3.2, the HH-polarized contrast measure was used.

As a baseline configuration, polynya observations having a separability described by $\delta^{\mu}_{HH} \leq -3dB$ were used as training and evaluation targets. Table 3.2 lists the number of polynyas corresponding to this criteria. Alternative contrast-based target definitions, including the absence of such, are explored in section 4.4.1.

4.2.4 Image Patching & Patch Sampling

As full-scene SAR images are generally too large to process at once with CNN models, a patch-based model training and inference approach was used. Patches of size 128x128 were taken based, at the lower bound, on the distribution of polynya sizes (Fig. 3.4), and at the upper bound on the flexibility to upsample patches as a feature resolution-enhancement method. Patches were generated with 30% overlap such that all polynyas with dimensions less than 38 pixels would be fully covered by at least one patch. Such patching is a common approach to aerial and spaceborne remote sensing image analysis, e.g. [101, 2].

Given the small and sparse nature of polynya targets, there is a significant imbalance ($\approx 10^2$) in the number of generated image patches that contain polynyas vs. those that do not — hereafter referred to as "positive" and "negative" patches, respectively. This imbalance in the classification of patches exists alongside a further imbalance in the number of polynya vs. ice pixels *within* most polynya-containing patches (also $\approx 10^2$). Training on all negative patches would thus lead to a high computational load for many redundant negative examples, and produce an extreme positive/negative class imbalance. On the other hand, sufficient negative examples must be provided during training to capture the true statistical distribution of the class, i.e., representing the spatial and temporal variability of landfast ice in the RoI.

To address these tradeoffs, a *dynamic negative patch sub-sampling* scheme was used. First, all patches generated from full-scene training images were classified as positive or negative based on their full coverage of, or total non-intersection with, a target polynya. At each epoch, all positive training patches were included, and a random sample of negative patches was taken at a ratio $K_{neg} \geq 0$ with respect to the number of positive patches: $N^{neg} = K^{neg} \cdot N^{pos}$. This produced a consistent number of training samples and positive/negative patch balance at each epoch, while capturing the diversity of the negative class over many epochs. This scheme, implemented with $K^{neg} = 2$, was found to outperform static sampling of an equivalent or larger number of negative patches in terms of precision and recall (defined in subsection 4.2.9), respectively.

4.2.5 Data Normalization

Min-Max normalization of images was performed, mapping HH, HV, and IA channel values from the clipped ranges [-30dB, 0dB], [-45dB, -15dB], and [19°, 49°], respectively, to the range [0, 1]. Ranges for backscatter values were based on the distributions observed in the datasets, as reported in chapter 3. This normalization scheme has been implemented successfully for other CNN-based sea ice classification studies, e.g. [12].

4.2.6 Data Augmentation

To increase the variance of land patterns in the SAR backscatter intensity images and thus attempt to lessen the risk of model overfit to polynya-correlated land formations, a set of affine data transforms was used. Illustrated in Fig. 4.4, these included horizontal and vertical flipping, random rotation of up to 45°, and a square crop up to half of the original image size, implemented to preserve all targets in the image and subsequently rescaled to the original resolution using bilinear upsampling. As a baseline configuration, all transforms were applied with a probability of 50%. The exploration of other configurations, including the absence of augmentation during training, is described in section 4.4.

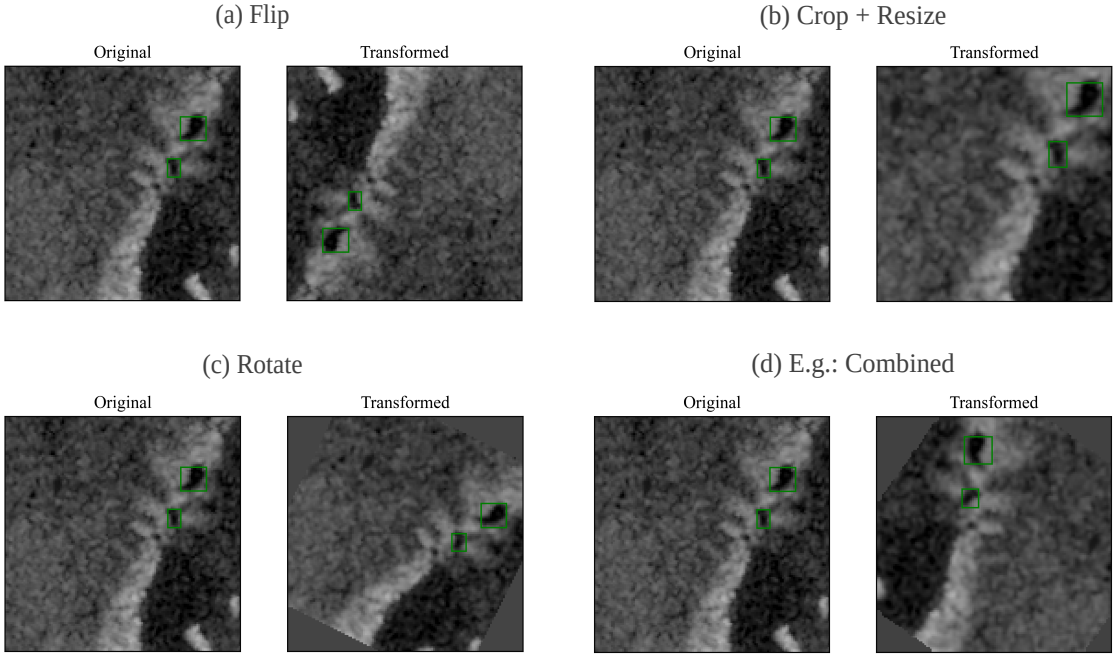

Figure 4.4: Illustration of affine data transforms applied to input image patches and labels: (a) horizontal and vertical flipping, (b) a square crop-and-resize policy that guarantees coverage of all targets in the original input image, (c) a rotation from 0 to 45°, and (d) an example of all augmentations applied probabilistically in composition.

4.2.7 Model Training

Model training was carried out using a standard stratified 5-fold cross-validation approach (e.g. [67]), whereby the entire population of training patches (all those generated from Sanikiluaq S1 training scenes) was split into 5 groups containing consistent ratios of positive to negative patches, with each group used in turn as a validation set for a model trained on the other 4 groups. All experiments followed an identical training scheme, training all sub-networks jointly and from scratch using a fixed number of epochs with a monotonically decaying learning rate. Specific training-related hyperparameters are detailed in table 4.3. Training was carried out on single 32GB NVIDIA Tesla V100 GPUs provided by the Digital Research Alliance of Canada (alliancecan.ca), leveraging experiment-tracking tools offered by the Weights & Biases platform (wandb.ai/site).

Table 4.3: Model training hyperparameters used throughout experiments.
* Batch size refers to the number of images, not bounding boxes, per batch.

Hyperparameter	**Value**
N. Epochs	60
Optimizer	Stochastic Gradient Descent, e.g. [115, 49]
Momentum	0.9
Weight Decay	None
Learning Rate	Cosine decay; 0.01 $\rightarrow$ 0.001
Batch size*	64

4.2.8 Full-Scene Model Inference

During evaluation, each evaluation scene was first divided into 30%-overlapping, 128x128-pixel patches. Patches having over 20% coverage beyond the floe edge or over 80% coverage of land were discarded. Model inference was performed on each remaining patch, then per-patch predictions were merged back into full-scene pixel space and subject to NMS (the same deduplication algorithm used within FRCNN) with an IoU threshold of 0.1. This overlap-and-deduplicate strategy ensured that every polynya measuring less than 38 pixels (largest dimension) was completely covered by at least one patch, with the highest-confidence detection from all target-covering patches taking precedence. After discarding predicted boxes below a user-defined score threshold, τ_{score}, full-scene detections were then

evaluated against target labels based on mapped polynyas and target criteria. Model detections were classified as TP or FP relative to target polynyas based on a configurable IoU threshold, τ_{IoU}. Target polynyas not detected by the model were categorized as FN.

It is emphasized that, following this full-scene inference scheme, models were evaluated on a heavily imbalanced number of negative (land and landfast ice only) versus positive (target polynya-containing) image patches. This imbalance was on the scale of 10^2 to 10^3 depending on the stringency of the target polynya criteria applied.

4.2.9 Evaluation Metrics

For a given τ_{score} and τ_{IoU}, precision (eq. 4.2) and recall (eq. 4.1) metrics describing the specificity (commission error) and sensitivity (omission error) of detections may be calculated. The f1-score (eq. 4.3) is defined as the harmonic mean between precision and recall.

$$recall,\ r = \frac{\#TP}{\#TP + \#FN} \tag{4.1}$$

$$precision,\ p = \frac{\#TP}{\#TP + \#FP} \tag{4.2}$$

$$f_1 = 2\frac{p \cdot r}{p + r} \tag{4.3}$$

While precision and recall metrics are of great practical use for understanding a model's hit, miss, and false alarm rates, they are sensitive to the chosen τ_{score}, which may capture the distribution of predicted box scores differently for individual models and thus provide limited utility for model intercomparison. For this reason, it is standard to assess object detection models across a range of score thresholds to understand their performance more holistically. This is commonly done graphically with precision-recall (PR) curves and quantitatively with the average precision (AP) score. PR curves plot a model's precision score at uniformly-sampled recall values from 0 to 1, while the AP metric takes the average of these precision scores. The standard AP metric introduced by Everingham et. al [34] is implemented using 11 precision scores corresponding to uniformly-sampled recall levels from 0 to 1. The precision scores are defined as $p_{interp}(r)$, describing the highest precision value achieved for a recall level equal to or higher than r [34]:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} p_{interp}(r). \tag{4.4}$$

All evaluation metrics reported in this thesis, including AP, were computed for τ_{IoU} = 0.1 across detections from all scenes in the respective evaluation datasets[1]. This low threshold allows predicted boxes having relatively low intersection with target boxes to be classified as true positives, adding resilience to label imprecision (a known challenge for small-target datasets, e.g. [147]).

4.3 Optimizing Key Detection Model Design Elements

Leveraging the adapted FRCNN model of section 4.1 and the training and evaluation strategies of section 4.2, this section reports the investigation of three important model design elements to optimize the detection of small, sparse landfast ice polynya targets. These elements, following advancements in the natural-scene and optical remote sensing small object detection fields discussed in chapter 2, include feature representation capacity, input image and feature map resolution, and loss function selection, parameterization, and balancing. Table 4.4 summarizes the experimental configurations pursued. Unless otherwise noted, all experiments described in this section were performed using {HH, HV, IA} input channels, affine data augmentation applied with a probability of 50% per patch, and target polynyas having a local separability in backscatter intensity adhering to $\delta^{\mu}_{HH} \leq -3dB$. These data elements are explored further in section 4.4.

4.3.1 Methods

Feature Extraction Network Depth Selection

Following the hypothesis that an effective set of features for landfast ice polynya detection in dual-polarized SAR imagery is less complex than that required for the multi-class natural-scene contexts for which EfficientNet and Faster-RCNN were originally developed, an investigation into the minimum feature extraction network depth required to yield effective performance on the polynya detection task was carried out. An ablation study was

[1]Note that τ_{IoU}, used for categorizing model detections during evaluation, is distinct from the IoU threshold hyperparameters used for categorizing boxes *within* the model during training.

Table 4.4: Key Faster-RCNN design elements investigated in this section, as discussed in section 4.3.1. Configurations with * indicate those used between experiments unless otherwise indicated.

Design Parameter	**Configurations**
Backbone network stages	S-1, S-2, S-3*, S-5, S-7, FPN-357
Input image resolution	128x128*, 256x256, 512x512, 128x128 + FPN-123
RPN loss balancing	CE_{16}, CE_{64}*, CE_{256}, $focal_{\gamma=0,\ \alpha=0.9}$, $focal_{\gamma=0,\ \alpha=0.99}$, $focal_{\gamma=2,\ \alpha=0.25}$, $focal_{\gamma=5,\ \alpha=0.25}$

performed whereby the EfficientNet-B0 backbone network (table 4.1) was truncated at progressively deeper stages from 1 to 7, with the final feature map being used by the RPN and RoI Head sub-networks for bounding box classification and regression. A configuration using FPN to reconstruct a stage-3 feature map from a stage-7 feature map was compared to the uni-directional configurations.

As the objective of this experiment was to ascertain the minimum feature *complexity* (network depth) sufficient for the RPN and RoI Head sub-networks to detect polynyas, the effect of changing feature map *resolutions* at different network stages due to internal downsampling operations was viewed as a confounding factor that might independently influence bounding box prediction performance. In particular, the resolution degradation at deeper network stages was expected to reduce the utility of feature maps at these stages for original-resolution input images, potentially under-representing the utility of more complex features for detecting polynyas. An attempt to alleviate this was pursued by upsampling input image patches for configurations using 5- and 7-stage backbone networks by factors of 2x and 4x, respectively. Thus, 3-, 5-, and 7-stage networks produced feature maps with equivalent dimensions for downstream bounding box prediction. For each configuration the network was trained from scratch, allowing it to learn features best suiting the available number of stages while being nonetheless constrained by the chosen depth. In these experiments, notation as "S-X" describes a backbone network without FPN that includes stages from 1 to X, and notation as "FPN-357" describes a backbone network with FPN where the largest number denotes the deepest backbone network stage reached and the smallest number indicates the shallowest stage reconstructed. Table 4.5 summarizes the number of learnable parameters associated with each network depth configuration.

Table 4.5: Number of feature extraction network parameters by stage utilisation. Each configuration represents the indicated EfficientNet stages plus an FPN lateral connection consisting of a 1x1 convolution mapping to a 128-channel feature map followed by a 3x3 anti-aliasing convolution, as used in the experimental protocol.
*Includes convolutional layers for FPN lateral connections and anti-aliasing operations

Stage Utilisation	N. Conv Blocks	N. parameters
S-1	2	0.15M
S-2	4	0.17M
S-3	6	0.22M
S-5	12	1.0M
S-7	17	3.8M
FPN-357	17+*	5.5M

Input Image Upsampling

A common method to address limitations in small-object detection results in natural scene and optical remote sensing contexts is that of super-resolution, i.e., increasing the resolution of an input image to enhance available small target information and thereby improve the chance of detection [53, 135]. While single-image super-resolution has been recently pursued with generative adversarial networks (GANs) trained to map low-resolution images or feature maps to high resolution [81, 4], this approach can be resource-intensive both in terms of computation as well as design and training times when adapting the method to a new domain. Given the geometric simplicity of polynya targets in SAR relative to object-dense, natural-scene RGB images with higher edge densities, colour details, and occlusions, a simple bilinear upsampling approach was used to investigate the influence of input image scale on detection performance results. Input image sizes of 128x128 (original), 256x256 (2x upsample), and 512x512 (4x upsample) were tested using a 3-stage backbone network (S-3 of table 4.5) based on results from feature depth experiments (section 4.3.3). A configuration having an input image size of 128x128 with an FPN-123 backbone configuration was also tested, reconstructing a feature map with an equivalent spatial resolution to a 3-stage network subject to a 512x512 input image. This explores the relative efficacy of the feature-level resolution merging of FPN versus input image-level upsampling.

RPN Loss Function Selection and Parameterization

To address the massive imbalance in positive and negative anchors, a balanced sampling scheme is typically used alongside CE loss in the RPN sub-network, whereby a fixed number

of boxes from each image are sampled for loss computation at a set ratio of negative to positive boxes. Both the number of boxes and balancing ratio are network hyperparameters (appendix A). If there are fewer positive boxes than that prescribed by the ratio, the sample is padded with sampled negative boxes. Given the extremely small and sparse polynya targets, there are typically only a handful of positive anchor boxes per positive input patch, thus the boxes sampled for loss computation are dominated by negatives. A larger sample size may lead to a higher diversity of negative examples and reduce model false positives, however the associated increase in imbalance may reduce the model's sensitivity to polynya targets. To investigate this imbalance problem, three CE loss sampling configurations were tested: "CE_{16}" (16 boxes), "CE_{64}" (64 boxes), and "CE_{256}" (256 boxes). These were investigated alongside the image patch-balancing parameter K_{neg} described in section 4.2.4, for which values of 2 and 6 were used.

