Arctic Sea Ice Property Retrieval from Synthetic Aperture Radar with Deep Learning Models

A thesis presented for the degree of

Doktor rer.nat.

by

Karl Kortum

Department of Physics

University of Bremen

Reviewer 1: Dr. Gunnar Spreen

Reviewer 2: Prof. Dr. Christian Haas

Submitted: 13.10.2023

Defended: 08.02.2024

Arctic Sea Ice Property Retrieval from Synthetic Aperture Radar with Deep Learning Models

Karl Kortum

Abstract

Current climate models are not capturing the intricate feedback mechanisms driving the accelerated warming of the Arctic. A great challenge the research is facing is the sparsity of high resolution observations. Satellite-borne synthetic aperture radar (SAR) instruments have the capability of monitoring Earth's sea ice masses at high resolution, unhampered by cloud coverage or the Arctic night. The measurements made by such instruments are made at scales of 10's of metres whilst still covering all of the Arctic in a matter of days. The combination of high resolution and frequent revisit time make the instruments a prime candidate for sea ice mapping. However, interpreting the radar signal to retrieve relevant sea ice information, such as roughness or ice age, is difficult. This is because of the complex interactions of the diverse and heterogeneous ice with the electromagnetic radar signal. Conventional neural network based algorithms have proven adept at leveraging contextual image data to make accurate predictions of surface ice properties comparable to those made by human experts. They are, however, dependent on large amounts of high-quality ground truth that is rare for the remote and dynamic sea ice. Thus, the full potential of high resolution classification that captures the intricacies of the SAR data, and robustly interprets sea ice properties from it, is yet to be unlocked. With the advent of the MOSAiC mission, large timeseries of SAR data and near-coincident ground measurements were made for an entire ice season for the first time. This thesis uses the unique opportunity provided by these data to analyse the behaviour of deep learning models. Seven months of data from the campaign is classified and analysed, using newly developed techniques to enable robust predictions across the timeseries. Core features are identified to facilitate robust and high-resolution classification. The final challenge of ground truth sparsity is then overcome using innovative network configurations that enable the training of $> 99.99\%$ of the model parameters without any ground truth data. The architecture-agnostic techniques developed as part of this thesis open up the space of sea ice property retrieval to big data technologies, where the model is able to improve with nothing but the abundantly available SAR data. These techniques enable the extrapolation of critically sparse reference data to a large space of sea ice conditions and enable high resolution and robust mapping of the Earth's region most affected by human-made climate change.

Contents

Chapter 1

Introduction

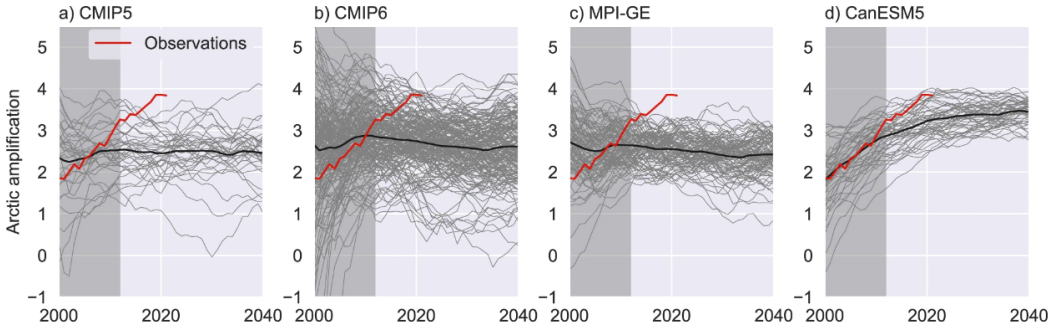
The Earth's surface is primarily covered by vast dark oceans. It is here where life as we know it originated, evolving creatures of ever-increasing complexity. And still, to this day, all life depends on it. Our relationship with the H_2O molecule is thus intrinsic to our nature. As we shift our gaze from the simple molecule to microorganisms living in the ocean, to large organisms such as ourselves dwelling on land, and finally all the way to the planet itself, the relationship with water remains important across all scales.

Since the industrial revolution around 200 years ago, humankind has been significantly impacting the atmosphere's composition with the continuous emission of greenhouse gases, dominantly carbon dioxide (CO_2) and methane (CH_4). For a current state of the climate and up-to-date evaluations of it's future, one may refer to the *IPCC, 2023: Climate Change 2023: Synthesis Report*. What would have been science fiction to those people who started to revolutionise production with the help of steam-powered machinery all those years ago has now become a reality. Humankind is rapidly changing the climate of the planet. Climate change has at its core a straightforward recipe: the Earth's energy imbalance. Around 150 million kilometres away, our sun fuses Hydrogen to Helium - acting like a furnace at our solar system's centre. A tiny little part of the energy radiated out into space arrives at Earth, yet it is still mighty enough to drive our planet's climate. Until now, this system has been reasonably stable - always returning to levels facilitating life on the Earth's surface. Recent research by Song, Kemp, et al., 2021 has shown, however, that both the magnitude and the rate of temperature change are significantly positively correlated with the rate of extinction of marine species. Whilst life has endured on earth, other planets at less favourable distance to the sun have not been so lucky. According to Way et al., 2016, our celestial neighbour Venus is thought to once have been covered by oceans too and it's atmosphere is now heavily dominated by dense carbon dioxide. We do not know what caused these transitions of our celestial neighbour, but we now find it as a barren wasteland - inhospitable to any form of life on the surface.

The amount of the sun's energy retained in the Earth's system is driven by radiative transfer. As shortwave radiation (.2μm - 3μm wavelength) hits the Earth's surface, they interact with the particles in two different ways. Whilst some surfaces directly reflect these electromagnetic waves, most waves are partially absorbed and re-emitted at higher wavelengths (4μm - 70μm). If these longer waves could travel back into space freely, the energy retention would be minimal. However, it still has to pass through Earth's atmosphere. Despite being very thin in comparison to the Earth's diameter, this conglomeration of gases trapped in the Earth's gravitational field makes a big difference: Some of the gases in the atmosphere can reabsorb or reflect these lower-frequency light waves that are coming from the Earth's surface. Most impactful are the suspended liquid water droplets of cloud formations, that absorb practically all light upwards of 25 μm. The resulting imbalance in the Earth's energy budget has slightly more energy being absorbed than emitted (quantified, for example, in Trenberth et al., 2009). These gases thus significantly affect the amount of energy retained in the Earth's systems and have fittingly been called greenhouse gases.

The 43-year Arctic amplification ratio in observations and climate models.

From: The Arctic has warmed nearly four times faster than the globe since 1979

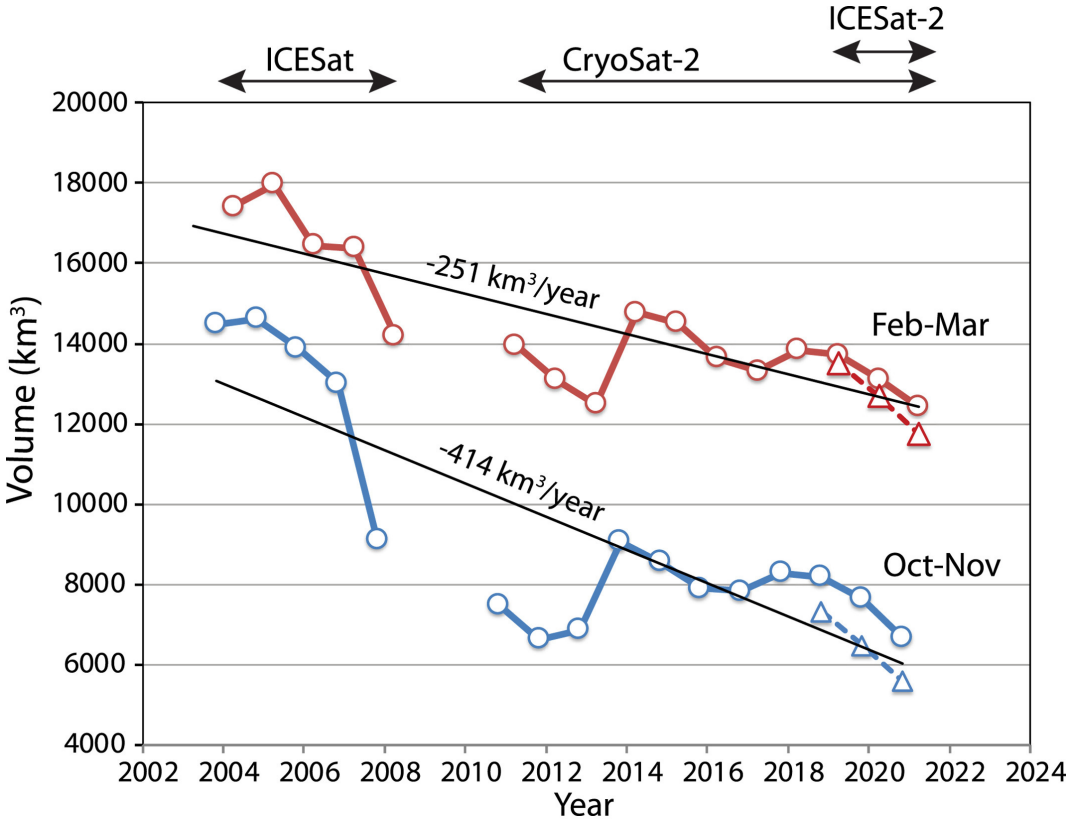


The 43-year Arctic amplification (AA) ratio derived from (**a**) CMIP5, (**b**) CMIP6, (**c**) MPI-GE and (**d**) CanESM5 realizations (thin grey lines) for all 43-year periods ending in 2000–2040. The x-axis represents the ending year of the 43-year AA ratios. Thick black lines represents the ensemble mean AA, calculated as a mean of ratios, not ratio of means. Observations (red lines) extend to 2021. 43-year AA ratios starting after 1970 and ending by 2040 are considered in the probability calculations (Section "Likelihood of observed Arctic amplification 1979–2021 in climate model simulations") and shown with light background. The Arctic is defined as the area north of 66.5°N.