Rather than relying on an imbalance-insensitive loss function and requisite sampling scheme, Lin et. al [85] proposed "focal loss", an alternative loss function for anchor scoring. This function is cross-entropy loss (equation 2.6) weighted by $\gamma \geq 0$, an exponential "focusing" parameter that emphasizes poorly-classified (e.g., rare) examples, and a rare-class balancing parameter $\alpha \in [0, 1]$. With these modifications, *all* negative and positive anchors are used for loss computation.

$$\mathcal{L}_{obj}^{focal} = \left\{ \begin{array}{ll} -(\alpha)(1-p_{obj})^{\gamma} log(p_{obj}) & if\ y = 1 \\ -(1-\alpha)(p_{obj})^{\gamma} log(1-p_{obj}) & otherwise \end{array} \right\} \tag{4.5}$$

Motivated by the success of focal loss in guiding effective anchor scoring under extremely imbalanced conditions in the one-stage RetinaNet detector [85, 90], several parameterizations of the function commensurate with ranges suggested by the authors [85] were attempted alongside CE-balancing experiments. These parameterizations are summarized in table 4.4. All experiments in this section used 128x128-pixel input images with a 3-stage backbone network (S-3) and a negative patch sampling ratio (K^{neg}) of 2.

4.3.2 Model Convergence During Training

Prior to examining inference results on the evaluation datasets, a brief case study is presented here illustrating model convergence during training. This is an important step in ensuring that models have been properly trained and that their evaluation results can indeed be used to draw realistic conclusions.

Training and validation loss curves of the four FRCNN loss functions (as described in section 2.4.3) are shown in Fig. 4.5 for the best-performing model configuration across

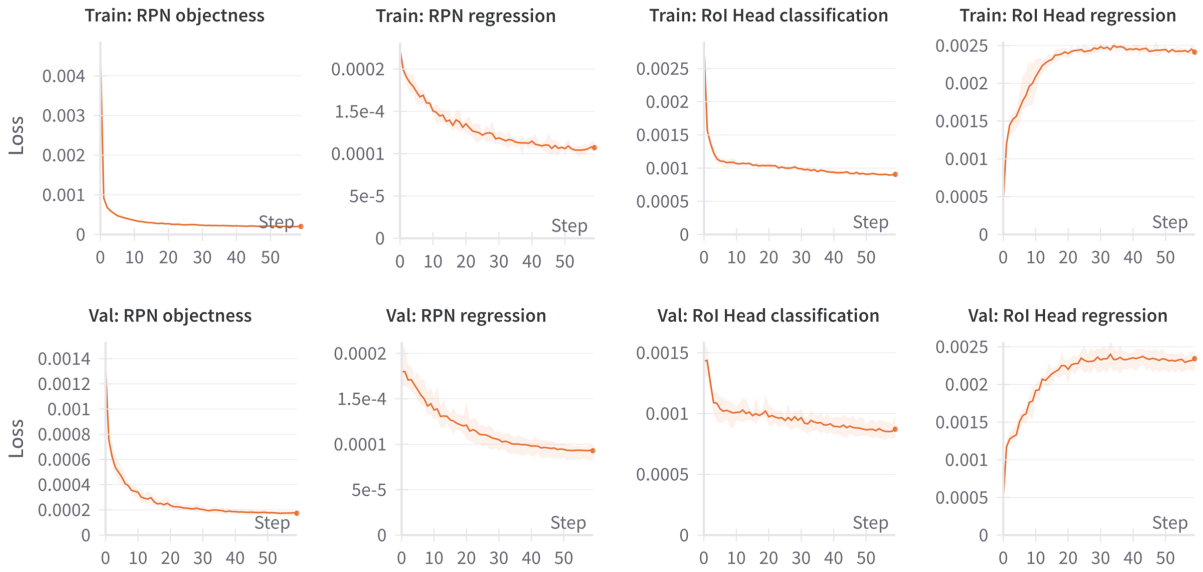

Figure 4.5: Training and validation loss curves of the 4 FRCNN sub-networks for the best-performing model configuration ("*") of table 4.4. Each loss curve indicates the mean (trend line) and min/max (shaded sleeve) loss values for 5 cross-validated models. Each step corresponds to one training epoch, i.e. a pass over the full training dataset.

experiments in this section. Loss values per epoch are summed and normalized by the number of image patches in the train or validation dataset. Both training and validation losses are calculated from a *subset* of all predictions in each patch: RPN objectness and RoI Head classification losses sample a fixed number of boxes per patch, while RPN and RoI Head regression losses are calculated only for positive boxes. RPN regression loss is normalized by the number of anchor locations in the patch, while RoI Head regression loss is normalized by the number of proposal boxes passed from the RPN.

RPN objectness loss is seen to decrease rapidly and smoothly in both training and validation, representing the RPN's improvement in predicting polynya vs. ice labels for highly imbalanced anchor boxes. RPN regression loss converges but plateaus in training and validation, suggesting a limit to the bounding box precision achievable by the model. RoI Head classification loss initially decreases rapidly but plateaus (with only marginal steady improvements) higher than RPN objectness loss in both training and validation. Since RPN passes boxes with the highest objectness scores to the RoI Head, the higher classification loss in the Head likely reflects the increased difficulty of the negative exam-

ples forwarded by the RPN. This is an implicit form of hard-negative mining for the Head sub-network, however it is clear only marginal improvements are made on classifying these "hard negatives" over the course of training. Finally, the RoI Head regression loss demonstrates a striking increase and plateau in both training and validation. Considering that this loss is calculated only for positive boxes forwarded from the RPN, however, it can be understood that as the RPN improves its classification performance, more positive boxes are forwarded and thus used for regression loss calculation in the RoI Head. Subsequently, little bounding box localization refinement is accomplished by the RoI Head on top of the regression predicted by the RPN.

These convergence patterns are representative of most models trained in this chapter. They illustrate the relative importance of the RPN objectness classification predictor, which converges the most significantly and influences downstream components of the model. They also suggest that the role of the RoI Head network in binary ice/water classification and localization may be somewhat redundant in relation to the RPN, although some improvement seems to be made in classifying difficult negative examples. Ultimately, the analysis of sub-network convergence patterns is limited by the relatively small sample sizes taken for loss computation, and thus subsequent assessment of model performance is carried out with respect to model inference results on held-out evaluation datasets.

4.3.3 Results & Discussion

Evaluation results for investigations into feature depth, input resolution, and loss balancing are presented and discussed in this section. Fig. 4.6 shows the mean precision-recall curves for each experiment, while table 4.6 records numerical evaluation metrics with mean and variance for five cross-validated models per experiment. AP scores (eq. 4.4) are used to illustrate the overall performance of models, while recall (eq. 4.1) and precision (eq. 4.2) metrics taken at τ_{score}= 0.8 illustrate results that can be achieved in terms of model sensitivity and specificity. This threshold on τ_{score} was informed by observed score distributions of true-positive boxes predicted by the detection models, and was taken for its yield of higher recall values than f1-optimal score thresholds. Alternative precision/recall metrics can be estimated visually using the precision-recall curves. Note that it is possible for one model to present a higher AP score but a lower precision at τ_{score} than another model; this may happen when the first model predicts higher-scoring boxes than the second model, leading to higher recall but lower precision for the same score threshold (the *average* precision over all score thresholds may still be higher for the first model).

Where differences in model performance responses to certain design elements occur

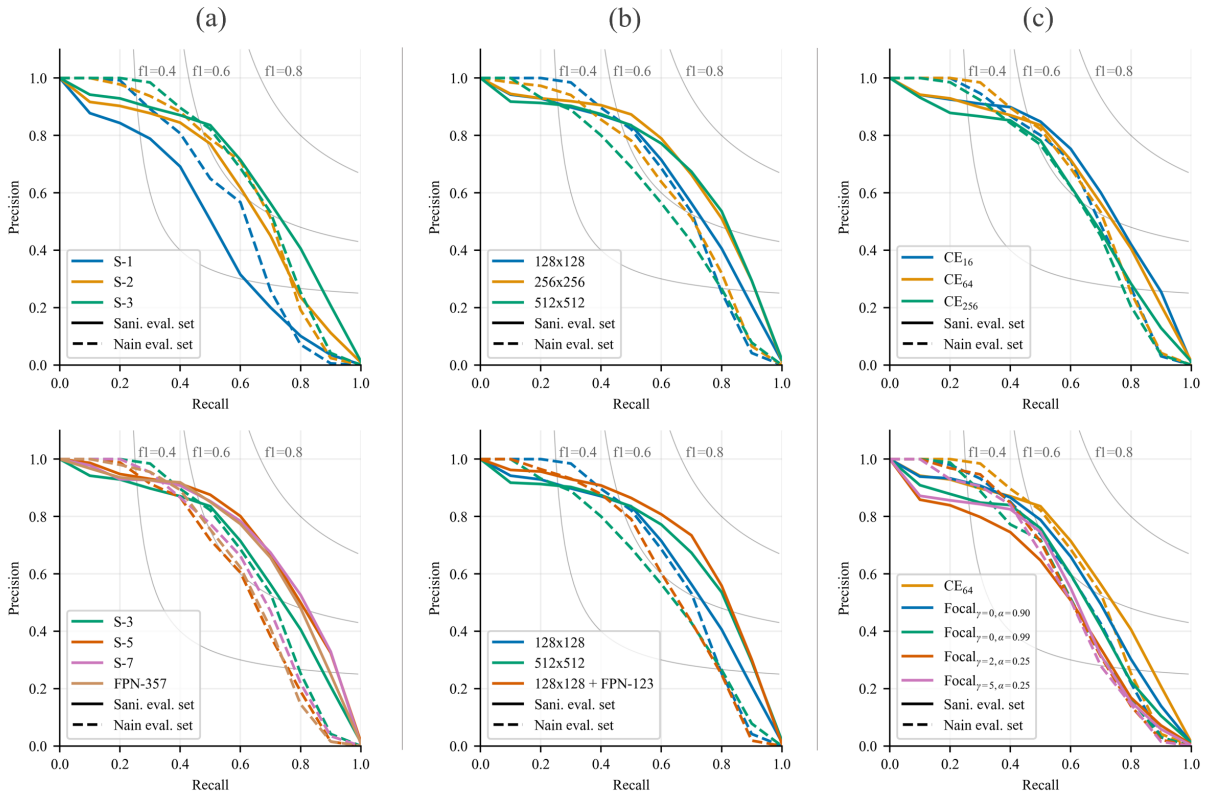

Figure 4.6: Precision-Recall curves at τ_{IoU}= 0.1 for investigations into (a) feature extraction network depth, (b) input image upsampling, and (c) RPN cross-entropy loss sampling and focal loss parameterization corresponding to results in table 4.6. Results are generated for $\delta^{\mu}_{HH} \leq -3dB$ targets with Affine_{p50} augmentation on HH, HV, IA input features.

between the two evaluation datasets (Sanikiluaq and Nain), interpretive preference is generally given to results on the Nain dataset since this region was completely held-out during training and thus better represents a model's ability to retrieve and classify relevant ice / water features with no aid from potential land correlations discussed in section 4.2.1. The inclusion of both datasets in subsequent analyses is useful for identifying instances where overfit to land is suspected.

Table 4.6: Evaluation metrics for investigations into model hyperparameterization, backbone feature utilisation, input image resolution, and loss function selection, sample balancing, and parameterization. Metrics means across 5 cross-validated models are reported in percentage points, followed by standard deviation in parentheses. All metrics are calculated for τ_{IoU}= 0.1. Precision and recall values are given for τ_{score}= 0.8. Boldface rows indicate the best-performing configuration for each experiment based on interpretations discussed in section 4.3.3. Configurations with * indicate those used between experiments.

Experiment	Sani. AP	Sani. Recall	Sani. Precision	Nain AP	Nain Recall	Nain Precision
Hyperparameters						
Original	52.1 (1.11)	46.6 (1.80)	61.0 (2.15)	56.2 (4.26)	53.9 (7.85)	56.1 (2.32)
*** Adapted**	**67.1 (0.59)**	**80.0 (0.93)**	**39.8 (2.70)**	**65.6 (2.02)**	**78.4 (1.85)**	**28.3 (2.15)**
Backbone Stages						
S-1	48.7 (1.75)	62.8 (2.33)	26.1 (2.44)	56.8 (2.11)	72.1 (4.12)	17.5 (1.37)
S-2	61.2 (2.69)	72.4 (1.12)	35.3 (3.07)	63.9 (2.37)	77.2 (2.51)	23.8 (4.78)
*** S-3**	**67.1 (0.59)**	**80.0 (0.93)**	**39.8 (2.70)**	**65.6 (2.02)**	**78.4 (1.85)**	**28.3 (2.15)**
S-5	72.4 (1.36)	86.1 (1.08)	39.9 (1.88)	60.6 (2.20)	75.0 (3.64)	26.6 (3.43)
S-7	72.0 (0.83)	85.7 (1.79)	39.5 (2.18)	63.5 (2.17)	75.8 (3.96)	29.1 (2.10)
FPN-357	70.6 (1.88)	81.6 (1.53)	39.7 (1.31)	61.6 (2.72)	75.0 (1.63)	30.5 (3.47)
Input Image Resolution						
*** 128x128 (orig.)**	**67.1 (0.59)**	**80.0 (0.93)**	**39.8 (2.70)**	**65.6 (2.02)**	**78.4 (1.85)**	**28.3 (2.15)**
256x256	71.3 (0.81)	85.7 (1.73)	38.0 (2.70)	64.3 (3.90)	81.6 (2.08)	22.8 (2.50)
512x512	70.3 (1.18)	84.9 (1.87)	40.8 (2.49)	59.7 (1.92)	79.9 (2.94)	24.7 (2.54)
128x128 + FPN-123	73.0 (1.06)	83.6 (2.78)	44.6 (4.06)	62.4 (2.49)	78.5 (1.79)	24.9 (2.52)
Patch & CE Loss Sampling						
K_{neg}=2; CE_{16}	68.7 (2.09)	87.9 (1.44)	27.9 (2.31)	64.7 (2.90)	81.1 (4.09)	20.9 (1.31)
*** K_{neg}=2; CE_{64}**	**67.1 (0.59)**	**80.0 (0.93)**	**39.8 (2.70)**	**65.6 (2.02)**	**78.4 (1.85)**	**28.3 (2.15)**
K_{neg}=2; CE_{256}	62.0 (1.57)	66.1 (0.97)	52.5 (3.96)	62.1 (1.71)	71.9 (1.73)	38.0 (0.73)
K_{neg}=6; CE_{16}	74.1 (2.65)	84.6 (2.57)	46.4 (4.85)	56.2 (4.65)	68.0 (4.77)	29.1 (3.53)
K_{neg}=6; CE_{64}	71.3 (0.92)	74.5 (1.01)	61.6 (2.14)	64.9 (0.85)	72.6 (3.37)	42.5 (5.69)
K_{neg}=6; CE_{256}	66.4 (2.18)	61.7 (3.47)	72.9 (1.68)	61.7 (3.69)	61.8 (7.57)	53.3 (2.91)
Focal Loss (K_{neg}=2 ∀)						
$F_{\gamma=0,\ \alpha=0.90}$	64.0 (2.00)	76.4 (1.94)	35.2 (4.44)	61.6 (2.27)	76.9 (4.45)	26.7 (3.30)
$F_{\gamma=0,\ \alpha=0.99}$	59.9 (1.48)	73.5 (1.91)	31.9 (2.95)	58.1 (4.06)	72.2 (4.43)	23.7 (2.50)
$F_{\gamma=2,\ \alpha=0.25}$	55.3 (2.45)	68.1 (2.07)	36.9 (2.80)	57.1 (2.86)	72.8 (3.43)	23.4 (3.41)
$F_{\gamma=5,\ \alpha=0.25}$	56.6 (1.02)	69.6 (1.62)	30.7 (6.20)	57.3 (3.18)	72.2 (5.14)	23.2 (2.13)

Feature Extraction Network Depth

Differences in performance patterns on Sanikiluaq and Nain evaluation datasets were observed in response to feature extraction network depth. For Sanikiluaq, evaluation performance increased significantly with network depth to a maximum at 5 stages, while evaluation performance on Nain varied less and peaked at a network depth of 3 stages. The 3-stage network also outperformed the baseline configuration of FPN-357 on the Nain

dataset, but not on the Sanikiluaq dataset. This difference is believed to be an artefact of polynya-land correlation; increased feature complexity offered by deeper backbone networks may allow these networks to partially "overcome" data augmentation strategies that seek to obscure polynya-land correlation, e.g. by learning rotation- and scale-invariant representations of specific land features that shallower networks cannot. Thus, following the signal from evaluation performance on the completely held-out Nain dataset, it is observed that a 3-stage EfficientNet-B0 network offers sufficient representational capacity to discriminate landfast ice and open water features for effective polynya detection. To illustrate the significance of this finding, the 3-stage EfficientNet-B0 network has nearly 20x fewer learnable parameters (0.2M) as the 7-stage network (3.6M — table 4.5), and was found to consume 50% less GPU memory during training.

Input Image Upsampling

Limited improvements in polynya detection performance were observed with upsampled input images using a 3-stage (S-3) backbone network. While a modest improvement on the Sanikiluaq evaluation dataset was observed with 2x-upsampled images, performance actually *decreased* with input image size on the Nain evaluation dataset. This decrease is attributed to a greater imbalance in the number of RPN anchors — for the same anchor sample spacing, a 4x-upsampled input image (512x512) has 16 times the number of anchors as the original image (128x128). As denser anchor spacing relative to original input dimensions was identified as a possible mechanism to *improve* detection performance, anchor spacing was not normalized to scale for these experiments (anchor sizes, naturally, were). The use of feature-level resolution merging with FPN slightly outperformed the equivalent image-upsampling approach (128x128 input image with FPN-123 backbone vs. 512x512 input image with S-3 backbone) on both datasets, however did not outperform the unmodified scale (128x128 with S-3) on the Nain evaluation set. Differences in performance between the datasets regarding scale may be due to the occurrence of smaller polynyas in the Sanikiluaq dataset (Fig. 3.4), however it is difficult to conclude that performance improvements on the Sanikiluaq dataset in response to upsampling are related to better discrimination of open water features more than they are to an increased capacity to learn correlated land features. Regardless, these experiments illustrate the ability of the FRCNN detection model to handle small (e.g. 10x10 pixel) targets in dual-polarized SAR imagery, and demonstrates the efficacy of feature-level resolution enhancement with FPN in this context relative to more computationally-intensive input image upsampling operations.

Loss Function Selection, Sampling & Parameterization

Findings related to RPN objectness loss function design revealed that the configuration used in the baseline FRCNN model (cross-entropy loss sampling 64 anchor boxes per image) outperformed modifications related to both CE loss sampling and focal loss parameterizations as interpreted from evaluation results on the Nain dataset. As shown in Fig. 4.6, smaller, more class-balanced samples (N=16) for CE loss increased polynya recall at the cost of precision, while larger (N=256) samples increased precision at the cost of recall. The same loss sampling configurations applied to an input image patch sampling scheme with 3x more negative patches (K_{neg}= 6) revealed the same tradeoff with higher precision and lower recall scores. While overall performance (AP_{10}) on the Sanikiluaq evaluation set was maximized for $K_{neg} = 6$, Nain evaluation results were stronger for $K_{neg} = 2$. A possible explanation for this is that models trained on less balanced data may learn to condition more strongly on land correlations that can be exploited on the Sanikiluaq evaluation set but not Nain. CE_{16} and CE_{64} sampling yielded similar overall performance results for Sanikiluaq and Nain datasets, demonstrating the efficacy of the dynamic negative patch sampling scheme described in section 4.2.4 in capturing sufficient diversity of the negative class. Based on the maximum AP_{10} score for the Nain evaluation set, CE_{64} with $K_{neg} = 2$ was identified as the strongest-performing configuration.

All focal loss parameterizations, evaluated for $K_{neg} = 2$, performed worse on both evaluation datasets than most sampled CE loss configurations. Interestingly, parameterizations equivalent to balanced cross-entropy (higher loss weighting for positive boxes) outperformed those using the focusing parameter γ. Based on these results, further tuning of the focusing parameter beyond the ranges recommended in [85] was not pursued.

4.3.4 Conclusions from Model Optimization Experiments

To summarize, three main findings were observed from the detection model optimization experiments described in this section:

1. A simplified 3-stage feature extraction network with 20x fewer parameters than the original 7-stage network demonstrated a similar capacity to detect landfast ice polynyas with less tendency to overfit to polynya-correlated land formations, validating the hypothesis that polynya detection requires a much simpler set of features than natural-scene object detection contexts;

2. Using a backbone network with fewer downsampling operations precluded the need for image- or feature-level resolution enhancement strategies to detect small targets;

Table 4.7: Various data scenarios investigated with the polynya-optimized FRCNN network. Configurations with † indicate those used for the model optimization experiments of section 4.3, and * those used between experiments in this section.

Design Parameter	**Values**
Input features	{HH}, {HV}, {HH, HV}, {HH, IA}*, {HH, HV, IA}†
Data augmentation scheme	No augmentation, Affine_{p50}*†, Affine_{p100}
Target polynya criteria (δ^{μ}_{HH})	≥+1dB, None (all targets), ≤-1dB, ≤-3dB*†, ≤-5dB

3. Sampled cross-entropy loss outperformed focal loss for objectness prediction in the RPN, with middle-ground sample sizes (64 boxes per image) outperforming larger (256) and smaller (16) samples.

4.4 Detection Model Evaluation under Different Data Scenarios

This section evaluates the performance of the optimized Faster-RCNN detection network in different data scenarios to provide a comprehensive assessment of the generalization capability of the model. Analysis on the performance variation also provides insights on feature importance and characteristics of the polynya detection problem. Specifically, this section aims to answer three key research questions:

1. Which dual-polarized C-band SAR input channels (of {HH, HV, IA}) provide utility for detecting polynyas in landfast ice?
2. Does data augmentation improve detection model generalizability between regions?
3. What are the performance drop-off characteristics of detection models trained to detect increasingly difficult polynya targets?

4.4.1 Methods

Experiments reported in this section leveraged the optimized FRCNN polynya detection model developed in section 4.3 (indicated by boldface rows in table 4.6) and the same training and evaluation strategies described in section 4.2.

Importance of SAR Polarization and Incidence Angle for Polynya Detection

Various combinations of decibel-scaled HH- and HV-polarized backscatter intensity (σ^0) and incidence angle (IA) features were used as input to the model, with an ablation carried out as summarized in table 4.7. The model architecture was unmodified between configurations except for the first convolutional layer to accommodate the varying number of input image channels.

Importance of Geometric Data Augmentation for Multi-Region Generalization

To evaluate the importance of data augmentation for generalizing from Sanikiluaq polynyas seen in training to Nain polynyas revealed only during evaluation, three variations on the data augmentation scheme described in section 4.2.6 were compared. In one, no transformations were performed; in another, all transforms were applied with a probability of 50%; in the third, the rotation and crop-then-rescale augmentations were applied every time while flipping was applied with a probability of 50%.

Detection Performance across Polynya Separability Levels

Given the variability and frequently low separability of polynya signatures in backscatter intensity, a key question in developing polynya detection models was the response in model performance to increasing polynya target difficulties in terms of backscatter separability. From a research standpoint such an investigation illustrates the capacity of CNNs to capture challenging polynya targets that intensity-based classifiers would struggle with, while from an application standpoint it provides an expectation for the capacities and limitations of model performance.

As discussed in section 4.2.3, the local backscatter intensity contrast measure δ^{μ}_{HH} was used as a simple proxy for the separability of polynya observations in the dataset. In this experiment, different thresholds on δ^{μ}_{HH} were applied to define several polynya target populations. These included all polynyas (no contrast requirement), polynyas having

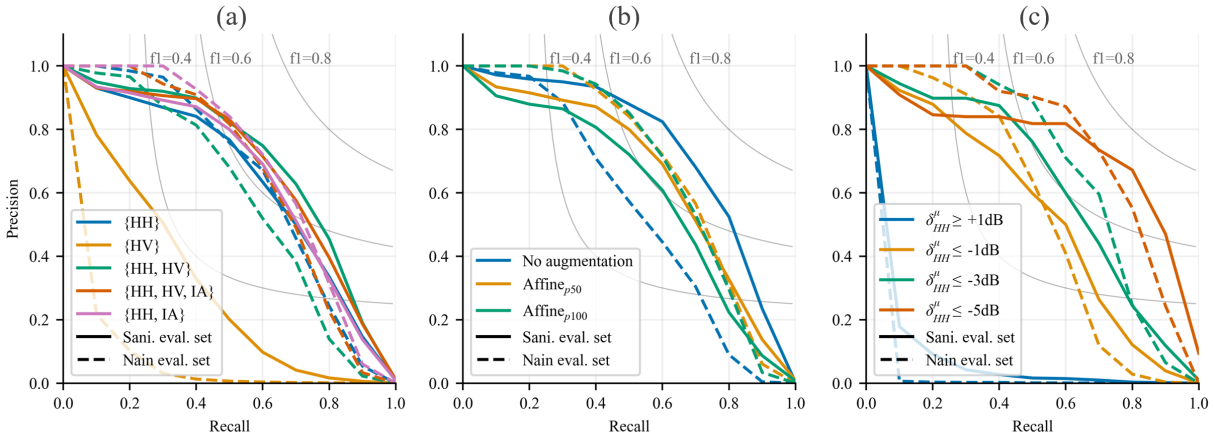

Figure 4.7: Precision-Recall curves at τ_{IoU}= 0.1 for investigations into (a) SAR input features, (b) data augmentation schemes, and (c) contrast-based polynya target definitions corresponding to results in table 4.8.

higher backscatter intensity than surrounding landfast ice ($\delta^{\mu}_{HH} \geq +1dB$), and polynyas having lower backscatter intensity than landfast ice with different minimum contrast levels ($\delta^{\mu}_{HH} \leq \{-1dB, -3dB, -5dB\}$). Models trained to detect both "dark" and "bright" polynyas performed less well than the more specialized models, and are not described further. For each configuration, a distinct model was trained end-to-end and evaluated on targets meeting the corresponding criteria. "Non-targets" not meeting contrast criteria were ignored in training and evaluation as described in section 4.1.3.

4.4.2 Results & Discussion

Evaluation results related to SAR input feature utility, the influence of data augmentation on regional generalization, and detection performance for varying polynya separability levels are discussed in this section. Fig. 4.7 shows the mean precision-recall curves for each experiment, while table 4.8 records numerical evaluation metrics.

Importance of SAR Polarization and Incidence Angle

HH and incidence angle (IA) input features were found to produce the strongest polynya detection performance for the baseline model on the held-out Nain evaluation dataset,

Table 4.8: Evaluation metrics for investigations into SAR input features, data augmentation scheme, and polynya target definition. Mean performance metrics across 5 cross-validated models are reported in percentage points, followed by standard deviation in parentheses. All metrics are calculated for $\tau_{IoU} = 0.1$. Precision and recall values are given for $\tau_{score} = 0.8$. Boldface rows indicate the best-performing configuration for each experiment based on interpretations discussed in section 4.4.2. Configurations with † indicate those used for model optimization experiments (sections 4.3.1 & 4.3.3), and * those used between experiments in this section.

Experiment	Sani. AP	Sani. Recall	Sani. Precision	Nain AP	Nain Recall	Nain Precision
Input Features						
† {HH, HV, IA}	67.1 (0.59)	80.0 (0.93)	39.8 (2.70)	65.6 (2.02)	78.4 (1.85)	28.3 (2.15)
{HH, HV}	68.6 (3.17)	81.0 (2.03)	41.2 (3.37)	58.0 (7.04)	70.4 (8.53)	31.1 (1.00)
{HH}	63.0 (0.95)	79.0 (0.92)	33.2 (1.60)	63.6 (3.90)	76.6 (4.08)	28.10 (2.28)
*** {HH, IA}**	**64.4 (1.79)**	**79.0 (1.73)**	**33.0 (2.63)**	**67.5 (1.63)**	**79.5 (2.33)**	**26.2 (1.58)**
{HV}	32.9 (3.34)	37.9 (3.40)	34.0 (3.23)	12.5 (1.00)	14.6 (1.64)	12.9 (3.50)
Augmentation						
No Augmentation	72.4 (2.57)	85.3 (2.79)	32.3 (2.37)	54.1 (5.20)	72.8 (3.68)	19.4 (2.71)
***† Affine$_{p50}$**	**64.4 (1.79)**	**79.0 (1.73)**	**33.0 (2.63)**	**67.5 (1.63)**	**79.5 (2.33)**	**26.2 (1.58)**
Affine$_{p100}$	59.4 (1.39)	73.6 (2.88)	33.5 (5.30)	66.8 (1.87)	79.6 (2.74)	27.8 (5.77)
Target Criteria						
$\delta^{\mu}_{HH} \geq +1$dB	12.8 (1.07)	6.90 (1.00)	21.1 (6.94)	9.30 (0.03)	0.00 (0.00)	20.0 (40.0)
None (all targets)	49.7 (1.33)	45.4 (2.27)	60.4 (4.37)	36.8 (1.68)	37.5 (1.56)	39.2 (3.51)
$\delta^{\mu}_{HH} \leq -1$dB	56.4 (2.26)	62.5 (1.16)	45.2 (4.32)	56.2 (1.86)	64.4 (0.91)	35.1 (6.07)
***†** $\delta^{\mu}_{HH} \leq -3$dB	64.4 (1.79)	79.0 (1.73)	33.0 (2.63)	67.5 (1.63)	79.5 (2.33)	26.2 (1.58)
$\delta^{\mu}_{HH} \leq -5$**dB**	**73.7 (1.50)**	**91.2 (0.94)**	**38.8 (2.75)**	**74.5 (1.91)**	**90.2 (2.06)**	**24.0 (1.33)**

while HH and HV produced better results for the Sanikiluaq dataset. We infer that the improvement on Sanikiluaq evaluation results with HV is an artefact of overfit to strong land signals in HV backscatter, which, despite the use of affine data augmentation, allows the model to overfit more to correlated land features in training and evaluation sets. On the Nain dataset the model is not able to leverage specific land features learned in Sanikiluaq-only training, and given the low backscatter intensity over landfast ice and water, HV effectively introduces a source of noise to model inference. This interpretation is strongly supported by the results from HV-only models, which perform much better for Sanikiluaq than Nain, giving a sense for the degree of land overfit that is *not* remedied by the Affine$_{p50}$ data augmentation policy. The modest performance gain in Nain evaluation results when including incidence angle information indicates some utility in conditioning HH backscatter intensities on IA for discriminating landfast ice and water features, and it is noted that the limited number of Nain polynya observations at low incidence angles may under-represent this utility.

Importance of Geometric Data Augmentation

Findings from the data augmentation experiment illustrated the tendency of models to overfit to specific land features highly correlated with polynya occurrence in single-region training and evaluation datasets. When no data augmentation scheme was used, the detection model performed very strongly on the Sanikiluaq evaluation dataset but poorly on the Nain evaluation dataset. Introducing affine transformations during training reduced performance on the Sanikiluaq evaluation set, but increased performance greatly on the Nain dataset. This suggests that in the scenario with no data augmentation, the model focuses "too much" on polynya-correlated land features present in the Sanikiluaq training and evaluation datasets but not present in the Nain dataset. The use of affine transformations, which jointly rotate, flip, crop and rescale polynya, ice and land geometries during training, weaken the prior available for the Sanikiluaq evaluation dataset but prompt the model to learn more general features better able to detect Nain polynyas. Increasing the augmentation frequency from 50% to 100% was not found to yield improvements on the Nain evaluation dataset, while further reducing performance on the Sanikiluaq evaluation dataset.

These findings emphasize the importance of data augmentation schemes in this and similar repeat-observation contexts, as well as the utility of evaluating on a completely held-out region to more accurately assess the true generalization potential of developed models. More involved augmentation strategies that could further decouple ice/water signatures from unwanted spatial correlations, such as the copy-paste augmentation [41], may warrant exploration in future works.

Detection Performance across Polynya Separability Levels

Detection model performance metrics showed a monotonic increase for both evaluation datasets with the imposed polynya/ice backscatter intensity contrast threshold measured by δ^{μ}_{HH}. In other words, models trained to only detect polynyas with $\delta^{\mu}_{HH} \leq -5dB$ achieved higher AP_{10} scores than models trained to detect polynyas with $\delta^{\mu}_{HH} \leq -3dB$, which scored better than those on $\delta^{\mu}_{HH} \leq -1dB$ targets. Nonetheless, the baseline FRCNN model was able to detect many challenging $-1dB$ polynyas, for instance achieving target recall values of nearly 70% for Sanikiluaq and 55% for Nain with reasonable precision (55% & 35%, respectively). Models trained to detect $-5dB$ targets achieved 90% recall on both datasets, with somewhat lower precision rates given the fewer true-positive targets used to normalize the number of false positives (eq. 4.2). Table 3.2 can be consulted for assurance that the $-1dB$-trained models did indeed learn to detect more polynyas than $-5dB$-trained

models. Models failed to detect most "bright" polynya targets with $\delta^{\mu}_{HH} \geq +1dB$. As there are very few polynya observations with $\delta^{\mu}_{HH} \geq +3dB$, the poor results here indicate the weak signature of these observations and the lower separability from landfast ice, perhaps due to signature similarities to areas of rough landfast ice (higher σ^0).