Figure 1.1: Figure 3 from Rantanen et al., 2022, showing Arctic amplification in the context of four different climate model ensembles. Published in Nature Communications Earth & Environment under a Creative Commons CC BY 4.0 license. ©Nature

The surface itself also plays as big of a role. Some materials, like the surface dominating liquid water, retain most of the sunray's energy. Others, like snow, reflect most of the sun's rays that have made their way through the atmosphere. As the ice and snow cover on the ocean lessens, the energy uptake is thus increased. This is commonly referred to as the albedo effect. With that in mind, it comes as no surprise that the Arctic is more affected by climate warming than all other regions on Earth. This is commonly described as Arctic amplification and has led to the Arctic warming four times as fast as the rest of the planet in recent years, as shown in Rantanen et al., 2022. The same work has also made clear that the strength of the amplification is not well described by current climate models (see figure 1.1), highlighting the need for more detailed observations of the Arctic environment.

As the Earth gets warmer, the sea ice cover decreases, and more ocean surface is exposed to solar radiation, leading to increased warming. This is the general feedback loop for the amplification. However, a variety of additional complex effects in play are not as well understood, as is made evident by the discrepancy of climate models and the historical record. Additionally, with perennial ice becoming increasingly rare in the Arctic, it will be dominated instead by thinner and younger ice. Only recently has the historical reduction of ice volume been parametrised: Recent work by Kacimi and Kwok, 2022 uses time-series of satellite altimeters to resolve not just the surface coverage but also the thickness of the ice. Results of that study are shown in figure 1.2. The increasingly prevalent younger and thinner ice develops differently both thermodynamically and under mechanical stress. These interactions with winds, waves and ocean currents break up the ice pack and open the ocean to interactions with the atmosphere (such as the emission of gases and aerosols, which can act as cloud-nucleating particles). To understand and extrapolate the behaviour



Arctic sea ice volume calculated from ICESat (2003–2008), CryoSat-2 (2010–2021) and ICESat-2 (2018–2021) ice thickness fields, for February–March (in red) and October–November (in blue). The bimonthly averages (October–November and February–March) are aligned with the campaign mode of ICESat. For the 3 years of overlapping CS-2 and IS-2, we also show ice volume using ice thickness estimates derived from satellite snow depths from differencing IS-2 and CS-2 freeboards (depicted by the triangles). Linear trends are calculated using estimates from longer time series of ICESat and CryoSat-2.

Figure 1.2: Figure 3 from the Kacimi and Kwok, 2022, showing the loss of arctic sea ice volume (from a spring and an autumnal perspective) as observed through satellite altimeters. Published in Geophysical Research Letters. ©John Wiley and Sons.

of the future Arctic, it is crucial to know how the responses to mechanical stress change with the ice age and thickness. The deformation that is left behind by these events has another effect later in its life cycle: Sea ice deformation in winter impacts the ice's coverage with meltwater in summer, as shown by Webster et al., 2015. The sea ice pack becoming increasingly made up of less deformed first-year ice could increase melt pond coverage and thus further amplify the Arctic warming. These mechanisms are not yet fully understood (and thus not included in climate models). Current climate models not capturing observed Arctic Amplification strongly suggests that these interactions are highly significant.

It is evident that the properties of the sea ice cover have a significant and diverse impact on the climatology of the Arctic system. As natural sciences are inherently empirical and the Arctic is a remote and hostile environment to the human organism, it is unsurprising

that the wealth of data to further our understanding is limited. The most comprehensive opportunity to monitor the Arctic sea ice cover at full scale thus comes from remote sensing using space-borne instruments. Specifically, observations at resolutions of 10s to 100s of metres, which reliably capture the local distribution of various ice types and open water, are needed. However, the lack of in-situ data is still the central complication in this field - hampering detailed observations and time series. Most importantly for this work, it also restricts available ground truth for models and retrieval methods.

These retrieval methods from spaceborne satellites need to perform well at high resolution and across the Arctic seasons to help us unravel the mechanisms of the sea ice and sharpen the predictions of climate models. Neural network based models are the primary data-driven resource to do such retrieval from synthetic aperture radar (SAR) satellites, as they excel at incorporating the wealth of image information to achieve accurate predictions. This method's complications revolve around a central conflict between using data-driven models and the sparsity of high resolution ground truth data available in the Arctic. In this thesis, deep-learning methodologies are analysed and advanced to define and overcome this challenge of sea ice property retrieval using synthetic aperture radar satellites. The aim is to move toward systems able to make robust observations near the resolution of the sensor to be able to advance the understanding of the sea ice system in a changing climate. To do so, established and newly developed neural network based methods are analysed on long timeseries of data acquired during the MOSAiC (Multidisciplinary Drifting Observatory for the Study of Arctic Climate) expedition.

The overarching question is, of course, if it is even possible to build data-driven deep learning models that perform accurately and robustly at high resolution, despite very little high resolution ground truth being available. To answer this, a series of studies on diverse data sets are presented in chapters 3 - 6 that answer aspects of that question empirically and quantitatively.

To test the performance of classifiers, research is currently limited to manual labelling (such as ice charts), and only sparse work has been done with validation from other measurements. A necessary step to establish the applicability of deep learning models is to test the robustness of these models over longer time-series and develop heuristics to enable robust predictions, even in situations where one has no data to test it systematically. This is tackled in chapter 3, using a long high-resolution TerraSAR-X X-Band time series acquired during MOSAiC.

The absence of comprehensive ground truth data sets poses an important and hitherto unanswered question. How do state of the art models perform on full resolution measured label data, and what is the impact and consequence of using coarser manual labels, which is common practice in the field currently? This question is answered in chapter 4 by drift correcting and co-locating twenty airborne laser scanning topography surveys with SAR data to construct a set of ground truth labels. It also establishes some differences between two different approaches to classification and the necessity for a model to be able make use

of intra-label dependencies.

To solve the ground truth sparsity for pan-Arctic classification, a core idea could be to use scenes without any labels. Is it possible to somehow leverage the vast number of unlabelled SAR scenes to inform retrieval models and improve their performance in the diverse range of ice conditions? This question is answered in chapter 5 by constructing an innovative physics-constrained adversarial network approach that uses the task of incidence angle dependence estimation as a proxy to ice type classification on Sentinel-1 C-Band data.

In the final chapter the results from the previous studies are pooled and incorporated into a classifier that can robustly extrapolate sparse amounts of altimeter derived ice properties to a diverse set of ice conditions on a larger scale, near the resolution and fidelity of the sensor. Thereby answering the question if it is possible to train advanced deep learning architectures for high fidelity sea ice property retrieval from SAR despite high resolution ground truth in the Arctic being critically sparse.

Chapter 2

Theory

This chapter introduces the nature of sea ice as a medium, as well as the core instruments and methods that will be used throughout the thesis to observe sea ice and its properties. It includes an overview of sea ice as a medium, an introduction to Synthetic Aperture Radar (SAR) technology and neural networks. This chapter will ease the reader into some fundamental theory on which the following work in the thesis is based. Also included are some observations and the qualitative interpretation of those, together with simple modelling schemes to back-up them up. Finally, the core conundrums of using deep learning models for sea ice retrieval are formulated, and a brief overview of existing work on the topic is given.

2.1 Sea Ice

To understand the challenges of sea ice property retrieval, one requires some knowledge about sea ice itself. Let us start this foray into the nature of sea ice with the humble yet powerful water molecule. Water is made up of two hydrogen atoms bonded with one oxygen. The most energy-effective way to bond these atoms results in an angle of approximately 104.5 degrees between the covalent bonds. Electrons dominate the side opposite the hydrogen cores. As the electrons in the system have a higher probability of being found on the far side, away from the hydrogen, the resulting molecule has a dipole moment. This intrinsic orientation of the molecule is of paramount importance for the properties of water in various states and interactions.

As water approaches its freezing point, it becomes increasingly likely to form hydrogen bonds without them being broken up by thermal energy states. A hydrogen bond is formed by orienting two molecules in such a manner that the hydrogen of one molecule is closest to the oxygen of another molecule. If multiple bonds are formed to a crystal, one oxygen, its two covalently bonded hydrogens and the two hydrogen bonds with hydrogen cores from other molecules form approximately a tetrahedral structure. The oxygen is in the centre, and the hydrogen atoms are at the vertices. The hydrogen will still be closer to one of the oxygen atoms rather than in the centre of the two, favouring the side with the covalent bond. Although this is the most energy-efficient way to organise the molecules, it is not the most space-efficient, which leads to the water density anomaly: At normal atmospheric pressure, water is densest in a liquid state at approximately 4 degrees Celsius. The most energy-efficient way to organise the hydrogen-bonded lattice with tetrahedral components results in hexagonal rings. This symmetry is preserved even at macroscopic scales and can be observed with the human eye in, for example, snowflakes. At 0 degrees Celsius (or 273.15K), the probability of forming these bonds becomes higher than breaking them through Brownian motion and the crystal is formed.

As they generally do not fit well into the specific natural structure, impurities in the crystal lattice of water are very energy inefficient. Because they cannot be incorporated efficiently, the ice formed from lake water is comparatively pure and much cleaner than the

water itself. Any impurities that do make their way into the ice are, hence, generally trapped in pockets or at interfaces in the polycrystalline structure.

If salt is added to the water, the behaviour of the solution changes. Seawater typically has salinities of around 35 parts per thousand (ppt). In the liquid state, the salt is hydrolysed into cations and anions. The water molecule's dipole leads to clusters around these ions, where the dipoles are oriented in the ions' electric fields. This increases its density in the liquid phase and lowers both boiling and freezing temperatures. As the temperature is lowered, the hydrogen bonds needed for solidification additionally compete with the electromagnetic interactions with the ions. Like impurities in lake water, the salts are largely emitted from the ice crystal during freezing and form in pockets or are expelled at the interfaces. The brine is liquid at much lower temperatures because the salt concentration in pockets inside the crystal is so high. Within the ice crystal, this salt will destroy the ice structure over time and form channels. As the ice gets older, it will desalinate. The dominant mechanism for this was identified as gravity drainage in Notz and Worster, 2009, where the colder, denser brine in the ice releases salt via exchange with the warmer and less saline (therefore less dense) seawater. So, whilst young ice can have salinities around 12-15 ppt, consolidated first-year ice typically ranges around 4-6 ppt, and multiyear ice is even lower than that at less than 2 ppt. An overview connecting the sea ice age or thickness with its salinity can be found in Kovacs et al., 1996.