4.4.3 Conclusions from Investigations into Data Scenarios

In summary, quantitative evaluation results from investigations into SAR input feature importance, data augmentation, and minimum target separability revealed the following:

1. HV provided limited utility in detecting polynyas from landfast ice, capturing land patterns but not ice/water signatures;
2. Affine data augmentation strategies significantly improved the generalization capability of polynya detection models to regions not seen in training, e.g. increasing the AP_{10} score on the Nain evaluation dataset from 0.54 to 0.68;
3. Models trained to detect only high-contrast polynyas outperformed those trained to detect more challenging polynyas, with AP_{10} scores on the fully held-out Nain evaluation dataset increasing from 0.37 for models trained to detect all polynyas, to 0.75 for models trained to only detect polynyas with $\delta^{\mu}_{HH} \leq -5dB$.

4.5 Inference Results on Evaluation Images

To better understand the capabilities and limitations of the developed polynya detection models, inference examples on held-out evaluation images are examined and discussed in this section. All predictions shown are from the optimized model architecture of section 4.3, indicated by the boldface rows of table 4.6.

Fig. 4.8 illustrates two scenes from the Sanikiluaq and Nain evaluation datasets where the model detected all high-contrast target polynyas (defined by $\delta^{\mu}_{HH} \leq -5dB$) and made very few false positive (FP) predictions. Target polynyas are delimited precisely and have high confidence scores. Both scenes occur later in the season (April 4 and May 6). HH-polarized backscatter over landfast ice is seen to be relatively high and fairly homogeneous, likely related to volume scattering mechanisms within a warming snowpack, and polynyas have high backscatter contrast. Under similar conditions in other evaluation scenes, detection model results had consistently high recall and precision.

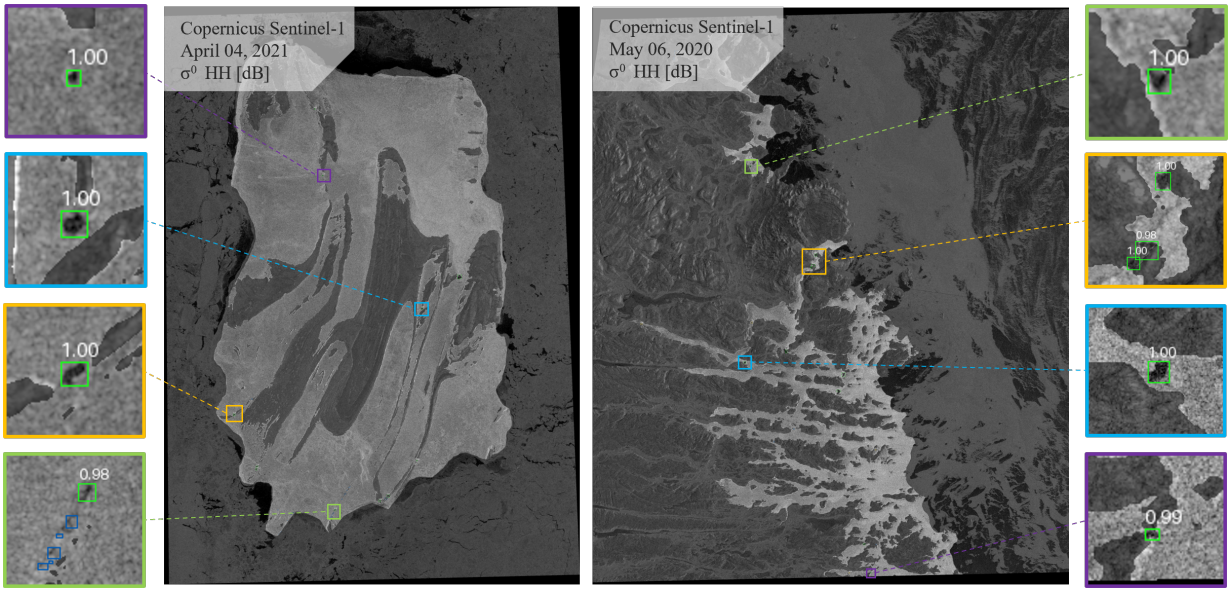

Figure 4.8: Example model predictions on held-out Sanikiluaq (left) and Nain (right) evaluation scenes. Both scenes demonstrate strong model performance for high-contrast polynya targets ($\delta^{\mu}_{HH} \leq -5dB$), having 100% target recall and near 100% precision.

Fig. 4.9 illustrates regions of a different evaluation scene from the Nain evaluation dataset alongside the closest available S2 observation, and compares the predictions of a model trained to detect target polynyas with $\delta^{\mu}_{HH} \leq -1dB$ and $\delta^{\mu}_{HH} \leq -5dB$. In particular, the occurrence of false positive predictions over areas that transition quickly between low and high backscatter intensity is demonstrated. Upon close inspection of the S2 imagery in Fig. 4.9c, these regions can be seen to correspond to boundaries between smooth and rough landfast ice in some cases. It was observed that such polynya-lookalike areas were more frequently predicted by models trained to detect polynyas with weaker signatures, e.g. $\delta^{\mu}_{HH} \leq -1dB$ targets, illustrating an inherent precision-recall tradeoff for the dual-polarized landfast ice polynya detection task.

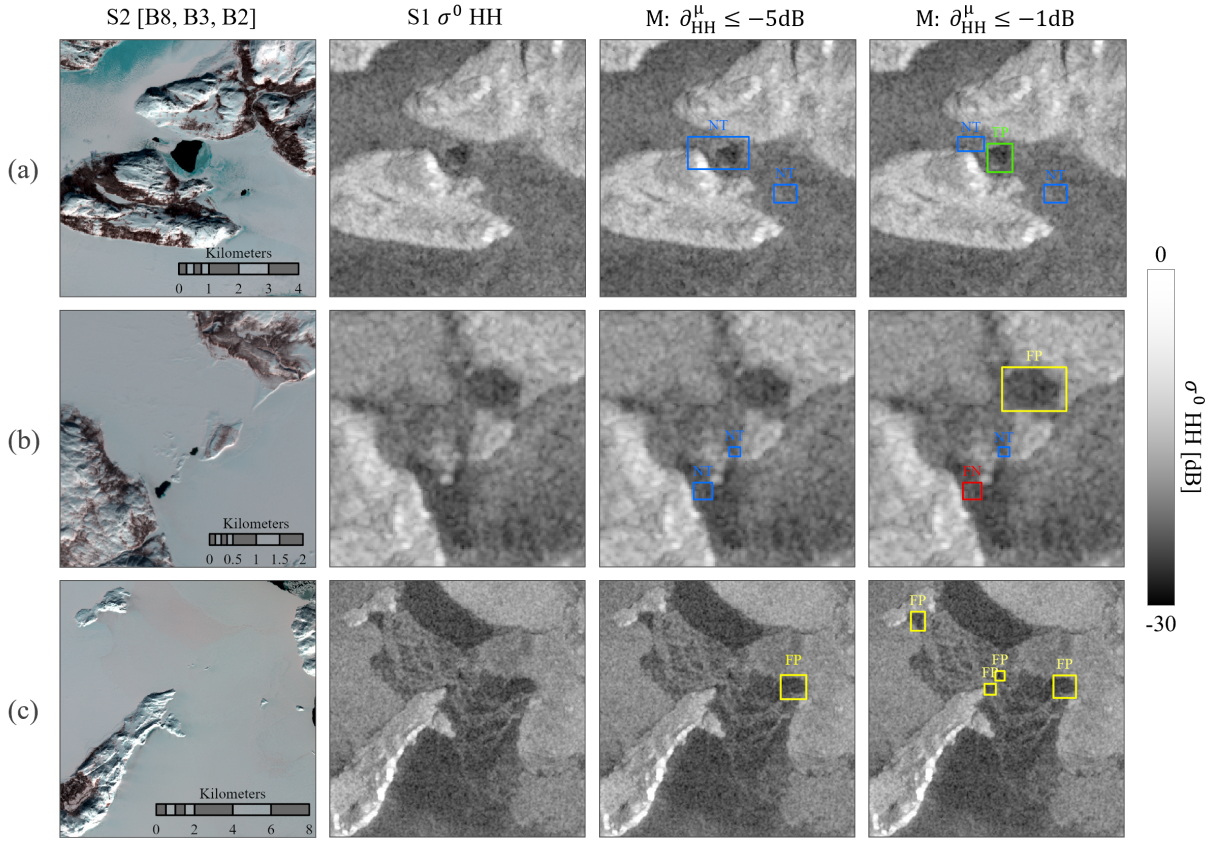

Figure 4.9: Close-up predictions from models (M) trained on different polynya target definitions on a Nain S1 evaluation scene acquired March 26, 2021, with a reference S2 image acquired March 30, 2021. True positive (TP), false positive (FP), false negative (FN), and non-target polynya (NT) boxes are shown. The model trained on more difficult targets ($\delta^{\mu}_{HH} \leq -1$dB) (a) detects polynyas that the $\delta^{\mu}_{HH} \leq -5$dB model does not, but (b) misses more targets, and (b & c) predicts more false positives. False positives tend to occur at boundaries between low and high backscatter intensity, e.g. in transitions between smooth and rough landfast ice as seen upon close inspection of the S2 image in (c).

Finally, Fig. 4.10 shows that missed polynyas (false negatives) consistently possessed weaker signatures (lower contrast) and smaller sizes than most detected (true positive) polynyas. Polynyas with local separability described by $\delta^{\mu}_{HH} \leq -5dB$ and bounding box long-edge lengths above 15 pixels (600m) were consistently detected in the evaluation datasets for both RoIs.

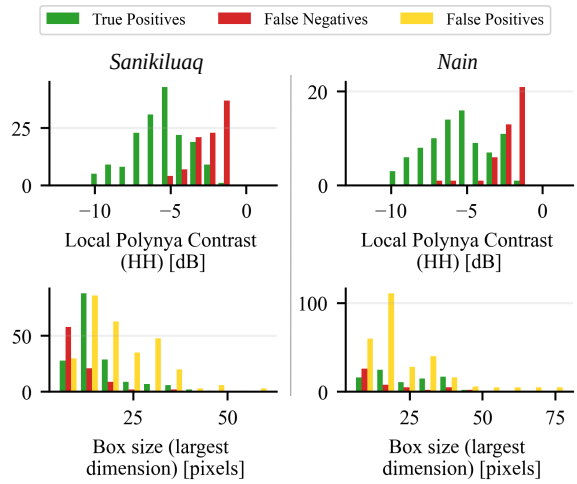

Figure 4.10: Distributions of categorized predictions (τ_{IoU}= 0.1, τ_{score}= 0.8) from a $\delta^{\mu}_{HH}\leq$ −1dB model on both evaluation datasets across local polynya contrast (δ^{μ}_{HH}) and bounding box sizes. False negatives occur at lower contrasts and smaller sizes.

4.6 Chapter Conclusions

This chapter described the development and evaluation of an FRCNN-based object detection network for landfast ice polynya detection in dual-polarized Sentinel-1 SAR imagery. Through several sensitivity and ablation studies carried out on the adapted FRCNN model, a peak in detection performance leveraging a relatively shallow (3-stage) feature extraction network was observed, alongside an agnosticism to input image upsampling on the ability to detect small targets and an empirical advantage of using sampled cross-entropy loss over several parameterizations of the imbalance-motivated focal loss. HH-polarized backscatter intensity and incidence angle information were found to yield the strongest detection results, and the use of affine data augmentation during training was found to significantly improve model generalization between regions. Models trained to detect polynyas having higher contrast in backscatter intensity with surrounding landfast ice outperformed those trained to detect polynyas having lower contrast.

The optimized detection model proved capable of detecting a substantial proportion of polynyas from observations not seen during training. For instance, the model located 79.5% of moderately difficult Nain polynyas ($\delta^{\mu}_{HH}\leq -3dB$), representing half of polynya observations across all imaging and environmental conditions over 3 years from a region not seen during training, with a precision of 26.2%. The same model achieved 79.0%

recall at 33.0% precision on the held-out Sanikiluaq evaluation scenes. Geometric data augmentation strategies (affine transformations) were found to play a critical role in model generalization between regions, counteracting the tendency of networks to overfit to region-specific, polynya-correlated land formations.

Ultimately, the variable separability and small sizes of landfast ice polynyas in medium-resolution, dual-polarized SAR imagery were found to limit the achievable performance of single-image detection models evaluated on all polynya targets under all imaging conditions. Even after optimization efforts, missed polynyas were found to consistently possess smaller sizes and lower contrast in backscatter intensity, while false positive predictions were often caused by lookalike patterns in landfast ice, e.g. at boundaries between smooth and rough ice. Nonetheless, under imaging and environmental conditions leading to clear polynya signatures in the SAR imagery (e.g., $\delta^{\mu}_{HH} \leq -5dB$), optimized models proved capable of consistently detecting sub-kilometer scale open water targets with high (e.g., $> 90\%$) recall and reasonable (e.g., $\geq 24\%$) precision.

Chapter 5

Generalizing Polynya Detection Networks from Sentinel-1 to Radarsat Constellation Mission SAR Imagery

Operationally-oriented SAR image classification systems that leverage the available diversities of satellite sensors and imaging modes with similar radar parameters are positioned to benefit from increased temporal resolution and more robust availability. For landfast ice polynya detection, increased temporal resolution of SAR observations offers an opportunity to produce more reliable estimates of polynya locations over a given time window, and to better track dynamic events such as polynya freeze-up or opening. The vulnerability in relying on a single SAR mission is illustrated in the polynya dataset described in chapter 3 by a significant reduction in Sentinel-1 (S1) imaging frequency for the 2022 season (e.g. Fig. 3.3). This reduction was in response to the failure of the Sentinel-1B satellite in December 2021. Meanwhile, the Canadian Space Agency's Radarsat Constellation Mission (RCM), comprised of three individual SAR satellites, has been providing dual-polarized C-band SAR imagery over Arctic coastal regions since 2019. The ability of landfast ice polynya detection models to operate on both S1 and RCM SAR imagery is thus of interest.

While a motivation for the calibration of SAR backscatter values *is* indeed sensor agnosticism at like frequencies and polarizations, different SAR beam modes, resolutions, noise characteristics, incidence angle ranges, and processing workflows available to SAR data users may introduce practical differences in the image products available. In the deep

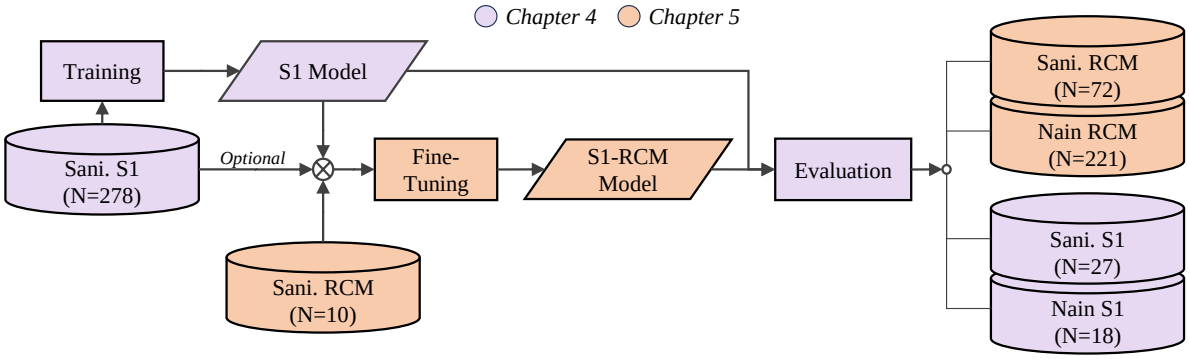

Figure 5.1: Conceptual overview of this chapter and its relation to methods and results presented in chapter 4.

learning context, few studies have investigated task-specific generalization characteristics of developed models between S1 and RCM imagery. In this chapter, the ability of CNN-based polynya detection networks trained on Sentinel-1 EW imagery to achieve similar performance on RCM ScanSAR 50-meter (SC50M) imagery is assessed. The evaluation performance of "original" S1-trained models developed in chapter 4 is compared to that of models additionally fine-tuned on a small amount of RCM imagery.

5.1 Data

5.1.1 RCM Imagery

Acquisition

RCM imagery was downloaded from Natural Resource Canada's Earth Observation Data Management System (EODMS) using the EODMS CLI library (https://github.com/eodms-sgdot/eodms-cli) version 3.3.1, both provided by Natural Resources Canada (NRCAN). Based on availability over both regions of interest, ScanSAR 50-meter resolution (SC50M) imagery ordered by the Canadian Ice Service was downloaded. While SC50M imagery was available over the Nain RoI for all three years of interest (2020-2022), the acquisition mode over the Sanikiluaq RoI transitioned in 2022 to the ScanSAR low-noise (SCLN) beam mode with 100m resolution. To provide consistency for this analysis, only

Table 5.1: Number of processed ScanSAR 50-meter (SC50M) images for each RoI by year

Year	N. SC50M Images	
	Sanikiluaq *(01/22 - 05/14)*	*Nain* *(03/19 - 05/14)*
2020	35	75
2021	47	76
2022	0	70

SC50M imagery available over the Sanikiluaq RoI for years 2020 and 2021 was downloaded. Image availability is summarized in table 5.1.

Processing

Processing of RCM products downloaded from EODMS to calibrated imagery was carried out using a processing codebase developed by the Canadian Ice Service and shared under a joint project agreement including the University of Waterloo research team. Processing steps included calibration to σ^0 and reprojection to UTM 17N and 21N zones for Sanikiluaq and Nain RoIs, respectively, using degree-2 polynomial interpolation to produce 40m pixels. No speckle filtering or incidence angle normalisation was performed, consistent with Sentinel-1 processing described in chapter 3. A noise subtraction algorithm accounting for scalloping effects in the azimuth dimension relevant to SC50M imagery was not available in the processor at the time of writing. No noise subtraction was performed, as 1D (range-only) subtraction does not account for these non-stationary characteristics in azimuth, e.g. near subswath boundaries. Since the S1 EW imagery used in this research *did* undergo noise subtraction, the resultant difference in noise characteristics between the RCM and S1 imagery used in this chapter is considered a variable to which model performance may be sensitive.

5.1.2 Polynya Labels

The same mapped polynya labels described in chapter 3 and used in chapter 4 were used for model evaluation on RCM imagery. These labels were verified and extended with RCM imagery for weeks where S1 coverage had not been available.

5.1.3 Polynya Signature Characterisation

Polynya signatures in RCM SC50M imagery were assessed following an identical methodology to that in chapter 3, using incidence angle-binned box plots and Jeffries-Matusita (JM) distances to evaluate separability in backscatter intensity on both polynya-local and full-scene scales. Fig. 5.2 illustrates the results of this analysis, showing patterns in backscatter intensity distribution and polynya/ice separability consistent with those observed in S1 imagery (Fig. 3.6). JM distances between polynya and ice classes are consistently below 0.5 in HH-polarized backscatter, and below 0.25 in HV. The RCM imagery includes observations at higher incidence angles than S1 (47°-55°), for which polynya/ice separability decreases in the Sanikiluaq dataset relative to SC50M observation at the S1 far range (e.g., 43°-47°). These findings reinforce the conclusion made in chapter 3 that polynya separability in dual-polarized backscatter intensity alone is not sufficient for the use of pixel-based classifiers.

5.2 S1-Trained Model Performance on RCM Imagery

5.2.1 Methods

The best-performing polynya detection model from chapter 4, summarized in table 5.2, was evaluated against the processed RCM imagery for both RoIs following three different protocols:

A. "Original" S1-trained models were evaluated directly on the RCM evaluation datasets;

B. S1-trained models were fine-tuned on a small number of RCM images from the Sanikiluaq RoI before being evaluated on the RCM and S1 evaluation datasets;

C. S1-trained models were fine-tuned on a mix of RCM and original S1 training images from the Sanikiluaq RoI before being evaluated on the RCM and S1 evaluation datasets.

All protocols followed the full-scene evaluation methodology described in section 4.2. Fine-tuning methods for protocols B and C are described subsequently.

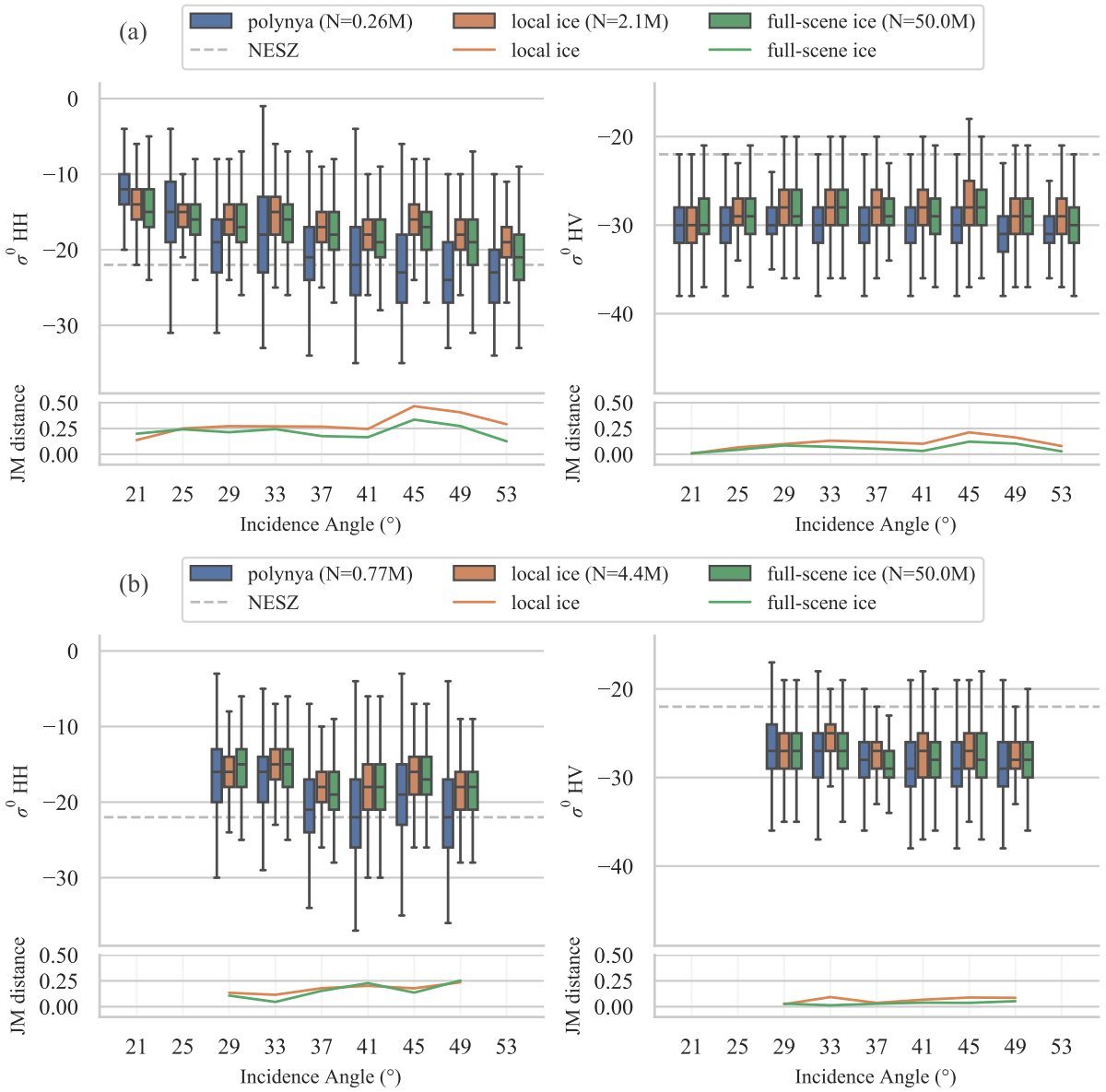

Figure 5.2: Backscatter intensity distributions in HH- and HV- polarized RCM SC50M SAR images of polynya pixels (left group) vs. polynya-local (center group) and large samples of all (right group) landfast ice pixels in the Sanikiluaq (a) and Nain (b) datasets. Box-plot whiskers indicate 1.5x the interquartile range. Jeffries-Matusita distances between polynyas and the two ice populations are calculated and plotted by the same incidence angle bins as those used for box-plotting.

Table 5.2: Model configuration and data inputs used for RCM generalization experiments described in this chapter. Design parameter notation can be reviewed in the referenced sections of chapter 4.

	Design Parameter	**Value**
Model (4.3)	Backbone network stages	S-3
	Input image scale	128x128
	RPN loss function & sampling	CE_{64}
Data (4.4)	Input Features	{HH, IA}
	Data augmentation scheme	$Affine_{p50}$
	Target polynya criterion	$\delta^{\mu}_{HH} \leq -3dB$

Fine-Tuning Data Selection & Sampling

Of the 82 RCM SC50M scenes available over the Sanikiluaq RoI, 72 were used for evaluation (all three protocols) and 10 were held out for fine-tuning (protocols B and C). The 10 scenes used for fine-tuning were selected in the same manner as the 27 S1 evaluation scenes for the Sanikiluaq RoI (section 4.2.2), providing representation across seasons, time of year, and target separability characteristics as measured by δ^{μ}_{HH} (equation 3.3). These images were patched into 30%-overlapping 128x128-pixel sub-images, and the negative patches were dynamically sampled during fine-tuning following the methodology described in section 4.2.4.

During fine-tuning for protocol C, both RCM and an equivalent number of S1 patches from the original training dataset were sampled. While these S1 patches had (likely, in the case of negatives) already been seen during training of the original model, they were re-used to constrain the re-parameterization of model to minimize loss on RCM imagery without compromising performance on S1 imagery.

Fine-Tuning Hyperparameter Selection

Fine-tuning was performed by first initializing FRCNN networks with the optimized model weights of chapter 4, then continuing end-to-end model training (updating all sub-networks jointly) for 10 epochs at an initial learning rate of 0.005 that decayed to 0.001 with a cosine schedule. These hyperparameters yielded similar convergence patterns and slightly better evaluation performance than training for 3 or 20 epochs, or the use of an initial learning rate of 0.01.

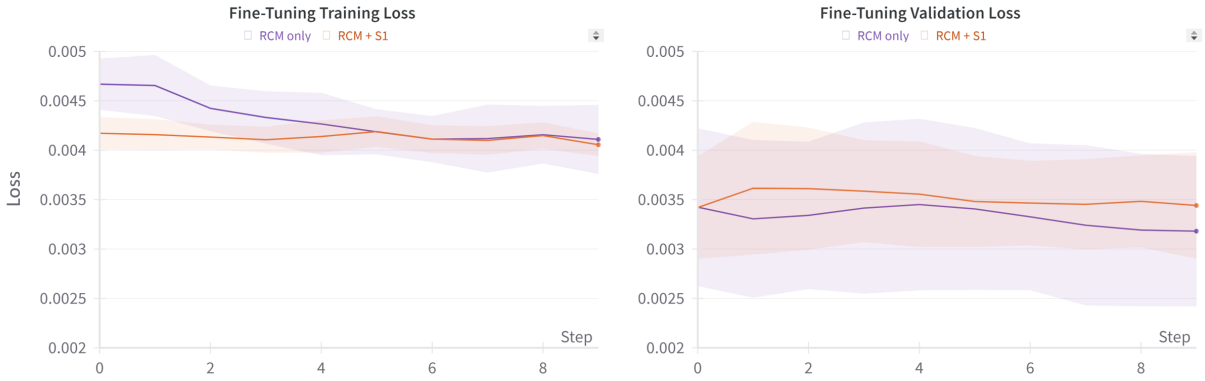

Figure 5.3: Fine-tuning training and validation loss for protocols B (RCM only) and C (RCM + S1). Each step corresponds to one epoch (a pass over the full fine-tuning dataset).

5.2.2 Results & Discussion

Fine-Tuning Convergence

Training and validation loss during fine-tuning is first inspected to indicate the degree to which model prediction patterns changed for protocols B and C. Loss is calculated the same way here as in section 4.3.2, where the total loss reported is a summation of the four sub-network losses described in section 2.4.3. As seen in Fig. 5.3, little additional model convergence was achieved in detecting $\delta^{\mu}_{HH} \leq -3dB$ polynya targets through fine-tuning on RCM imagery. While training loss decreased somewhat over the 10 epochs for protocol-B models fine-tuned only on RCM imagery, training loss for protocol-C models seeing both S1 and RCM imagery did not decrease, and validation loss for both protocols did not decrease significantly. These patterns indicate that predictions made for bounding boxes *sampled for loss* (limited negatives) by the original S1-trained model did not change significantly in response to fine-tuning on RCM imagery. This might be due to the similarity of features extracted by the "original" backbone network on S1 and RCM imagery, and the resultant efficacy of the existing decision boundaries within the RPN and RoI Head sub-networks on those features. In other words, the lack of further convergence during fine-tuning suggests a degree of baseline generalizability from S1 EW to RCM SC50M imagery for polynya detection without modification. Given the limited data used for fine-tuning and the nature of loss sampling, however, these observations must be supported with evaluation results before conclusions can be drawn.

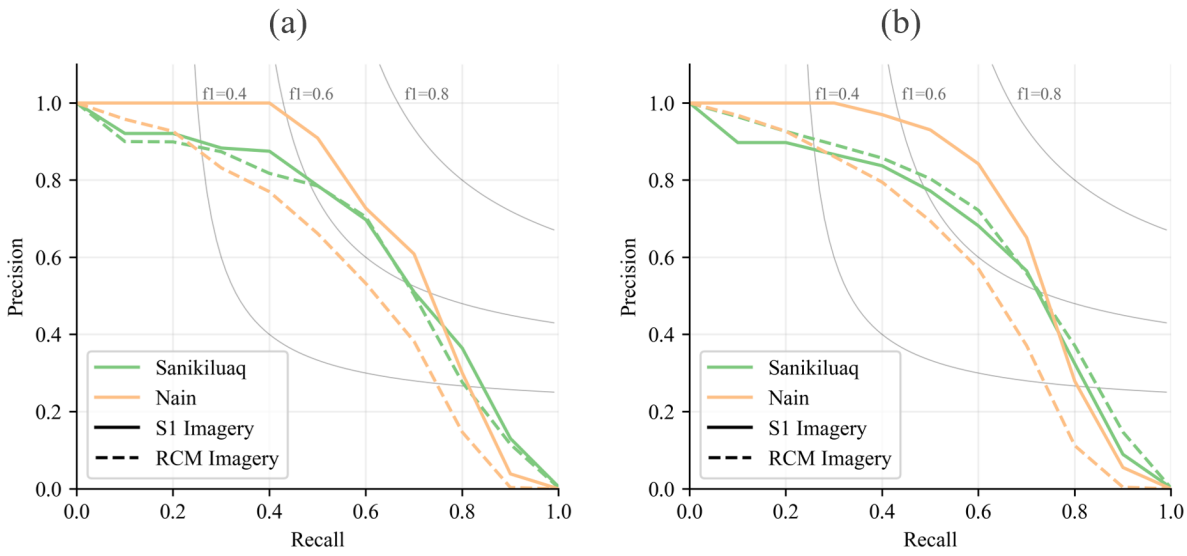

Figure 5.4: Precision-Recall curves for polynya detection models on S1 and RCM imagery, (a) without fine-tuning (protocol A) and (b) after fine-tuning with RCM and S1 image patches (protocol C).

Quantitative Evaluation Results

Evaluation metrics on S1 and RCM imagery for both RoIs are given for all protocols in table 5.3 and illustrated for protols A and C in Fig. 5.4. The "original" model (protocol A) performs similarly well at detecting polynyas in RCM imagery as S1 for the Sanikiluaq RoI, however performs worse on RCM imagery for the Nain dataset with an 11-point decrease in AP_{10} compared to S1. After fine-tuning this model only on RCM image patches (protocol B), performance on S1 imagery drops severely (17.6 points in AP_{10}) on the Sanikiluaq dataset and to a lesser degree (4.4 points) on the Nain dataset, while increasing modestly on RCM imagery for both RoIs (1.9 and 0.7 points, respectively). Fine-tuning on both RCM and S1 image patches (protocol C) not only alleviates the performance drop on S1 imagery for the Sanikiluaq RoI (4.4 points in AP_{10}), but also increases performance in S1 on the Nain dataset (2.2 points) in addition to increasing performance on RCM imagery for both datasets to a similar degree as protocol B (2.7 and 0.7 points).