Because salinity increases the density of brine, but temperature decreases the density, a temperature gradient in the ice results in an inverse salinity gradient in the brine to maintain equal density (phase equilibrium). As ice cools, the brine fraction decreases and, in turn, its salinity rises. In thermal equilibrium at -5 degrees Celsius and with a seawater salinity of 34 parts per mil, Assur, 1960 finds solid ice fractions of ca 65 % under laboratory conditions in a closed volume. As discussed, in nature, desalination via exchange with the ocean occurs naturally during freeze-up, and the brine volume is thus significantly lower than that of ice grown in a closed system. Still, this measurement shows that sea ice cannot be thought of as a solid but needs to be considered a mushy layer of polycrystalline ice and brine.

Initial sea ice freezing starts off with the formation of tiny crystals that eventually form a slush-like layer on the top of the ocean. These crystals are known as frazil ice and are typically on length scales of millimetres. The frazil ice either consolidates into a relatively homogenous ice sheet called a nila - appearing first translucent, then grey - or is compressed under the action of ocean turbulence forming pancake ice, which are rounded plates that typically have raised edges from mutual collisions. When the ocean-atmosphere interface is entirely covered by ice, growth can continue only in the vertical direction. The speed of growth is related to the ice's thermal conductivity, as the ocean's latent heat is dissipated into the atmosphere for the ice to grow. Younger, more saline ice is not only thinner but also more conducive to heat - leading to accelerated ice growth of younger ice sheets in contrast to dry and less saline multi- and second-year ice.

New ice growth on the bottom is often found to have a columnar nature. A theory for

the origin of this formation is found in the desalination process of ice upon freeze-up: The rejection of salt and latent heat results in a salty layer at the ocean-atmosphere interface during initial freeze-up and at the ice-ocean interface after the ice has consolidated. To maintain thermal equilibrium there, a heat flux away from that layer is a direct consequence. However, the heat flux happens faster than the dissipation of the salt and this layer is constantly supercooled. Thus, any perturbation or protrusion into this layer can quickly lead to more ice growth, catalysing vertical structures. For columnar ice growth, this requires relative calm in the ocean, which does not disturb the supercooled layer mechanically. Due to increased ocean dynamics in the Antarctic, columnar ice is much sparser here (e.g. Eicken, 2003). A comprehensive monograph for the formation and structure of sea ice by W.F. Weeks and S.F. Ackley, U.S. Army Cold Regions Research and Engineering Laboratory, 1982 may serve as a detailed reference.

Snow is the other major contributor to heat flux - and thus to ice growth. Snow is much less heat conductive than the underlying ice and acts as a blanket - insulating the ice pack from the atmosphere. Warming the underlying ice also affects its dielectric properties and, therefore, its appearance on microwave remote sensing instruments, as was explored, for example, by Kim et al., 1984.

Having discussed the thermodynamic growth and some basic relationships of external conditions with the ice formation, dynamic ice events also need to be taken into account. These affect the evolution of the properties of sea ice across all scales. Keeping in mind that sea ice is not a solid crystal but rather a mushy layer of crystals and brine (as well as some gas) inclusions - ice is more plastic than one might naively assume. For example, this means it can compress under stress without breaking. A complete understanding/parametrisation of how ice deforms has not yet been achieved. There are, however, observable dynamic features in the Arctic and Antarctic environments. These features are born of different kinds of dynamical events: mainly divergence, convergence or shear. As ice converges, it is placed under increasing stress, which can eventually lead to the breaking of floes. If the stress is high, it may be pushed together and break up the ice, creating ridges along the fault line of the break. These pressure ridges act as sails on the ice, increasing the coupling to the atmosphere's winds and the amount of energy that can flow in such a manner from one system to the other. If the ice diverges after a breaking event, it will open up a lead of open water in the ice pack. Leads are not only hotspots of interaction between ocean and atmosphere but also artefacts of the ice dynamics and, therefore, one of the most investigated sea ice phenomena.

2.2 Remote Sensing of Sea Ice

As was mentioned in the introduction of the thesis, remote sensing is an important tool to monitor the polar regions. Intrinsic to the remote sensing aspect of sea ice, is its response to an external electromagnetic field. For a bulk material, this is described by the strength

of the induced field in the material (permittivity) and the amount of energy transferred to free ions (conductivity). Both of these variables are combined into a single complex variable $\epsilon_r = \epsilon + i\sigma$, commonly referred to as the relative permittivity or the dielectric constant. It appears in the exponent of the wave solution to the (one-dimensional) free field equations. For an undampened field ($\sigma = 0$) propagating along x,

$$\frac{\partial^2}{\partial x^2} E = \mu\epsilon \frac{\partial^2}{\partial t^2} E, \tag{2.1}$$

$$\text{with} \qquad E \propto e^{i\omega t} E_x, \tag{2.2}$$

$$\Rightarrow \qquad E_x = e^{-i\omega\sqrt{\mu\epsilon}x}. \tag{2.3}$$

If we consider a dampening effect due to the conductivity σ of the material, the wave equation in turn becomes

$$\frac{\partial^2}{\partial x^2} E = \mu\epsilon_r \frac{\partial^2}{\partial t^2} E. \tag{2.4}$$

Again, choosing a propagation only along direction x and looking only at the spatial dependence the solution becomes a little more complex, as

$$E_x = e^{-i\alpha x} e^{-\sigma\beta x}, \tag{2.5}$$

$$\text{with} \qquad \beta = \frac{\omega\mu}{2\alpha}, \tag{2.6}$$

$$\text{and} \qquad \alpha = \omega\sqrt{\mu\epsilon}\left[\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}}\right]^{\frac{1}{2}}. \tag{2.7}$$

Here the imaginary part of the complex permittivity σ (conductivity) contributes as multiplication with a negative exponential function, thus parameterising the drop of the electric field strength in the medium. A property that is often referenced is the penetration depth

$$\delta = \sigma\beta = \frac{1}{2}\sqrt{\frac{\mu}{\epsilon}}\left[\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}}\right]^{-\frac{1}{2}}. \tag{2.8}$$

The real part ϵ (permittivity) governs the strength of the response, which will be shown below. Electromagnetic properties are largely influenced by the amount of brine present in the ice and vary across frequencies of the external electric field. As discussed, the brine volume changes over time, and thus so does the response on electromagnetic remote sensing instruments. Whilst the conductivity discussed above describes the absorption of an electromagnetic wave in a dielectric medium, we still need to look at the strength of the reflection (and transmission). A macroscopic description is given by the Fresnel equations. The symmetry-breaking element in the scattering mechanism is the orientation of the surface. This results in different scattering for different polarisations and incidence angles. The scattering mechanism/matrix can be derived from the wave equations and upholding continuity at the interface boundary. This scattering cannot lead to changes in polarization. Polarisation flips require geometries with multiple bounces.

Assuming a horizontally polarised incident wave E_i^H, B_i, with incidence angle θ_i, reflected wave E_r^H, B_r and transmitted wave E_t^H, B_t at transmitted angle θ_t and requiring continuity

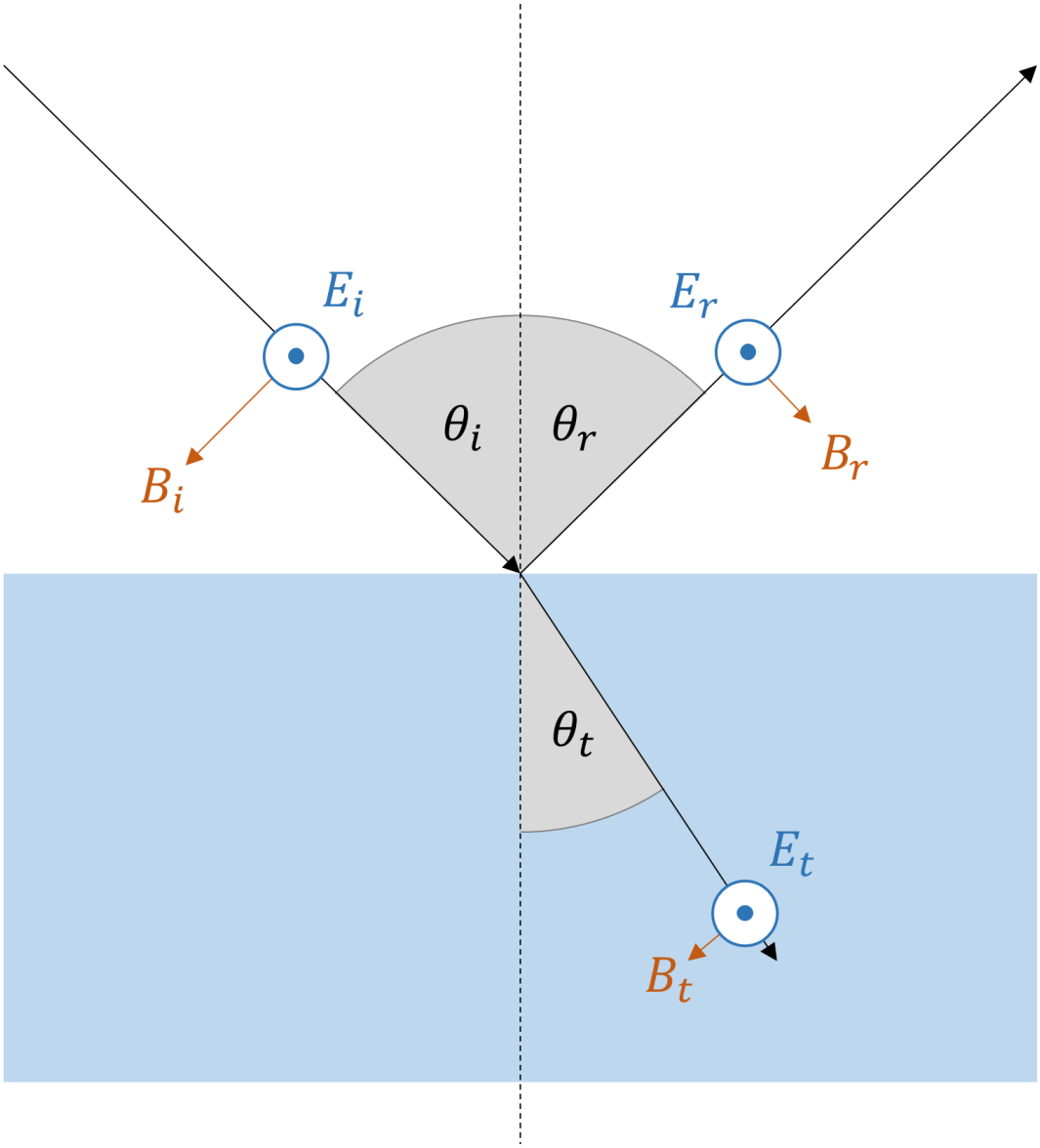


Figure 2.1: Fresnel scattering at a single interface between two homogeneous substances. The wave travels from a substance of permittivity ϵ_i and is scattered at an interface with permittivity ϵ_t.