While fine-tuning on both RCM and S1 imagery (protocol C) demonstrates the strongest quantitative results, the related performance gains on detecting polynyas in RCM imagery are modest compared to the original S1-trained model performing "out-of-the-box". This

Table 5.3: Evaluation metrics for original and fine-tuned models on RCM and S1 images over Sanikiluaq and Nain. For protocols B and C, mean performance metrics across 5 models cross-validated on the fine-tuning (FT) training dataset are reported in percentage points, followed by standard deviation in parentheses. Differences in mean AP relative to protocol A are also indicated. All metrics are calculated for τ_{IoU} = 0.1. Precision and recall values are given for τ_{score} = 0.8.

Protocol (Model)	**Eval. Dataset**	**AP**	**Recall**	**Precision**
A (Original)	Sanikiluaq S1	64.4	79.0	33.0
	Sanikiluaq RCM	62.5	78.8	30.1
	Nain S1	67.5	79.5	26.2
	Nain RCM	56.5	75.1	24.2
B (FT_{RCM})	Sanikiluaq S1	46.8 (0.46) −17.6	52.2 (2.5)	50.2 (2.0)
	Sanikiluaq RCM	64.4 (0.64) +1.9	71.2 (1.3)	48.0 (0.94)
	Nain S1	63.0 (1.3) −4.5	71.7 (0.72)	26.9 (3.3)
	Nain RCM	57.2 (1.9) +0.7	69.4 (1.4)	33.9 (2.2)
C (FT_{RCM+S1})	Sanikiluaq S1	60.0 (2.1) −4.4	70.9 (1.3)	40.5 (2.2)
	Sanikiluaq RCM	65.2 (0.8) +2.7	76.6 (0.9)	40.1 (1.4)
	Nain S1	69.7 (0.4) +2.2	75.7 (0.6)	28.0 (3.2)
	Nain RCM	57.2 (1.1) +0.7	72.1 (1.5)	29.1 (2.2)

supports the interpretation put forward in convergence analysis that the S1-trained model generalizes relatively well to SC50M imagery without any modification. However, the significant drop in AP_{10} on the Nain RCM dataset (12.5 points lower than the Nain S1 dataset for protocol C) prompts investigation.

Qualitative Evaluation Results

Fig. 5.5 shows an example of protocol-C model predictions on an RCM SC50M evaluation image from the Nain dataset. All target polynyas were detected precisely in this scene, despite the model having never seen the Nain region during training or fine-tuning in either S1 or RCM imagery. However, the scene also contains several FP predictions that occur near land and often in areas of radar shadow. These shadows occur at high incidence angle (IA), i.e., oblique viewing geometries, when vertical terrain blocks radar illumination of adjacent down-range surfaces [14]. As illustrated by Fig. 5.6a, the RCM SC50M imagery over the Nain RoI has a higher IA range and a greater density of images at these high IAs

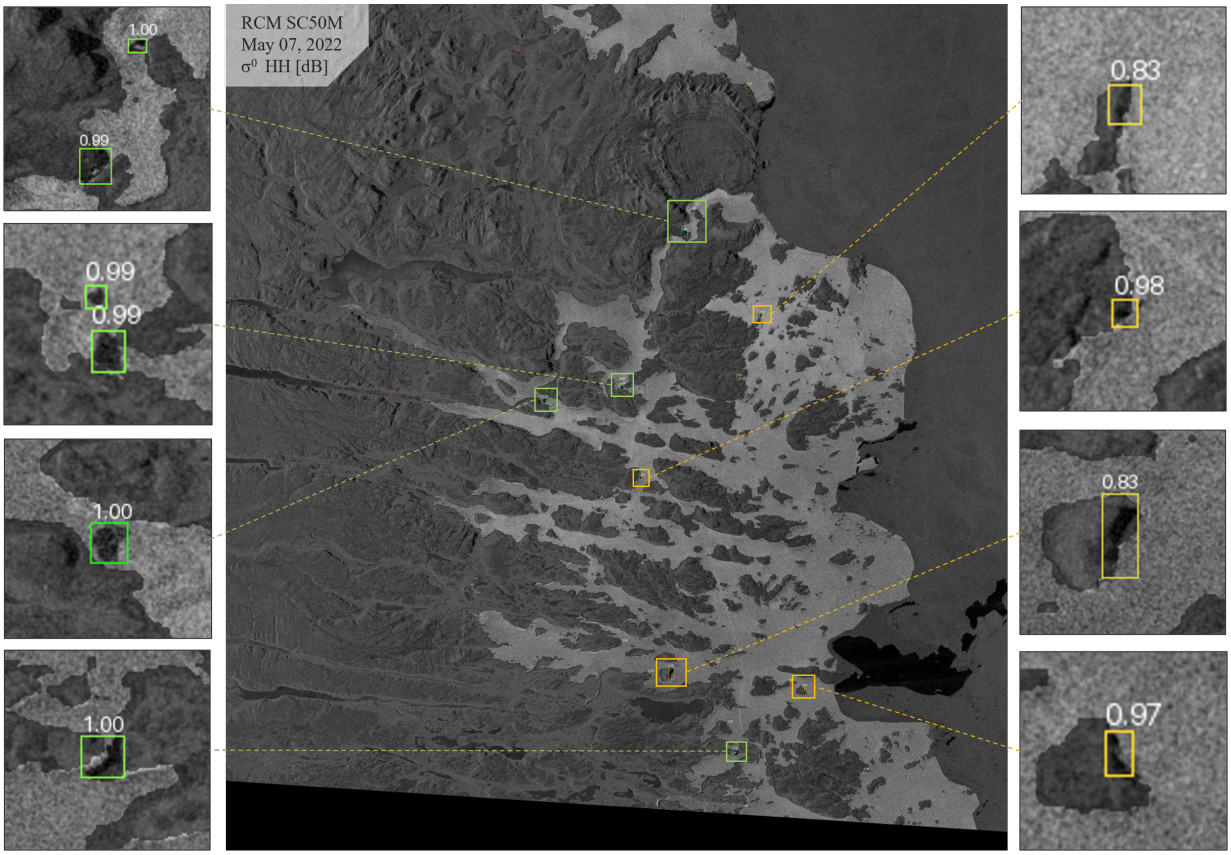

Figure 5.5: Example of fine-tuned model predictions on an SC50M RCM image captured May 7, 2022, over the Nain RoI. The image has an incidence angle range of 39° to 44° (left to right). True positive and false positive predictions are shown in the left and right columns, respectively. Predicted confidence scores are indicated above each bounding box.

compared to the S1 imagery. A higher number of false positives is seen for images at far range, and many of these false positives were observed to correspond to terrain shadow. This is believed to be a significant contributing factor to the reduced performance observed on the Nain RCM dataset. The impact on the Sanikiluaq dataset is believed to be less pronounced in part due to the development of polynyas in areas that are less shadowed by terrain in the east-west direction, i.e., approximately the range dimension of polar-orbiting SAR satellites.

While radar shadows from land imaged at high incidence angles are mostly covered by the dataset's landmask, it is difficult to discard FP predictions that partially cover land

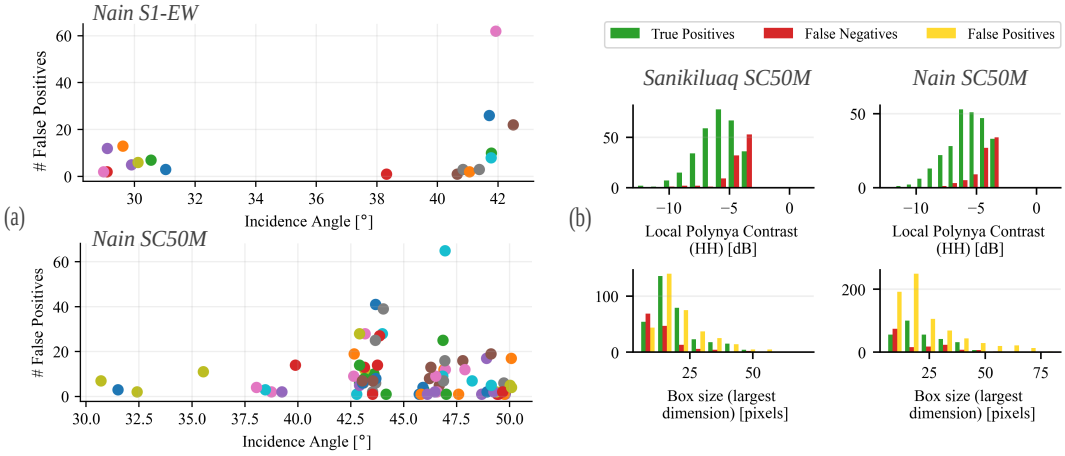

Figure 5.6: Distributions of fine-tuned model predictions over evaluation image and target parameters. (a) Number of FP predictions per scene against each scene's median incidence angle; (b) categorized predictions by local polynya contrast (δ^{μ}_{HH}) and bounding box size.

without discarding TP boxes that also often do so. In the evaluation approach used thus far, only boxes which are *centered* on land are discarded. It may be possible to discard a greater number of FP without compromising TPs through more detailed post-processing, e.g. by analyzing the fractional land coverage of FP versus TP boxes and identifying a threshold to use for filtering. More generally, for regions with significant vertical terrain features and possible shadowing of polynyas, it may be worth considering the use of radiometric terrain correction based on a digital elevation model (DEM) to alleviate the impact of geometric distortions on the SAR backscatter signal. For instance, calibration to γ^0 rather than σ^0 normalizes measured backscatter to the observed surface in the SAR geometry based on a digital elevation model, rather than to a "flat" surface based on an ellipsoid model of Earth [126]. Even such calibration cannot, however, reconstruct backscatter over completely shadowed areas. Another, more limiting, solution might be to flag and discard model predictions under conditions with a high risk of shadow, such as observations near land at high incidence angles (e.g. above 40° based on Fig. 5.6a) or more precisely by estimating shadowed areas with a DEM and known imaging geometry and masking model outputs over these regions.

Consistent with observations of model performance on the S1 evaluation datasets, both the "original" S1-trained model and the fine-tuned model missed polynyas with small sizes (< 15 pixels) and low contrast ($\delta^{\mu}_{HH} < -5dB$) in the RCM imagery for both RoIs. This is illustrated by the distributions in Fig. 5.6b.

5.3 Chapter Conclusions

This chapter investigated the ability of polynya detection models trained on Sentinel-1 EW imagery to generalize to RCM ScanSAR 50-meter imagery, and assessed the utility of fine-tuning on a small amount of data to do so. The following conclusions are made:

1. Models trained on Sentinel-1 EW imagery maintained similar performance in detecting polynyas in RCM SC50M imagery without any modification;
2. Fine-tuning jointly on a small number of RCM images and previously-seen S1 images offered modest performance improvements on RCM without degrading performance on S1;
3. False positive predictions were observed to occur over areas of terrain-induced radar shadow, reducing model precision at high incidence angles ($\geq 40°$) near land formations with significant elevation.

Overall, the generalization performance of polynya detection models between dual-polarized S1 and RCM image products indicates a level of feature extraction and classification invariance to sensor noise characteristics, and supports their utility for multi-sensor application. Such application unlocks significant increases in temporal resolution and reduces availability risk of derivative information products.

Chapter 6

Conclusion

6.1 Summary

This thesis has explored the use of CNN-based object detection networks for detecting landfast ice polynyas relevant to coastal communities from dual-polarized SAR imagery. Chapter 3 described the synthesis of datasets of SAR polynya observations across several seasons, and demonstrated the challenging signatures of small, sparse polynya targets in landfast ice whose separability varies with imaging incidence angle, windspeed, and seasonality. Chapter 4 presented the development and evaluation of a Faster-RCNN-based object detection network for landfast ice polynya detection. The sensitivity of detection model performance to key design elements inspired by advancements in the natural-scene small object detection field was investigated. A lightweight 3-stage feature extraction network was found to produce favourable results for polynya detection, while input image upsampling and alternative balanced loss functions were not found to provide utility. The optimized model was evaluated under several data scenarios which demonstrated the importance of geometric data augmentation techniques for inter-region model generalization and the low utility of noise-dominated HV backscatter over landfast ice for ice/water discrimination. In chapter 5, the capability of detection models trained on S1 SAR imagery to detect polynyas in RCM imagery with similar radar parameters was evaluated. Detection performance transferred well without modification, with similar qualitative prediction patterns in RCM as in S1. Decreases in quantitative evaluation metrics in the RCM dataset over the Nain RoI reflected the susceptibility of the model to terrain shadow effects adjacent to landfast ice occurring at high incidence angle ranges.

Overall, optimized detection models were able to retrieve a significant portion of polynya

targets with moderate precision characteristics. For instance, on an evaluation dataset of Sentinel-1 polynya observations from the completely held-out Nain region, models detected ∼80% of target polynyas representing half of observations under all imaging and environmental conditions, at ∼25% precision. For both regions and SAR sensors, missed polynyas consistently possessed lower separability in HH backscatter intensity (local contrast less than 5dB) and smaller sizes (less than 600m), while false predictions often occurred at the boundaries of smooth and rough ice or near land.

While the diverse optimization efforts and comprehensive analyses of model predictions described in this thesis yield models with reasonable performance, they simultaneously point to fundamental limits in the sensitivity and specificity that can be achieved from single-image predictions with medium-resolution, dual-polarized C-band SAR. While further single-image detection network optimization is not expected to yield substantial benefits given these limitations, other approaches are suggested for the improvement of landfast ice polynya detection models developed in this thesis.

6.2 Future Work

6.2.1 Leveraging Polynya Spatio-Temporal Patterns

Given their persistent and recurrent nature, polynyas can be viewed as *spatio-temporal* features, i.e., possessing signatures in both space and time. The detection models developed in this thesis have only operated on the former of these dimensions. In contrast, models that consider a series of consecutive SAR images may be able to exploit differences in the temporal evolution of polynya versus landfast ice backscatter signatures and thus widen the decision boundary between the two classes. Indeed, the use of learned spatial and temporal features for sea ice classification has been investigated previously, e.g. with separate CNN and LSTM models in [129] and with a combined convolutional LSTM model in [128]. Key challenges in the application of such models to polynya detection include the non-uniform timesteps between SAR observations (e.g. Fig. 3.3) as well as the identified variability in polynya separability between images.

An alternative spatio-temporal approach might operate on model *predictions* rather than on images directly. On a region-by-region basis, the belief that a polynya exists at a given location, as derived from previous predictions, offers a strong prior regarding the probability that a polynya currently exists at that location. Parallels to video object detection and tracking emerge from this formulation; for instance, the spatio-temporal small

object detection network proposed by Bosquet et. al [11] correlates model detections between video frames with the goal of strengthening true positive detections while discarding false positives. Key differences in the polynya detection context are that polynyas do not move significantly between frames, but that their ability to be detected may vary between observations depending on imaging and environmental conditions.

6.2.2 Leveraging Multifrequency & Next-Gen SAR

Several analyses presented in this thesis have made it clear that a fundamental challenge encountered in this research is the frequent overlap of landfast ice and open water signatures in dual-polarized C-band SAR backscatter intensity, as well as the small sizes of polynyas in the medium-resolution imagery products available on an operational basis over Arctic coastal regions. Meanwhile, the utility of L-band SAR has received significant attention for sea ice and water classification, and studies combining both C- and L-band imagery have shown the complementary potential of such multifrequency analysis, e.g. [97, 95]. In particular, L-band backscatter has been found to have utility for mapping deformed sea ice features [26] as well as identification of early-stage ice types [97]. These sensitivities could perhaps be leveraged to identify areas of rough ice that contribute to false positives in C-band model predictions, or increase the contrast of polynyas with thin ice cover otherwise missed in C-band. Future availability of L-band imagery provided by missions such as NISAR [68] and ROSE-L [33] represents a possible avenue for advancements in polynya and other ice hazard detection systems.

More broadly, continued improvements in SAR sensor resolutions and noise characteristics will benefit applications like landfast ice hazard detection operating at the limits of both. Ultimately, the provision of community-relevant SAR-based sea ice information products depends on the regular coverage of northern coastal regions, requiring ongoing investments from data providers, governments, and users.

References

[1] IPCC, 2019: Summary for Policymakers. Technical report, Cambridge University Press, 2019.

[2] Fatih Cagatay Akyon, Sinan Onur Altinuc, and Alptekin Temizel. Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 966–970, Bordeaux, France, October 2022. IEEE.

[3] Shun-ichi Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5):185–196, June 1993.

[4] Yancheng Bai, Yongqiang Zhang, Mingli Ding, and Bernard Ghanem. SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, volume 11217, pages 210–226. Springer International Publishing, Cham, 2018. Series Title: Lecture Notes in Computer Science.