of the electric field at the interface gives two equations for the amplitudes:

$$E_t^H \stackrel{!}{=} E_i^H + E_r^H \tag{2.9}$$

$$B_t \cos\theta_t \stackrel{!}{=} B_i \cos\theta_i - B_r \cos\theta_r. \tag{2.10}$$

For an electromagnetic wave,

$$B = E\sqrt{\epsilon\mu}. \tag{2.11}$$

Constraining ourselves to non-magnetic media, the combination of the above equations yields a relationship for the ratio of reflected to the incident wave, as

$$\frac{E_r^H}{E_i^H} = \frac{\sqrt{\epsilon_i}\cos(\theta_i) - \sqrt{\epsilon_t}\cos(\theta_t)}{\sqrt{\epsilon_i}\cos(\theta_i) + \sqrt{\epsilon_t}\cos(\theta_t)}. \tag{2.12}$$

One thing that is immediately obvious is that the higher permittivity of the transmission medium leads to lower amplitude of the reflected wave, as one would expect. One thing we have not yet exploited is the relationship of incident and transmission angles, governed by Snell's law. Assuming $\epsilon_i \approx 1$, as for the atmosphere, we obtain the formulation

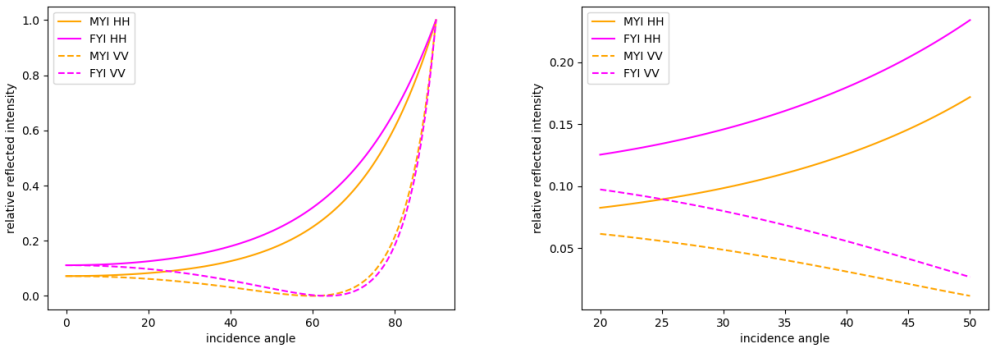
$$\frac{E_r^H}{E_i^H} = \frac{\cos(\theta_i) - \sqrt{\epsilon_t - \sin^2(\theta_i)}}{\cos(\theta_i) + \sqrt{\epsilon_t - \sin^2(\theta_i)}}. \tag{2.13}$$

Equivalently for vertical polarisation, one obtains

$$\frac{E_r^V}{E_i^V} = \frac{-\sqrt{\epsilon_t}\cos(\theta_i) + \sqrt{1 - \sin^2(\theta_i)/\epsilon_t}}{\sqrt{\epsilon_t}\cos(\theta_i) + \sqrt{1 - \sin^2(\theta_i)/\epsilon_t}}. \tag{2.14}$$

For an example of the Fresnel scattering at substances with permittivities similar to first-year and multiyear ice over various incidence angles, see figure 2.2a.

One might have noticed that the Fresnel description only covers specular reflection. If we operate an active satellite at a slanted incidence angle, very little or no incident wave would be returned to the satellite via spectral reflection on a relatively level surface. One crucial aspect that we have been missing in our description is that of surface deformation, which of course also influences the local incidence angle of the wave. To simplify interactions across multiple scales, one can subdivide the influence of deformation into three separate scattering regimes. In the first case, the scatterer (or correlation length of surface deformations) can be significantly larger than the wavelength and we can treat the scattering as largely specular reflection. Commonly this is termed the optical regime. This would then be quite well described by the Fresnel equations introduced above. Note that energy returning to an active instrument viewing from a non-nadir angle would need either multiple bounces or an optimally slanted surface to see a reflected signal. Thus in this regime, the return is largely determined by surface geometry. In a second case, the scatterer is significantly smaller than the wavelength, in which case we enter Raleigh scattering regime. Note, however, that there are only a few scatterers in sea ice that fit this criterion and the scattering strength is quite weak compared to the other contributions. Thus this type of scattering can be considered



(a) Full IA range

(b) SAR IA range

Figure 2.2: Ratio of reflected compared to initial wave energy from Fresnel scattering at a single interface of a substance of permittivity $\epsilon_t = \{3, 4\}$ for 'first-year ice' (FYI) and 'multiyear ice' (MYI) respectively. The permittivity of the initial medium is set to 1. The right figure shows a cutout to the incidence angle range, as typical for a SAR satellite.

negligible in our case. It is a reason why higher frequency microwave sensors can no longer penetrate cloud cover, as suspended water particles facilitate significant Raleigh scattering. Finally, if the scatterers are of similar size to the wavelength, this is described by Mie Theory and is the most complicated of the three cases. We will not discuss it in detail here, but it is valuable to note that scatterers with characteristic lengths near the wavelength lead to a significant increase in radar backscatter. It is also important that Mie scattering allows for non-zero reflection in all directions.

Having discussed the origin of varying scattering contributions, we are still lacking a description of bulk scattering processes. We will give a brief overview here, condensing what is commonly found in textbooks such as Woodhouse, 2017 on this subject. With an electric field vector separated into an orthogonal basis

$$E = \begin{pmatrix} E^H \\ E^V \end{pmatrix}, \tag{2.15}$$

a common way to rewrite the four components of this field - two amplitudes and two phases - is the Stokes vector

$$I = \begin{pmatrix} I_H \\ I_V \\ U \\ V \end{pmatrix} = \begin{pmatrix} \langle E_H \overline{E}_H \rangle \\ \langle E_V \overline{E}_V \rangle \\ 2\ \mathrm{Re}\langle E_H \overline{E}_V \rangle \\ 2\ \mathrm{Im}\langle E_H \overline{E}_V \rangle \end{pmatrix}. \tag{2.16}$$

Whilst this decomposition seems quite arbitrary, it is historically motivated, as these components were directly measured in original experiments. In this basis, the scattering becomes a matrix multiplication with a matrix M of the form

$$I_s = M I_i, \tag{2.17}$$

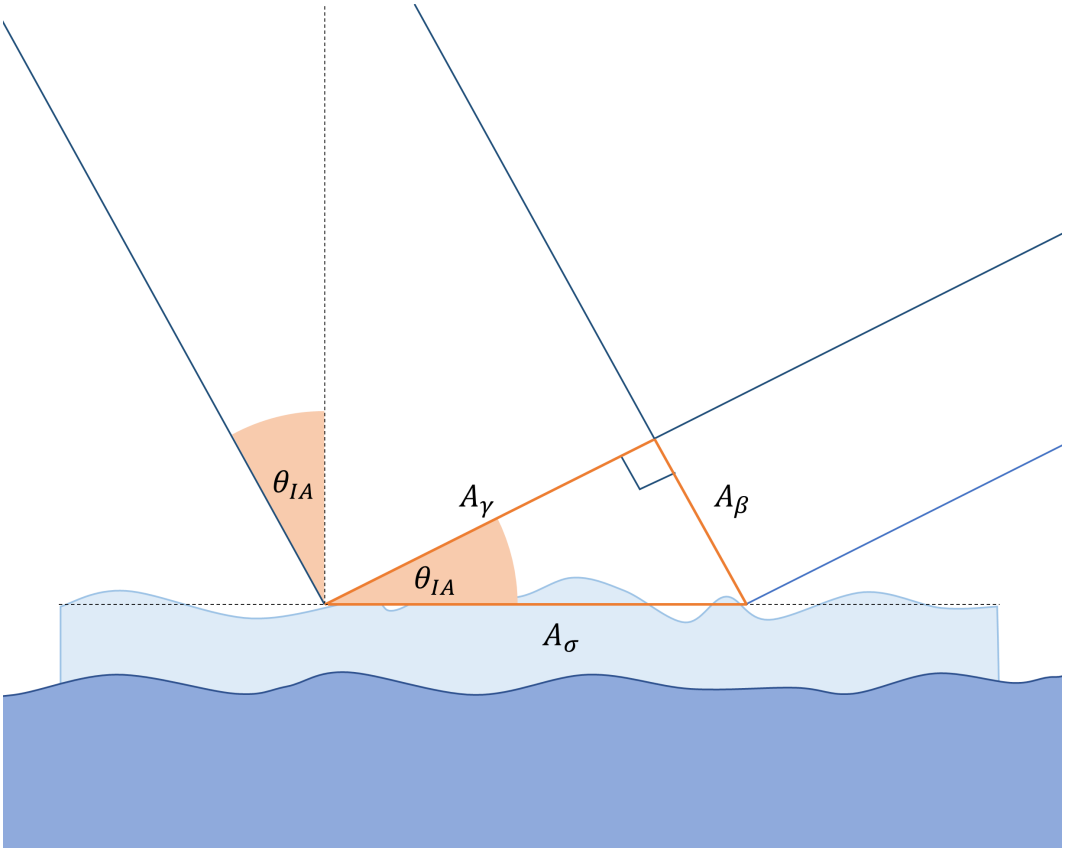


Figure 2.3: Ground geometry of the radar footprint transmitted normal to area A_γ under incidence angle θ_{IA}. A_σ is the area in the ground projection, A_β is the area in slant range and A_γ is perpendicular to the slant range.

with indices denoting scattered and initial wave properties. With an observer at a range r_0 and assuming isotropic scattering, the amplitude is attenuated under spherical symmetry, such that a received Stokes vector

$$I_r = \frac{1}{4\pi r_0^2} M I_i, \tag{2.18}$$

where r_0 is the distance from the observer to the scatterer. A convenient way to express the scattering of an incident field, particularly when considering that most radar instruments make two separate measurements for the horizontal and vertical polarisations, is to define the scattering matrix

$$S = \begin{pmatrix} S^{VV} & S^{VH} \\ S^{HV} & S^{HH} \end{pmatrix}. \tag{2.19}$$

The components describing same receiving and transmitting channels HH and VV are called *copol channels* and the others *crosspol channels*. The scattering matrix relates incoming and scattered fields as

$$\begin{pmatrix} E_s^H \\ E_s^V \end{pmatrix} = \frac{e^{-ikr}}{r_1} \begin{pmatrix} S^{VV} & S^{VH} \\ S^{HV} & S^{HH} \end{pmatrix} \begin{pmatrix} E_i^H \\ E_i^V \end{pmatrix}, \tag{2.20}$$

where r is the range coordinate and r_1 the distance of the scatterer. As per the reciprocity principle, the $S^{HV} = S^{VH}$ and thus often a vector of the scattering components

$$k = \begin{pmatrix} S^{VV} \\ S^{HV} \\ S^{HH} \end{pmatrix}, \tag{2.21}$$

is defined, so that the coherence of the channels can be written as the covariance matrix

$$C = k\overline{k}^T = \begin{pmatrix} \left|S^{VV}\right|^2 & S^{VV}\overline{S}^{HV} & S^{VV}\overline{S}^{HH} \\ S^{HV}\overline{S}^{VV} & \left|S^{HV}\right|^2 & S^{HV}\overline{S}^{HH} \\ S^{HH}\overline{S}^{VV} & S^{HH}\overline{S}^{HV} & \left|S^{HH}\right|^2 \end{pmatrix}. \tag{2.22}$$

The amount of energy that is returned to the radar is described by the radar cross-section

$$\sigma = \frac{I_r}{I_i} 4\pi r_0^2, \tag{2.23}$$

where the loss over distance is corrected as if the scattering was isotropic. In a spatial context, it is often sensible to relate the cross-section to the area that it was acquired over. This is done by dividing by the area A_σ of the measurement (footprint of the radar), giving the normalised radar cross-section

$$\sigma_0 = \frac{\sigma}{A_\sigma}. \tag{2.24}$$

An Earth-observing satellite radar does not natively measure in the plane of Earth's surface. In the satellite geometry, the radar measures a cross-section β over a surface that is perpendicular to the radar-looking direction A_β. As the intensity is the same in both coordinate systems, the cross-sections and the respective areas are related by the expression

$$\sigma_0 A_\sigma = \beta_0 A_\beta. \tag{2.25}$$

The two areas are related via the incidence angle θ (see fig. 2.3), via the geometric relation

$$\sin\theta = \frac{A_\beta}{A_\sigma}. \tag{2.26}$$

Combining the two equations gives the conversion from the satellite measured β_0 to the ground projected σ_0, as

$$\sigma_0 = \beta_0 \sin\theta. \tag{2.27}$$

The geometry we have used here inherently assumes a flat earth surface. In words, the normalized radar cross setion σ_0 describes, under the assumption of isotropic scattering, the ratio of the area that was measured versus that of a flat surface.

2.3 Scattering Simulations in 1D

To gain some intuition how surface roughness can influence radar backscatter, some simple 1D simulations of radar scattering from a random periodic surface are presented. The underlying assumption of the work presented here, is that the backscatter at the surface is

largely specular at the top of the surface and that the length scales (correlation lengths) of the roughness are significantly larger than the wavelength, so interference effects/Mie scattering can be ignored. The random surfaces considered are simply randomly generated Fourier series with diverse correlation lengths and surface heights. The energy returned to the instrument depends on the angle of the surface (for single bounce scattering). In the following we assume a strongly anisotropic scattering at every point of the surface, where most energy is returned if the surface normal is parallel to the incident wave and the intensity drops of exponentially as the the angle between surface normal and incident wave changes. An example for scattering at such a random surface can be seen in figure 2.4.

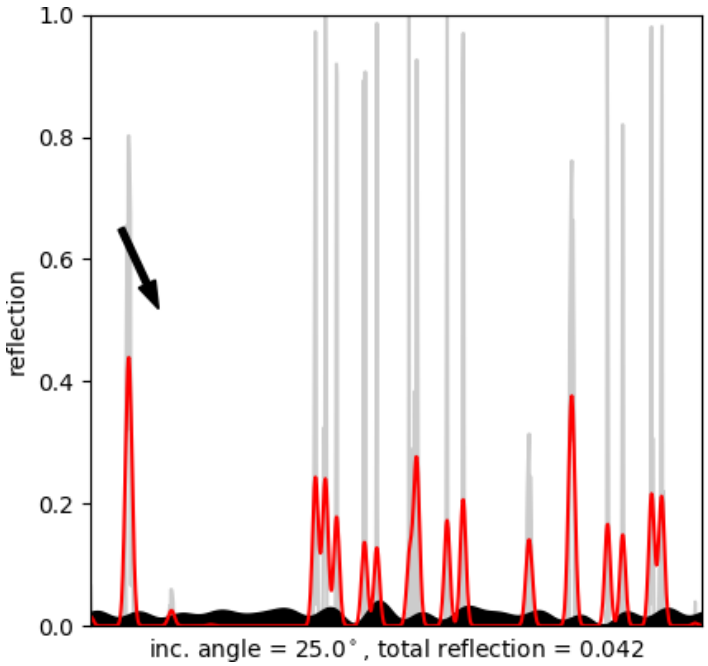


Figure 2.4: Scattering simulation of a random surface (seen at the bottom in black). The individual contributions of the subsampled surface are shown in a light grey in the background, the red line shows an aerial accumulation, i.e. the approximate local normalised backscatter at every point of the surface. The total reflected energy is also given. The arrow denotes direction if the incident wave. Some double bounce contribution can be seen in the rightmost peak of the triple-peaked contributions just left of the centre of the image. It must be double bounce that contributes here as the surface normal points away from the incident wave.

To investigate the influences of surface roughness correlation lengths, five different random surfaces are simulated with decreasing correlation lengths at incidence angle steps of a single degree. The results of these simulations can be seen in figure 2.5. There it is shown, that the lower the roughness correlation lengths are, i.e. the surface is 'smoother', the more anisotropic the scattering becomes. It might come as a surprise that a nadir incidence does not lead to the strongest radar backscatter. This is because for the sinusoidal oscillations used to describe the surface, the points where the slope is equal to zero (resulting in strong

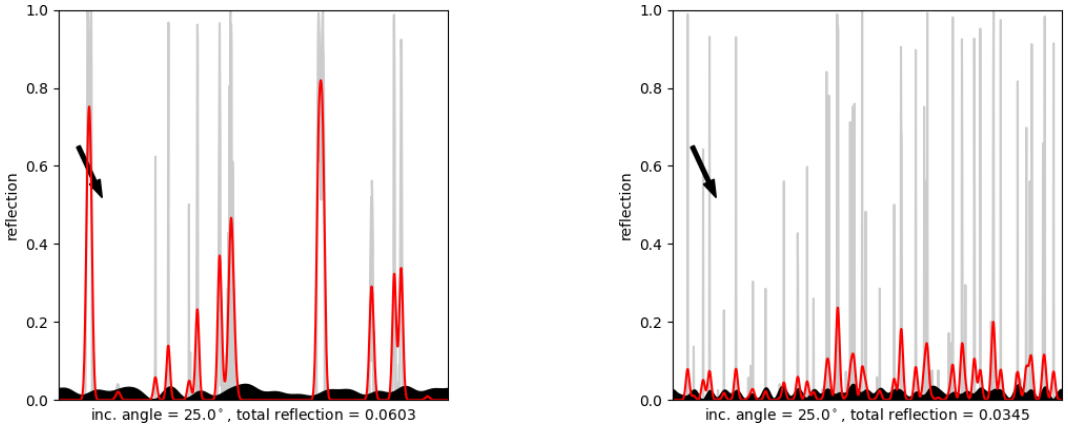
nadir reflection) are those were the changes in the slope are most extreme. Therefore the area where the slopes are close 0 is small. As expected the brightness falls of quickly when moving away from the incidence angles of strongest return as the strong slopes required to scatter back a signal at higher incidence angles are increasingly rare in smoother landscapes.

Finally, it is time to look at the region of incidence angles most interesting to us: the typical incidence angle range of a synthetic aperture radar satellite. The same results as in the polar plot are shown for this regime in figure 2.6. Here it is clearly visible how the gradients get shallower as the roughness correlation lengths decrease. The smoother surfaces drop off very quickly with increased incidence angles. This is also observed for open water in satellite images. In regards to total backscatter there does seem to be an optimal amount of roughness, as the most rough surface's backscatter, indicated in yellow-orange in the plot, is not as strong as other surfaces. In this simulation the reason for that behaviour is, that the regions of optimal slopes for radar returns lie in the shadow of other peaks. It must be noted that these simulations do not take into account backscatter with more than two bounces, which occur more often in these regimes, therefore one should be careful when interpreting the simulation results for very high roughness.

To summarise the results of the simulations: The rougher a surface (i.e. smaller roughness correlation lengths at the same surface height variance) is, the more it tends scatter more isometrically, meaning the radar return intensity does not drop of as quickly as the incidence angle increased. Also, across the incidence angles that are typically measured with a SAR satellite, rougher surfaces tend to be brighter, especially at high incidence angles. The incidence angle window where smooth and rough surfaces have approximately the same brightness does seem to lie within the incidence angle of the satellite (this is confirmed for open water visually, see figure 3.7 for an example).