[5] D. G. Barber and Ellsworth F. LeDrew. SAR Sea Ice Discrimination Using Texture Statistics: A Multivariate Approach. *Photogrammetric Engineering & Remote Sensing*, 57(4):385–395, April 1991.

[6] D. G. Barber and J. Yackel. The physical, radiative and microwave scattering characteristics of melt ponds on Arctic landfast sea ice. *International Journal of Remote Sensing*, 20(10):2069–2090, January 1999.

[7] D.G. Barber and R.A. Massom. Chapter 1 The Role of Sea Ice in Arctic and Antarctic Polynyas. In *Elsevier Oceanography Series*, volume 74, pages 1–54. Elsevier, 2007.

[8] L. Beaulieu, A. Arreak, R. Holwell, S. Dicker, O. Qamanirq, L. Moorman, K. Wilson, R. Segal, S. Crichton, and T. Bell. Indigenous self-determination in cryospheric

science: The Inuit-led Sikumik Qaujimajjuti ("tools to know how the ice is") program in Inuit Nunangat, Canada. *Frontiers in Earth Science*, 11:1076774, February 2023.

[9] Trevor Bell, Robert Briggs, Ralf Bachmayer, and Shuo Li. Augmenting Inuit knowledge for safe sea-ice travel - The SmartICE information system. In *2014 Oceans - St. John's*, pages 1–9, St. John's, NL, September 2014. IEEE.

[10] Breanna Bishop. *Respecting ontology: Documenting Inuit knowledge of coastal oceanography in Nunatsiavut*. PhD thesis, Dalhouse University, December 2019.

[11] Brais Bosquet, Manuel Mucientes, and Víctor M. Brea. STDnet-ST: Spatio-temporal ConvNet for small object detection. *Pattern Recognition*, 116:107929, August 2021.

[12] Hugo Boulze, Anton Korosov, and Julien Brajard. Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks. *Remote Sensing*, 12(13):2165, July 2020.

[13] Girija Kalyani Burada, Adrian McDonald, James Renwick, and Ben Jolly. Delineating Polynya Area Using Active and Passive Microwave Sensors for the Western Ross Sea Sector of Antarctica. *Remote Sensing*, 15(10):2545, May 2023.

[14] Canada Centre for Mapping and Earth Observation. Fundamentals of Remote Sensing, August 2019.

[15] Natalie Carter, Charlotte Buttle, Gita Ljubicic, Regena Sinclair, and Emmelie Paquette. Results of a community survey on environmental forecasting uses and needs. Technical report, December 2023.

[16] Natalie Carter, Charlotte Buttle, Gita Ljubicic, Regena Sinclair, and Emmelie Paquette. Results of a community survey on environmental forecasting uses and needs: Sanikiluaq, Nunavut. Technical report, December 2023.

[17] Natalie Ann Carter and Gita Ljubicic. Weather and Society in Inuit Nunangat: User Needs, Service Provider Efforts, and Opportunities for the Future. Technical report, Hamilton, ON, October 2022.

[18] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open MMLab Detection Toolbox and Benchmark. 2019. Publisher: arXiv Version Number: 1.

[19] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. 2014. Publisher: arXiv Version Number: 4.

[20] Xinwei Chen, Muhammed Patel, Linlin Xu, Yuhao Chen, K. Andrea Scott, and David A. Clausi. Weakly Supervised Learning for Pixel-Level Sea Ice Concentration Extraction Using AI4Arctic Sea Ice Challenge Dataset. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[21] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva. SAR image despeckling through convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5438–5441, Fort Worth, TX, July 2017. IEEE.

[22] David A. Clausi. Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery. *Atmosphere-Ocean*, 39(3):183–194, September 2001.

[23] Mohammed Dabboor, Stephen Howell, Mohammed Shokr, and John Yackel. The Jeffries–Matusita distance for the case of complex Wishart distribution as a separability criterion for fully polarimetric SAR data. *International Journal of Remote Sensing*, 35(19), October 2014.

[24] Aiguo Dai, Dehai Luo, Mirong Song, and Jiping Liu. Arctic amplification is caused by sea-ice loss under increasing CO2. *Nature Communications*, 10(1):121, January 2019.

[25] Jackie Dawson, Winfried Hoke, Machiel Lamers, Daniela Liggett, Gita Ljubicic, Brian Mills, Emma Stewart, and Rick Thoman. Navigating Weather, Water, Ice and Climate Information for Safe Polar Mobilities. Technical report, World Meteorological Organization, Geneva, 2017.

[26] W. Dierking and T. Busche. Sea ice monitoring by L-band SAR: an assessment based on literature and comparisons of JERS-1 and ERS-1 imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4):957–970, April 2006.

[27] Wolfgang Dierking. Sea Ice Monitoring by Synthetic Aperture Radar. *Oceanography*, 26(2), June 2013.

[28] Rohit Venkata Sai Dulam, Kelsey Kaplan, and Chandra Kambhamettu. Deep Learning-Based Sea Ice Lead Detection from WorldView and Sentinel SAR Imagery.

In Jean-Jacques Rousseau and Bill Kapralos, editors, *ICPR 2022 International Workshops and Challenges*, volume 13645, pages 330–342, Cham, 2023. Springer Nature Switzerland. Series Title: Lecture Notes in Computer Science.

[29] ESRI. Arcgis Pro, 2023.

[30] European Space Agency. Sentinel-1 SAR User Guide.

[31] European Space Agency. Sentinel-2 MSI User Guide.

[32] European Space Agency. Sentinel-1 Product Definition, March 2016.

[33] European Space Agency. ROSE-L (Radar Observing System for Europe in L-band), September 2023.

[34] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[35] M. Fleming. Sea Ice in the Belcher Islands, Nunavut, Canada. 2010. Publisher: NSIDC.

[36] J. D. Ford, D. Clark, T. Pearce, L. Berrang-Ford, L. Copland, J. Dawson, M. New, and S. L Harper. Changing access to ice, land and water in Arctic communities. *Nature Climate Change*, 9(4):335–339, April 2019.

[37] James D. Ford, Graham McDowell, and Tristan Pearce. The adaptation challenge in the Arctic. *Nature Climate Change*, 5(12):1046–1053, December 2015.

[38] James D Ford, Barry Smit, Johanna Wandel, Mishak Allurut, Kik Shappa, Harry Ittusarjuat, and Kevin Qrunnut. Climate change in the Arctic: current and future vulnerability in two Inuit communities in Canada. *The Geographical Journal*, 174(1):45–62, March 2008.

[39] Shari Fox Gearheard, editor. *The meaning of ice: people and sea ice in three Arctic communities*. International Polar Institute Press, Hanover, 2017. OCLC: 1022589824.

[40] T Geldsetzer and J J Yackel. Sea ice type and open water discrimination using dual co-polarized C-band SAR. *Canadian Journal of Remote Sensing*, 35(1):73–84, January 2009.

[41] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D. Cubuk, Quoc V. Le, and Barret Zoph. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. 2020. Publisher: arXiv Version Number: 2.

[42] N. Ghoggali and F. Melgani. Automatic Ground-Truth Validation With Genetic Algorithms for Multispectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2172–2181, July 2009.

[43] Jagvijay P.S. Gill, John J. Yackel, Torsten Geldsetzer, and M. Christopher Fuller. Sensitivity of C-band synthetic aperture radar polarimetric parameters to snow thickness over landfast smooth first-year sea ice. *Remote Sensing of Environment*, 166:34–49, September 2015.

[44] Ross Girshick. Fast R-CNN. 2015. Publisher: arXiv Version Number: 2.

[45] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. 2013. Publisher: arXiv Version Number: 5.

[46] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[47] Google Earth Engine. Sentinel-1 Algorithms. Technical report, March 2024.

[48] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, December 2017.

[49] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour, 2017. Version Number: 2.

[50] Jörg Haarpaintner. The Storfjorden polynya: ERS-2 SAR observations and overview. *Polar Research*, 18(2):175–182, January 1999.

[51] Dorothy K. Hall. Remote Sensing of Snow and Ice Using Imaging Radar. In Robert A. Ryerson, Floyd Henderson, and Anthony J. Lewis, editors, *Principles & Applications of Imaging Radar*, volume 2 of *Manual of Remote Sensing*, pages 677–698. 3 edition, 1998.

[52] Martti Hallikainen and Dale P. Winebrenner. The physical basis for sea ice remote sensing. In Frank D. Carsey, editor, *Geophysical Monograph Series*, volume 68, pages 29–46. American Geophysical Union, Washington, D. C., 1992.

[53] Wei Han, Jia Chen, Lizhe Wang, Ruyi Feng, Fengpeng Li, Lin Wu, Tian Tian, and Jining Yan. Methods for Small, Weak Object Detection in Optical High-Resolution Remote Sensing Images: A survey of advances and challenges. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):8–34, December 2021.

[54] Yanling Han, Yekun Liu, Zhonghua Hong, Yun Zhang, Shuhu Yang, and Jing Wang. Sea Ice Image Classification Based on Heterogeneous Data Fusion and Deep Learning. *Remote Sensing*, 13(4):592, February 2021.

[55] Joel P. Heath, H. Grant Gilchrist, and Ronald C. Ydenberg. Regulation of stroke pattern and swim speed across a range of current velocities: diving by common eiders wintering in polynyas in the Canadian Arctic. *Journal of Experimental Biology*, 209(20):3974–3983, October 2006.

[56] Floyd Henderson and Anthony J. Lewis. Principles and Applications of Imaging Radar: Introduction. In Robert A. Ryerson, editor, *Principles and Applications of Imaging Radar*, volume 2 of *Manual of Remote Sensing*, pages 1–7. John Wiley & Sons, Inc., 3 edition, 1998.

[57] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia De Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, July 2020.

[58] T. Hollands, V. Haid, W. Dierking, R. Timmermann, and L. Ebner. Sea ice motion and open water area at the Ronne Polynia, Antarctica: Synthetic aperture radar observations versus model results: SEA ICE MOTION AND OPEN WATER AREA AT THE RONNE POLYNIA. *Journal of Geophysical Research: Oceans*, 118(4):1940–1954, April 2013.

[59] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989.

[60] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-Excitation Networks. 2017. Publisher: arXiv Version Number: 4.

[61] Wenjun Huang, Anzhu Yu, Qing Xu, Qun Sun, Wenyue Guo, Song Ji, Bowei Wen, and Chunping Qiu. Sea Ice Extraction via Remote Sensing Imagery: Algorithms, Datasets, Applications and Challenges. *Remote Sensing*, 16(5):842, February 2024.

[62] Zhongling Huang, Zongxu Pan, and Bin Lei. What, Where, and How to Transfer in SAR Target Recognition Based on Deep CNNs. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4):2324–2336, April 2020.

[63] Inuit Circumpolar Council - Canada. The Sea Ice is Our Highway: An Inuit Perspective on Transportation in the Arctic. Technical report, March 2008.

[64] Inuit Circumpolar Council - Canada. The Sea Ice Never Stops: Circumpolar Inuit Reflections on Sea Ice Use and Shipping in Inuit Nunaat. Technical report, December 2014.

[65] Larysa Istomina, Henrik Marks, Marcus Huntemann, Georg Heygster, and Gunnar Spreen. Improved cloud detection over sea ice and snow during Arctic summer using MERIS data. *Atmospheric Measurement Techniques*, 13(12):6459–6472, December 2020.

[66] Steven A. Jacobson. *Yupik Eskimo dictionary.* Alaska Native Language Center, Univ. of Alaska, Fairbanks, Ak, 2012.

[67] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, editors. *An introduction to statistical learning: with applications in R.* Number 103 in Springer texts in statistics. Springer, New York, 2013. OCLC: ocn828488009.

[68] Jet Propulsion Laboratory. NISAR: NASA-ISRO SAR Mission, 2023.

[69] Mingzhe Jiang, Linlin Xu, and David A. Clausi. Sea Ice–Water Classification of RADARSAT-2 Imagery Based on Residual Neural Networks (ResNet) with Regional Pooling. *Remote Sensing*, 14(13):3025, June 2022.

[70] T. Kailath. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Transactions on Communications*, 15(1):52–60, February 1967.

[71] Joe Karetak, Frank J. Tester, and Shirley Tagalik, editors. *Inuit Qaujimajatuqangit: what Inuit have always known to be true*. Fernwood Publishing, Halifax ; Winnipeg, 2017. OCLC: ocn987796398.

[72] J. Karvonen, M. Simila, and M. Makynen. Open Water Detection From Baltic Sea Ice Radarsat-1 SAR Imagery. *IEEE Geoscience and Remote Sensing Letters*, 2(3):275–279, July 2005.

[73] J.A. Karvonen. Baltic Sea ice SAR segmentation and classification using modified pulse-coupled neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1566–1574, July 2004.

[74] Salman Khaleghian, Habib Ullah, Thomas Kræmer, Nick Hughes, Torbjørn Eltoft, and Andrea Marinoni. Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks. *Remote Sensing*, 13(9):1734, April 2021.

[75] Mate Kisantal, Zbigniew Wojna, Jakub Murawski, Jacek Naruniec, and Kyunghyun Cho. Augmentation for small object detection. 2019. Publisher: [object Object] Version Number: 1.

[76] Ryan Kruk, M. Christopher Fuller, Alexander S. Komarov, Dustin Isleifson, and Ian Jeffrey. Proof of Concept for Sea Ice Stage of Development Classification Using Deep Learning. *Remote Sensing*, 12(15):2486, August 2020.

[77] Gita J. Laidler, Tom Hirose, Mark Kapfer, Theo Ikummaq, Eric Joamie, and Pootoogoo Elee. Evaluating the Floe Edge Service: how well can SAR imagery address Inuit community concerns around sea ice change and travel safety?: Evaluating the Floe Edge Service. *The Canadian Geographer / Le Géographe canadien*, 55(1):91–107, March 2011.

[78] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.

[79] Steven Leigh, Zhijie Wang, and David A. Clausi. Automated Ice–Water Classification Using Dual Polarization SAR Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5529–5539, September 2014.

[80] Anthony J. Lewis, Floyd Henderson, and Derrold W. Holcomb. Radar Fundamentals: The Geoscience Perspective. In Robert A. Ryerson, Floyd Henderson, and Anthony J. Lewis, editors, *Principles & Applications of Imaging Radar*, volume 2 of *Manual of Remote Sensing*, pages 131–176. 3 edition, 1998.

[81] Jianan Li, Xiaodan Liang, Yunchao Wei, Tingfa Xu, Jiashi Feng, and Shuicheng Yan. Perceptual Generative Adversarial Networks for Small Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1951–1959, Honolulu, HI, July 2017. IEEE.

[82] Wenwen Li, Chia-Yu Hsu, and Marco Tedesco. Advancing Arctic sea ice remote sensing with AI and deep learning: now and future. preprint, Sea ice/Remote Sensing, January 2024.

[83] Zeyu Liang, Xiaoping Pang, Qing Ji, Xi Zhao, Guoyuan Li, and Yizhuo Chen. An Entropy-Weighted Network for Polar Sea Ice Open Lead Detection From Sentinel-1 SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[84] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection, April 2017. arXiv:1612.03144 [cs].

[85] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal Loss for Dense Object Detection. 2017. Publisher: arXiv Version Number: 2.

[86] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common Objects in Context. 2014. Publisher: [object Object] Version Number: 3.

[87] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*, 128(2):261–318, February 2020.

[88] Xiaomin Liu, Tiantian Feng, Xiaofan Shen, and Rongxing Li. PMDRnet: A Progressive Multiscale Deformable Residual Network for Multi-Image Super-Resolution of AMSR2 Arctic Sea Ice Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022.

[89] Xiaomin Liu, Tiantian Feng, Yushi Yang, and Rongxing Li. Characterization of north Greenland polynyas with super-resolved passive microwave sea ice concentration. *GIScience & Remote Sensing*, 61(1):2300222, January 2024.

[90] Yang Liu, Peng Sun, Nickolas Wergeles, and Yi Shang. A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, 172:114602, June 2021.

[91] Johannes Lohse, Anthony P. Doulgeris, and Wolfgang Dierking. Mapping sea-ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle. *Annals of Glaciology*, 61(83):260–270, December 2020.

[92] Johannes Lohse, Anthony P. Doulgeris, and Wolfgang Dierking. Incident Angle Dependence of Sentinel-1 Texture Features for Sea Ice Classification. *Remote Sensing*, 13(4):552, February 2021.

[93] M. Meredith, M. Sommerkorn, S. Cassota, C. Derksen, A. Ekaykin, A. Hollowed, G. Kofinas, A. Mackintosh, J. Melbourne-Thomas, M.M.C. Muelbert, G. Ottersen, H. Pritchard, and E.A.G Schuur. *Polar Regions*. IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. Cambridge University Press, 1 edition, 2019.

[94] Grant J. Macdonald, Stephen F. Ackley, Alberto M. Mestas-Nuñez, and Adrià Blanco-Cabanillas. Evolution of the dynamics, area, and ice production of the Amundsen Sea Polynya, Antarctica, 2016–2021. *The Cryosphere*, 17(2):457–476, February 2023.

[95] Grant John Macdonald, Randall K. Scharien, Parnian Rezania, Christian Haas, Stephen Howell, and Alexander S. Komarov. Assessing Sea Ice Surface Roughness and Thickness in the Canadian Arctic Archipelago Using C-band and L-band Synthetic Aperture Radar (SAR). In *AGU '22*, volume 2022, pages C11C–01, December 2022. Conference Name: AGU Fall Meeting Abstracts ADS Bibcode: 2022AGUFM.C11C..01M.

[96] Mallik S. Mahmud, Torsten Geldsetzer, Stephen E. L. Howell, John J. Yackel, Vishnu Nandan, and Randall K. Scharien. Incidence Angle Dependence of HH-Polarized C- and L-Band Wintertime Backscatter Over Arctic Sea Ice. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11):6686–6698, November 2018.

[97] Mallik S. Mahmud, Vishnu Nandan, Suman Singha, Stephen E.L. Howell, Torsten Geldsetzer, John Yackel, and Benoit Montpetit. C- and L-band SAR signatures of Arctic sea ice during freeze-up. *Remote Sensing of Environment*, 279:113129, September 2022.

[98] Marko Makynen and Juha Karvonen. Incidence Angle Dependence of First-Year Sea Ice Backscattering Coefficient in Sentinel-1 SAR Imagery Over the Kara Sea. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6170–6181, November 2017.

[99] M.P. Makynen, A.T. Manninen, M.H. Simila, J.A. Karvonen, and M.T. Hallikainen. Incidence angle dependence of the statistical properties of C-band HH-polarization

backscattering signatures of the Baltic Sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2593–2605, December 2002.

[100] Walter N. Meier, Julienne Stroeve, and Florence Fetterer. Whither Arctic sea ice? A clear signal of decline regionally, seasonally and extending beyond the satellite record. *Annals of Glaciology*, 46:428–434, 2007.

[101] Benoit Montpetit, Benjamin Deschamps, Joshua King, and Jason Duffe. Assessing the Parameterization of RADARSAT-2 Dual-polarized ScanSAR Scenes on the Accuracy of a Convolutional Neural Network for Sea Ice Classification: Case Study over Coronation Gulf, Canada. *Canadian Journal of Remote Sensing*, 49(1):2247091, January 2023.

[102] M. A. Morales Maqueda, A. J. Willmott, and N. R. T. Biggs. Polynya Dynamics: a Review of Observations and Modeling. *Reviews of Geophysics*, 42(1), March 2004.

[103] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1):6–43, March 2013.

[104] Vishnu Nandan, Randall Scharien, Torsten Geldsetzer, Mallik Mahmud, John J. Yackel, Tanvir Islam, Jagvijay P.S. Gill, Mark C. Fuller, Grant Gunn, and Claude Duguay. Geophysical and atmospheric controls on Ku-, X- and C-band backscatter evolution from a saline snow cover on first-year sea ice from late-winter to pre-early melt. *Remote Sensing of Environment*, 198:425–441, September 2017.

[105] Junhyug Noh, Wonho Bae, Wonhee Lee, Jinhwan Seo, and Gunhee Kim. Better to Follow, Follow to Be Better: Towards Precise Supervision of Feature Super-Resolution for Small Object Detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9724–9733, Seoul, Korea (South), October 2019. IEEE.

[106] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019. Publisher: arXiv Version Number: 1.

[107] Tristan Pearce, Barry Smit, Frank Duerden, James D. Ford, Annie Goose, and Fred Kataoyak. Inuit vulnerability and adaptive capacity to climate change in Ulukhaktok, Northwest Territories, Canada. *Polar Record*, 46(2):157–177, April 2010.

[108] Pikialasorsuaq Commission. People of the Ice Bridge: The Future of the Pikialasorsuaq. Technical report, November 2017.

[109] Paul-Louis Pröve. MobileNetV2: Inverted Residuals and Linear Bottlenecks, April 2018.

[110] Zhixin Qi, Anthony Gar-On Yeh, Xia Li, and Zheng Lin. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. *Remote Sensing of Environment*, 118:21–39, March 2012.

[111] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. 2015. Publisher: [object Object] Version Number: 5.

[112] J. Francis Reintjes and Geoffrey T. Coate. *Principles of Radar*. McGraw-Hill Book Company Inc., 3 edition, 1952.

[113] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 2015. Publisher: arXiv Version Number: 3.

[114] Yibin Ren, Xiaofeng Li, Xiaofeng Yang, and Huan Xu. Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[115] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400–407, September 1951.

[116] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986.

[117] Stein Sandven, Gunnar Spreen, Georg Heygster, Fanny Girard-Ardhuin, Sinéad L. Farrell, Wolfgang Dierking, and Richard A. Allard. Sea Ice Remote Sensing—Recent Developments in Methods and Climate Data Sets. *Surveys in Geophysics*, April 2023.

[118] Randall Kenneth Scharien and Sasha Nasonova. Incidence Angle Dependence of Texture Statistics From Sentinel-1 HH-Polarization Images of Winter Arctic Sea Ice. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[119] Peter Schledermann. Polynyas and Prehistoric Settlement Patterns. *Arctic*, pages 292–302, 1980.

[120] Alessandro Sebastianelli, Maria Pia Del Rosso, Silvia L. Ullo, and Paolo Gamba. A Speckle Filter for Sentinel-1 SAR Ground Range Detected Data Based on Residual Convolutional Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:5086–5101, 2022.

[121] Rebecca A. Segal, Randall K. Scharien, Frank Duerden, and Chui-Ling Tam. The Best of Both Worlds: Connecting Remote Sensing and Arctic Communities for Safe Sea Ice Travel. *ARCTIC*, 73(4):461–484, January 2021.

[122] Mohammed Shokr and Nirmal K. Sinha. *Sea Ice Physics and Remote Sensing*. Number 209 in Geophysical Monograph. American Geophysical Union + John Wiley & Sons, Inc., 2015.

[123] Markku Similae. SAR segmentation by a two-scale contextual classifier. pages 434–443, Rome, Italy, December 1994.

[124] Natasha Simonee, Jayko Alooloo, Natalie Ann Carter, Gita Ljubicic, and Jackie Dawson. Sila qanuippa? (how's the weather?): Integrating Inuit Qaujimajatuqangit and environmental forecasting products to support travel safety around Pond Inlet, Nunavut in a changing climate. *Weather, Climate, and Society*, August 2021.

[125] S. Sinha, C. Jeganathan, L. K. Sharma, and M. S. Nathawat. A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology*, 12(5):1779–1792, May 2015.

[126] David Small. Flattening Gamma: Radiometric Terrain Correction for SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(8):3081–3093, August 2011.

[127] Pierre Soille. Opening and Closing. In *Morphological Image Analysis*, pages 105–137. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[128] Wei Song, Wen Gao, Qi He, Antonio Liotta, and Weiqi Guo. SI-STSAR-7: A Large SAR Images Dataset with Spatial and Temporal Information for Classification of Winter Sea Ice in Hudson Bay. *Remote Sensing*, 14(1):168, December 2021.

[129] Wei Song, Minghui Li, Wen Gao, Dongmei Huang, Zhenling Ma, Antonio Liotta, and Cristian Perra. Automatic Sea-Ice Classification of SAR Images Based on Spatial

and Temporal Features Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12):9887–9901, December 2021.

[130] Timbo Stillinger, Dar A. Roberts, Natalie M. Collar, and Jeff Dozier. Cloud Masking for Landsat 8 and MODIS Terra Over Snow-Covered Terrain: Error Analysis and Spectral Similarity Between Snow and Cloud. *Water Resources Research*, 55(7):6169–6184, July 2019.

[131] Ian Stirling and Holly Cleator. *Polynyas in the Canadian Arctic*. Number 45 in Occasional paper / Canadian Wildlife Service. Ottawa, 1981.

[132] Takeshi Tamura and Kay I. Ohshima. Mapping of sea ice production in the Arctic coastal polynyas. *Journal of Geophysical Research: Oceans*, 116(C7):2010JC006586, July 2011.

[133] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019. Publisher: arXiv Version Number: 5.

[134] Frank James Tester and Peter Irniq. Inuit Qaujimajatuqangit: Social History, Politics and the Practice of Resistance. *Arctic*, 61:48–61, 2008. Publisher: Arctic Institute of North America.

[135] Kang Tong, Yiquan Wu, and Fei Zhou. Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, 97:103910, May 2020.

[136] Yi Wang, Syed Muhammad Arsalan Bashir, Mahrukh Khan, Qudrat Ullah, Rui Wang, Yilin Song, Zhe Guo, and Yilong Niu. Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Systems with Applications*, 197:116793, July 2022.

[137] Yi-Ran Wang and Xiao-Ming Li. Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning. *Earth System Science Data*, 13(6):2723–2742, June 2021.

[138] Pål Wessel and Walter H. F. Smith. A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research: Solid Earth*, 101(B4):8741–8743, April 1996.

[139] W.J. Williams, E.C. Carmack, and R.G. Ingram. Chapter 2 Physical Oceanography of Polynyas. In *Elsevier Oceanography Series*, volume 74, pages 55–85. Elsevier, 2007.

[140] Katherine Wilson, Andrew Arreak, Jamesie Itulu, Sikumiut Community Management Committee, Gita J. Ljubicic, and Trevor Bell. "When We're on the Ice, All We Have is Our Inuit Qaujimajatuqangit": Mobilizing Inuit Knowledge as a Sea Ice Safety Adaptation Strategy in Mittimatalik, Nunavut. *ARCTIC*, 74(4):525–549, January 2022.

[141] Katherine Wilson, Andrew Arreak, Sikumiut Committee, Trevor Bell, and Gita Ljubicic. The Mittimatalik Siku Asijjipallianinga (Sea Ice Climate Atlas): How Inuit Knowledge, Earth Observations, and Sea Ice Charts Can Fill IPCC Climate Knowledge Gaps. *Frontiers in Climate*, 3:715105, October 2021.

[142] Wilson, Katherine. The Sikumiut model: a cross-cultural decolonizing research approach for see ice travel safety in Mittimatalik, Nunavut. 2022. Publisher: Memorial University of Newfoundland.

[143] Iain H. Woodhouse. *Introduction to microwave remote sensing*. Taylor&Francis, Boca Raton, 2006.

[144] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? 2014. Publisher: arXiv Version Number: 1.

[145] Natalia Zakhvatkina, Anton Korosov, Stefan Muckenhuber, Stein Sandven, and Mohamed Babiker. Operational algorithm for ice–water classification on dual-polarized RADARSAT-2 images. *The Cryosphere*, 11(1):33–46, January 2017.

[146] Natalia Yu. Zakhvatkina, Vitaly Yu. Alexandrov, Ola M. Johannessen, Stein Sandven, and Ivan Ye. Frolov. Classification of Sea Ice Types in ENVISAT Synthetic Aperture Radar Images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2587–2600, May 2013.

[147] Haoran Zhu, Chang Xu, Wen Yang, Ruixiang Zhang, Yan Zhang, and Gui-Song Xia. Robust Tiny Object Detection in Aerial Images amidst Label Noise. 2024. Publisher: arXiv Version Number: 1.

[148] Jelle Zitman. Detecting plume-driven polynyas from dual-pol SAR imagery. Master's thesis, TU Delft, August 2022.

[149] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 111(3):257–276, March 2023.

APPENDICES

Appendix A

Model Hyperparameters

Table A.1: Faster-RCNN hyperparameters used for the adapted polynya detection models in this thesis. Modifications to FRCNN defaults are discussed in section 4.1.2.
[†]This hyperparameter is modified in the experiments of section 4.3.1

Hyperparameter	**FRCNN [113] default**	**Adapted**
Region Proposal Network		
Conv. kernel size	3x3	3x3
Conv. depth	1	1
Anchor sizes	{128, 256, 512}	{8, 16, 32}
Anchor aspect ratios	{1:1, 1:2, 2:1}	{1:1, 1:2, 2:1}
Positive anchor IoU threshold	0.7	0.5
Negative anchor IoU threshold	0.3	0.1
NMS threshold	0.7	0.7
Post-NMS Top-N (train)	2000	100
Post-NMS Top-N (inference)	1000	100
Obj. loss samples per image	256	64[†]
Obj. loss positive fraction	0.5	0.5
RoI Pooling		
Pooling window	7x7	3x3
RoIAlign sampling ratio	2	2
RoI Head Network		
N. FC layers	2	2
N. neurons / layer	1024	128
Positive box IoU threshold	0.5	0.5
Negative box IoU threshold	0.5	0.3
NMS threshold	0.5	0.5
Cls. loss samples per image	512	32
Cls. loss positive fraction	0.25	0.5
Top-N per image (inference)	100	100

Appendix B

Evaluation Scenes

Table B.1: All Sentinel-1 EW scenes held out during training and used for polynya detection model evaluation in chapters 4 and 5.

Sanikiluaq Dataset					*Nain Dataset*				
		N. Targets with $\delta^{\mu}_{\mathbf{HH}} \leq$					**N. Targets with $\delta^{\mu}_{\mathbf{HH}} \leq$**		
Date	**IA [°]**	**-1dB**	**-3dB**	**-5dB**	**Date**	**IA [°]**	**-1dB**	**-3dB**	**-5dB**
01-25-2017	19.2 - 29.2	1	1	1	03-19-2020	34.1 - 46.8	2	0	0
02-13-2017	41.4 - 46.8	2	2	2	03-30-2020	20.7 - 38.1	2	1	1
03-21-2017	41.4 - 46.7	0	0	0	03-31-2020	34.1 - 46.8	9	6	5
04-12-2017	24.3 - 36.0	16	13	7	04-11-2020	20.7 - 38.1	3	2	2
02-11-2018	19.2 - 27.5	9	8	5	04-23-2020	20.7 - 38.1	5	3	1
02-20-2018	41.4 - 46.8	5	3	1	04-24-2020	34.1 - 46.8	14	11	9
03-08-2018	24.3 - 36.0	26	10	8	05-05-2020	20.7 - 38.1	8	4	2
04-02-2018	33.6 - 43.1	24	18	8	05-06-2020	34.1 - 46.8	10	10	9
05-07-2018	24.3 - 36.0	13	10	0	03-25-2021	20.7 - 38.1	7	5	3
02-01-2019	24.3 - 36.0	0	0	0	03-26-2021	34.1 - 46.8	10	4	2
02-15-2019	41.4 - 46.7	6	5	2	04-06-2021	20.7 - 38.1	1	0	0
03-11-2019	41.4 - 46.7	5	4	2	04-07-2021	34.1 - 46.8	8	5	3
04-02-2019	24.3 - 36.0	3	0	0	04-18-2021	20.7 - 38.1	14	6	4
05-01-2019	19.2 - 27.5	13	8	2	04-19-2021	34.1 - 46.8	19	10	6
01-29-2020	41.4 - 46.7	4	3	3	04-30-2021	20.7 - 38.1	7	4	3
03-09-2020	24.3 - 36.1	16	10	8	05-01-2021	34.1 - 46.8	9	6	5
03-20-2020	19.2 - 27.5	8	5	4	03-21-2022	34.1 - 46.8	4	2	1
04-19-2020	19.2 - 27.7	12	8	6	04-02-2022	34.1 - 46.8	5	4	3
05-04-2020	41.3 - 46.7	12	11	8	*Total*		*137*	*83*	*59*
01-27-2021	24.3 - 36.1	18	10	6					
02-15-2021	33.5 - 43.1	23	10	15					
02-27-2021	33.5 - 43.1	10	10	6					
03-16-2021	24.3 - 36.1	3	0	0					
04-04-2021	33.5 - 43.1	18	14	11					
04-29-2021	41.3 - 46.7	7	6	5					
01-29-2022	33.5 - 43.1	14	11	7					
03-16-2022	19.2 - 27.7	3	3	1					
Total		*271*	*183*	*108*					

Glossary

hyperparameters Design elements of a model usually chosen by hand and fixed during the optimization (model training) process 21, 40, 43, 49

natural-scene Refers to optical-wavelength images, usually RGB, taken from a human perspective 18, 38