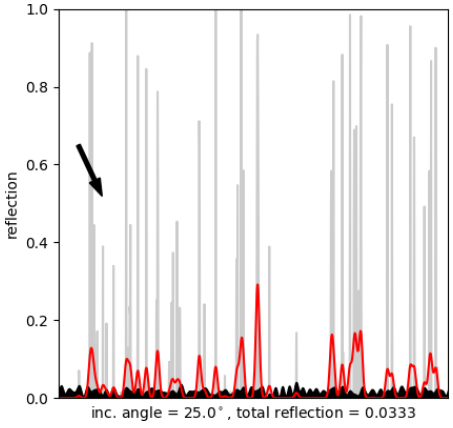
2.4 Observations at C-Band from Scatterometer and SAR

In real observations of the sea ice cover, one will almost always have contributions from multiple scattering regimes, as well as multiple bounces and it is very difficult to separate contributions of individual scatterers. Backscatter intensity over sea ice can be seen to be approximately linearly related to incidence angle (over incidence angle ranges seen by a satellite, see 2.2b for Fresnel scattering in these regions) and can qualitatively be related to the nature of the ice. As shown by the simulations above, level young ice and open water will have a strong specular response and thus the brightness will quickly fall off with increasing incidence angle. Frost flowers growing on freshly frozen leads or old, desalinated and deformed multiyear ice however, will have some scatterers near the wavelength of the sensor and/or allow for multiple bounces, leading to smaller dependence on the incidence angle. The aim of this section is to give some intuition of how to relate observations of radar backscatter to sea ice properties.

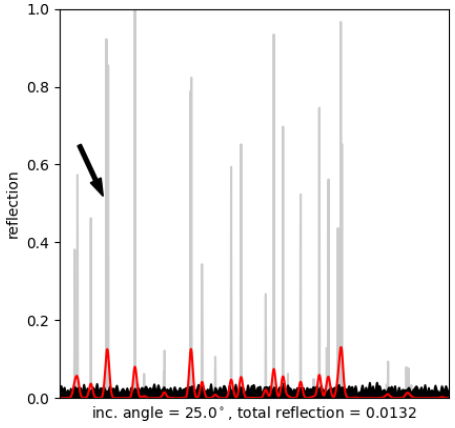


(a)

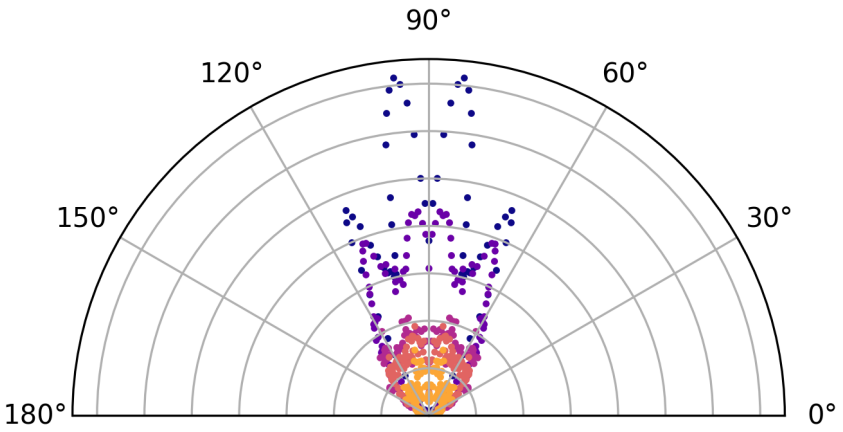
(b)



(c)



(d)



(e)

Figure 2.5: Scattering simulation of random surfaces with different correlation lengths (surfaces are seen at the bottom in black in a-d). The backscatter was simulated at every incidence angle between 0 and 180 degrees and the total reflection is shown on the radial axis of plot e. The colors go from blue to orange with increasing roughness (decreasing correlation length).

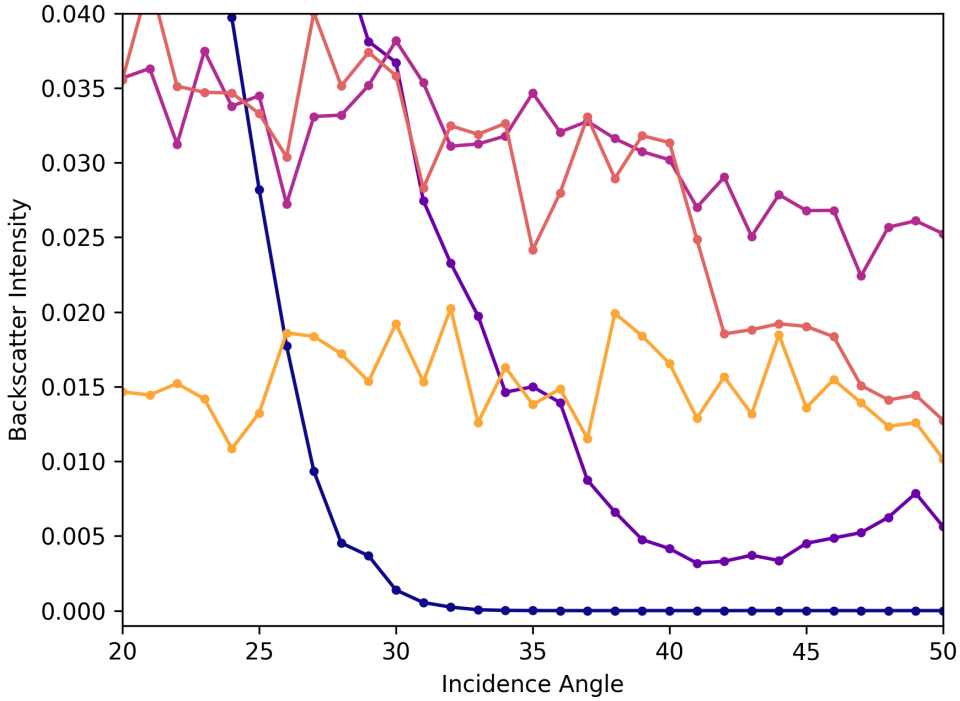
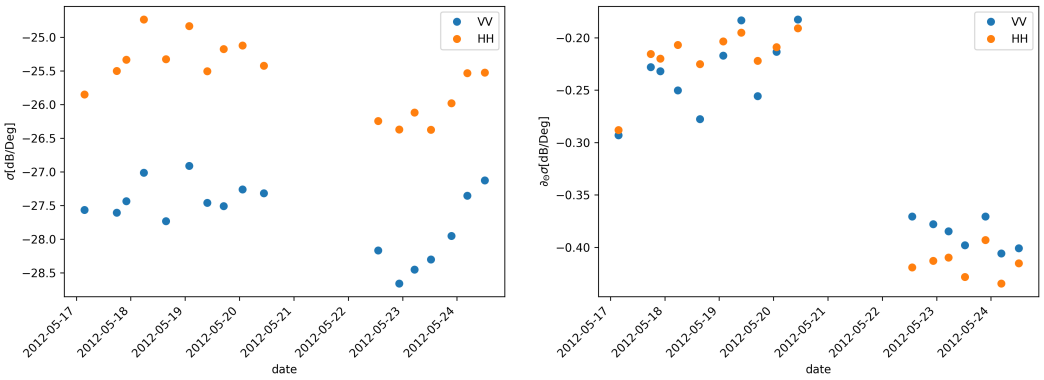
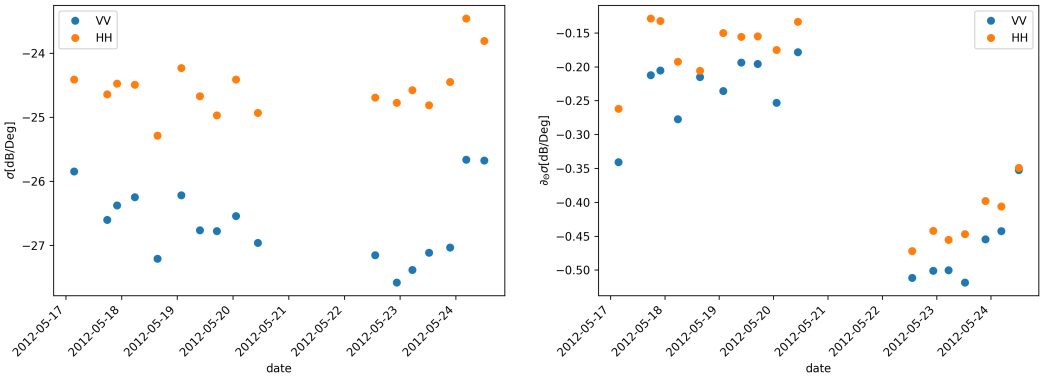


Figure 2.6: Same simulation as in figure 2.5, shown for the incidence angle window typical for satellite SAR sensors. As above, the colors go from purple to orange with increasing roughness (decreasing correlation length).



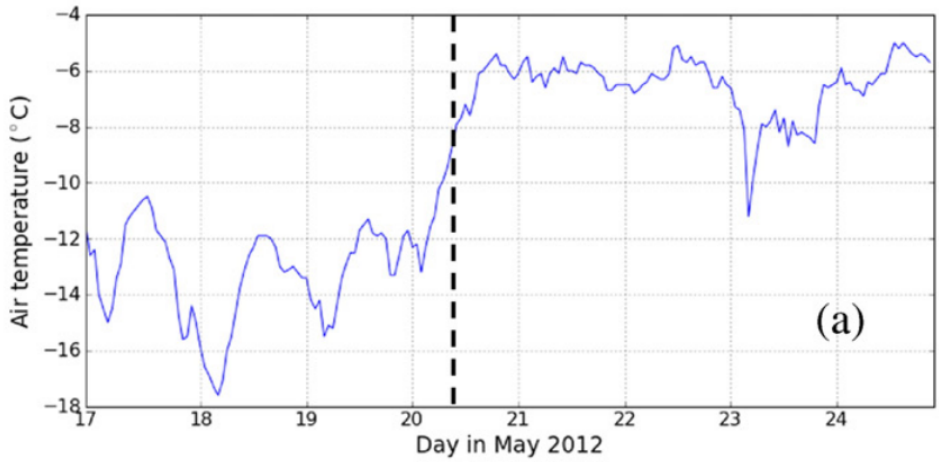
(a) mean backscatter

(b) slope



(c) mean backscatter in reduced IA range

(d) slopes from values in reduced IA range



(e) coincident air temperature from Nandan et al., 2017 fig. 4a

Figure 2.7: Scatterometer data from a warming event in Resulte Bay, Canada. The data is at C-band over a 16cm saline snow cover on top of landfast first-year ice. Research utilising this data is published in Nandan et al., 2017. The lower two scatterplots show the values if restricted to the incidence angle range 20-50 degrees, which is typical for SAR satellite imagery. The last figure at the bottom shows the air temperature and is a copy of figure 4a in Nandan et al., 2017. ©Elsevier.

Scatterometer A time series of real measurements made with an on-ice scatterometer during a warming event in Resolute Bay in Canada in May 2012 is shown in figure 2.7. Research has already been published on this data in Nandan et al., 2017, relating the instrument's backscatter to atmospheric conditions. The slopes' behaviour during the warming event is also interesting to us, but not yet explored in the paper. Especially when restricting oneself to the incidence angle range of SAR satellites, the signal of the warming event is much more significant in the slopes than the backscatter. This demonstrates the sensitivity of the incidence angle slopes to physical environmental changes. The observed ice in the experiment was first-year ice covered with a saline snow-pack of 16 cm depth. As shown in Nandan et al., 2017, models suggest that the snow depth is larger than the penetration depth of the radar, meaning the backscatter is dominated by the scattering from the snow layer. This is a rare case in sea ice remote sensing at these frequencies and is facilitated by the layer's high salinity. Typically the snow is mostly transparent at the frequencies of common SAR systems. So far mainly surface scattering (at the ice-atmosphere surface) has been discussed. For more realistic deductions to be made, contributions from scatterers inside the volume also need to be considered. This volumetric scattering is commonly considered quasi-random, as it originates from a multitude of differently oriented surfaces inside the volume. A possible explanation for the drastic impact on the slopes in the measurements could be that the warming of the snowpack homogenized the snow layer and destroyed the interior structure, responsible for the volumetric scattering (quasi-random multi-bounce, leading to more isotropic scattering behaviour). With such structures gone, the signal is dominated by specular reflections, which leads to a stronger incidence angle slope. To put it concisely, the warming probably smoothed the internal structures in the snowpack. Aside from demonstrationgh that incidence angle dependent slopes are valuable for insight into structural change, these observations also demnonstrate how radar response can change without changing the ice type. In this case this was due to a saline and damp snowpack on top of landfast ice, which is not commonly encountered on non-fast ice, however.

Synthetic Aperture Radar An example of sea ice's backscatter and incidence angle dependence as measured with a C-Band SAR sensor (Sentinel-1) can be seen in figure 2.8. The slopes in these scenes are computed using a technique published in Cristea et al., 2020 and is thus only a secondary product, not a real measurement. If we assume it to be representative of the real slopes, we can make some derivations about the structure of the observed ice. First, it is notable that the first- and multiyear ice (FYI and MYI) are best separated in the HV band and the HH slopes. In the HV band, multiple bounces are required to obain a radar response. As the brightness between FYI and MYI regions is comparable in HH band but different in HV it thus follows that MYI has more surface deformation and/or higher penetration depths of the signal to facilitate multiple bounces (High penetration depth would allow scattering at more interfaces in the crystal). The HH slope for MYI being less steep than for FYI also suggests that we have increased volumetric scattering. Because of the quasi random nature of volumetric scattering, these contributions should

exist similarly in the HV band. Hence, the surface scattering must be stronger for FYI for both ice types to have similar total backscatter. This is slightly atypical, as multi-year ice often has significantly stronger backscatter in the HH band.

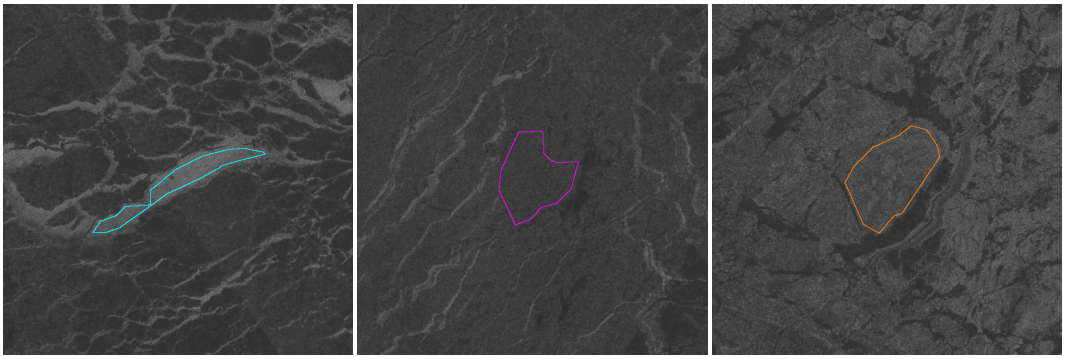
In summary, regarding the FYI and MYI differences, the data suggests the FYI section has stronger surface scattering contributions, whilst the observed MYI has more volumetric scattering. The young ice in this example has high backscatter in both channels but steep negative slopes. The high backscatter of young ice (YI) combined with low penetration depth due to increased salinity in the upper layers there, suggest the presence of some structures on top of the ice scattering near the instrument's wavelength (5,5 centimetres in this case of Sentinel-1). As the slopes are very steep, these structures are not randomly oriented and seem to have much stronger preferential orientation than the MYI contributions. One such phenomenon that fits this contribution is the presence of frost flowers: thin ice crystals on top of young ice. These grow through desublimation of the water in the warm air over the ice. This air is warmed by the latent heat of the ice growth below and drops in temperature rapidly in a substantially cooler atmosphere.

2.5 Radiative Transfer Modelling

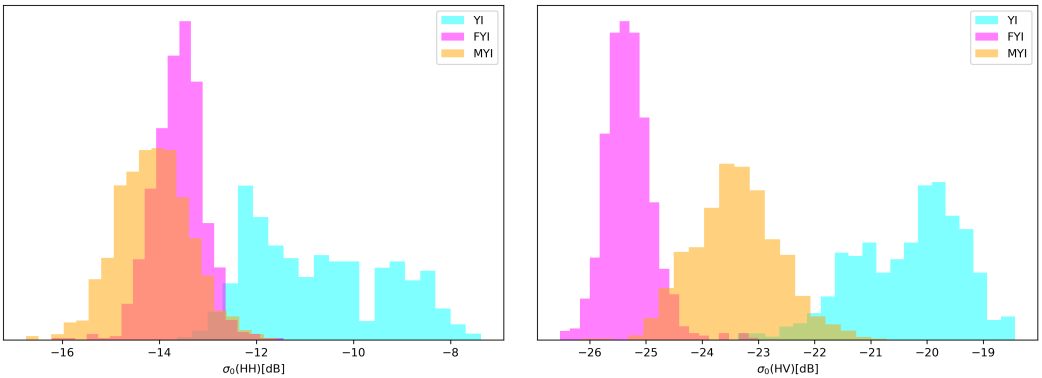
Although this work focuses on the observation of sea ice across various spatial and temporal scales, to comprehend the challenges of sea ice as a medium, we will briefly introduce a modelling approach, summarising work from Golden et al., 1998 and giving appropriate background information, that can be found in e.g. Woodhouse, 2017. It is only one of many ways that sea ice can be modelled, but it is an intuitive approach that outlines the difficulty of the modelling efforts. This illustrates why retrieval by inversion of forward models is currently not a feasible approach to monitoring the sea ice masses of our planet.

One property that can be readily derived from the permittivity and conductivity is the penetration depth of an electromagnetic wave: the distance in the medium where the amplitude of the electric field has decreased by 1/e (see eq. 2.8). These calculations, however, assume a homogenous medium, which sea ice is very much not. Similar problems of relating sea ice's dynamic and highly variable nature to bulk properties is challenging. Still, it cannot be avoided when incorporating sea ice into models. This process is termed homogenization, as one tries to describe the behaviour of a very heterogeneous material by a homogenous one. Because the interactions being modelled are highly nonlinear regarding these variables, simply taking a mean of some local parameters is not a successful approach. In the case of electromagnetic wave interaction, one example of such homogenization approaches is to extend the extinction coefficient $\kappa = 1/\delta$ where δ is the penetration depth, by an additional scattering term to account for scattering along interfaces of the polycrystalline structure, brine channels and air bubbles.

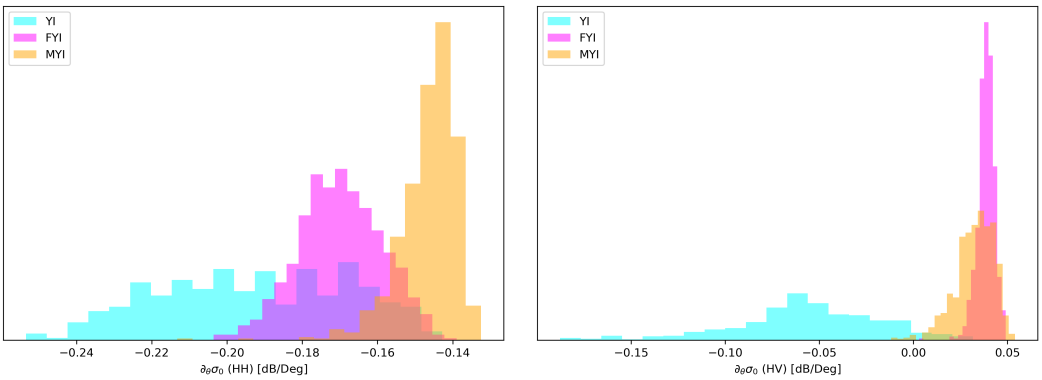
First, a general case is described, wherein a transparent volume of area A and depth h encloses scatterers without orientation, fully defined by their radar cross section σ with



(a) YI (b) FYI (c) MYI



(d) (e)



(f) (g)

Figure 2.8: Histograms of backscatters and incidence angle dependence slopes for the three different ice regions (first row - HV) labelled manually. The slopes where determined using a Gaussian clustering-based method, published in Cristea et al., 2020. The data is from a Sentinel-1 EW mode acquisition of the central Arctic in November 2019.

density ρ. Defining thus an effective extinction coefficient

$$\kappa_{\text{eff}} = \rho(\kappa_a + \kappa_s), \tag{2.28}$$

with the extinction coefficient from the scatterers material is κ_a and the newly appended scattering term κ_s. The effective opacity of the volume along a path $s = [0, h]$ is

$$\kappa \cdot s = \kappa \frac{h}{cos\ \theta}. \tag{2.29}$$

Computing the scattering from a beam with footprint B, incident at angle θ, we sum over all contributions over the path s and the area, to obtain the cross-section

$$\begin{aligned} \sigma_{\text{vol}} =& \frac{B}{\cos\theta} \int_0^h \rho\sigma e^{-2z\kappa_{\text{eff}}/\cos\theta} dz \\ =& \frac{B}{\cos\theta} \frac{\rho\sigma}{2\kappa_{\text{eff}}} \cos\theta \left[-e^{-2z\kappa_{\text{eff}}/\cos\theta} \right]_0^h \\ =& B \frac{\rho\sigma}{2\kappa_{\text{eff}}} (1 - e^{-2h\kappa_{\text{eff}}/\cos\theta}). \end{aligned} \tag{2.30}$$

With the area relation $B = A\cos\theta$, where A is the area on the ground, we can now compute the normalised radar cross section

$$\sigma_0 = \frac{\sigma_{\text{vol}} \cos\theta}{B} = \cos\theta \frac{\rho\sigma}{2\kappa_{\text{eff}}} (1 - e^{-2h\kappa_{\text{eff}}/\cos\theta}). \tag{2.31}$$

The problem with this model is that such scatterers with cross sctions independent of orientation are not realistic for sea ice, where brine channels and ice growth typically have preferred directions. For simpl air bubble inclusions, this could be suitable, but the effect of the bulk medium have not yet been taken into consideratoin.

In the next stage, the assumptions about the orientation of the scatterers are relaxed, and the absorption of the bulk medium of the volume is also taken into account. As covered in the introduction of this work, sea ice growth undergoes various stages, from frazil ice consolidation to layering of nilas to formation of pancake ice and eventual columnar ice growth. The snow on top forms its own diverse structures, as older snow at the bottom can be compressed or even flooded. Due to the different sea states, growth/development differs not only regionally in the polar regions but especially when comparing Arctic and Antarctic conditions. The openness of the Antarctic, in contrast to the sheltered Arctic, is the central reason for these disparities.

Inspired by the different growth regimes, the following model will approximate the sea ice cover as multilayered random media. For example, these layers could be: Fresh snow, compressed snow, consolidated young ice, columnar ice and porous bottom ice encapsulated by the atmosphere and ocean. Scattering in these layers will be approximated by enclosed scatterers with given distributions of shape and orientation embedded in a medium with given background permittivity. The model also does not incorporate effects of coherency, interference or resonance - essentially reducing the fields to particle-like interactions.

We define the model as having $N + 1$ layers given by layer number $n = \{0, 1, \ldots, N\}$ with boundaries located along the z axis at $z = \{0, -d_1, \ldots, -D_N\}$. Each layer has a

background permittivity ϵ^n and is populated by a number of different types of scatterers M_n, each of which is parametrised by a fractional volume $p^s_{n_j}$, permittivity $\epsilon^s_{n_j}$, as well as size and orientation distributions. The interfaces can be parametrised with a roughness (standard deviation) and a correlation length.

The differential equation for the intensity of a field in layer n, propagating along θ, ϕ is given by the differential equation

$$cos(\theta)\frac{d}{dz}I(\theta,\phi,z) = -\kappa(\theta,\phi)I_n(\theta,\phi,z) + \int_{4\pi} d\Omega' P_n(\theta,\phi,\theta',\phi')I_n(\theta',\phi',z). \tag{2.32}$$

I_n is the Stokes vector, as introduced previously - P_n is the phase matrix and κ is the extinction coefficient of the bulk medium. To recap, the Stokes vector in an orthogonal basis denoted by polarisation H, V is given with the characteristic impedance η is

$$I_n = \begin{pmatrix} I^H \\ I^V \\ U \\ V \end{pmatrix} = \frac{1}{\eta}\begin{pmatrix} \langle E^H \overline{E}^H \rangle \\ \langle E^V \overline{E}^V \rangle \\ 2\ \mathrm{Re}\langle E^H \overline{E}^V \rangle \\ 2\ \mathrm{Im}\langle E^H \overline{E}^V \rangle \end{pmatrix} \tag{2.33}$$

.

Scattering of the incident field $E^i = (E^H, E^V)$ in direction (θ^i, ϕ^i) to scattered field $E^s = (E^H, E^V)$ in direction (θ^s, ϕ^s) is given by a scattering matrix

$$S = \begin{pmatrix} S^{VV} & S^{VH} \\ S^{HV} & S^{HH} \end{pmatrix}. \tag{2.34}$$

The corresponding stokes vector of the incident and scattered field are given by a 4x4 matrix $L(\theta^i, \phi^i, \theta^s, \phi^s)$ which is dependent on S. From the matrix L, the phase matrix P is computed by averaging (ignoring coherency effects) over the parametrisations of the random scatterers with given joint probability distributions f. For example, consider random scatterers of cylindrical shape, with size parametrisation a, b and orientation parametrised by α, β. The joint distribution $f(a, b, \alpha, \beta)$ then defines the phase matrix

$$P(\theta^i,\phi^i,\theta^s,\phi^s) = \rho \int da \int db \int d\alpha \int d\beta\ f(a,b,\alpha,\beta)L(\theta^i,\phi^i,\theta^s,\phi^s), \tag{2.35}$$

where ρ is the number of scatterers (per unit volume).

Considering that the bounds of size and orientation are free parameters as well as the probability distribution itself having to be parametrized, it is clear that even such a simplified contribution in a single layer already results in a large amount of parameters that would have to solved when trying to invert such a model. Combined with the effects of other layers it is immediately obvious how relaxing some of the super simplistic assumptions from the initial modelling approach results in a model with very high number of parameters. This is despite still being a large simplification of reality with coherency effects and geometry of the interfaces not being taken into account. Especially when regarding that often SAR acquisitions have only one frequency and two polarization channels, inversion is thus not remotely feasible.

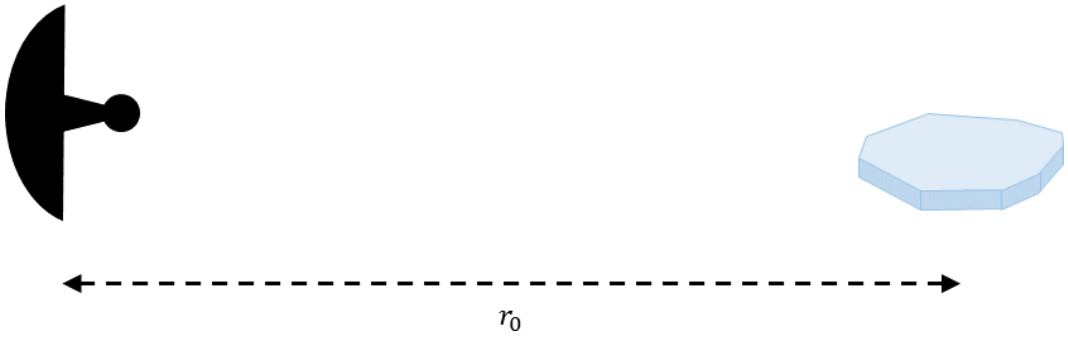


Figure 2.9: The basics of radar feature a receiver and a target. The fundamental measurements that a radar makes are the electromagnetic field and time.

2.6 Introduction to Synthetic Aperture Radar

To observe the Arctic at high resolution (10s of meters) from space, requires the use of active sensors. As energy should not be lost to atmospheric absorption, microwave frequencies in the microwave atmospheric window (roughly $10^{-3} - 10^{1}$ metres) are used most commonly. The optimal product should yield a two dimensional mapping of the surface, to achieve high spatial coverage. One technology that enables an acqusition of that type is a radar system with synthetic aperture. The core feature of such an instrument is that the observed backscatter can be observed in two dimensions, where the distance to the scatterer has no influence on the measured signal (bar incidence-angle effects).

2.6.1 Radar Basics

Before turning to radars with synthetic apertures, the basic principles of standard radar are established. Modern techniques started to be developed in the 1930s in several countries independently. A historical overview is given in Skolnik, 1985. Radar (RAdio Detection and RAnging) is a technique to measure the distance of far away objects by sending a signal and timing the reflected echo (fig. 2.9). If the transmitted and received signals differ by a time Δt, then the range r_0 of the reflected target is

$$r_0 = \frac{c\Delta t}{2}, \tag{2.36}$$

where c is the speed of light in the respective medium and the factor 2 arises because the signal travels the distance twice (to the target and back). Because a pulse of finite, non-zero length is transmitted, the radar has only finite resolution in measuring the range. Subjects need to be apart from one another by at least one *pulse length* t_p, so the range resolution δ_r is

$$\delta_r = \frac{c\ t_p}{2}. \tag{2.37}$$

It is much more common, however, to characterise radar systems not by their pulse length but their *bandwidth* Δf, which is the difference of the maximum and minimum transmitted

frequency

$$\Delta f = f_{max} - f_{min}. \tag{2.38}$$

How bandwidth and pulse length are related is not obvious. In literature, one often finds the relation

$$\Delta f = \frac{1}{t_p}. \tag{2.39}$$

To understand this relation, we need to look at the signal in frequency space. A pulse of length t_p in an electric circuit is defined by a voltage

$$u(t) = \begin{cases} 0 & \text{for } |t| < \frac{t_p}{2}, \\ 1 & \text{else.} \end{cases} \tag{2.40}$$

Note that such a pulse has Energy $E_p = It_p$, with current A. For simplicity we choose the current $I = 1$ and thus

$$E_p = t_p. \tag{2.41}$$

A pulse from an antenna is much easier described in the frequency space of electromagnetic waves. Such a rectangular pulse as we are describing here is made up of an infinite amount of frequencies, described by the Fourier transform

$$\hat{u}(f) = \int_{-\infty}^{\infty} u(t) e^{i2\pi ft} dt. \tag{2.42}$$

Inserting $u(t)$ reduces the integral and one obtains the following expression:

$$\hat{u}(f) = \int_{-\frac{t_p}{2}}^{\frac{t_p}{2}} e^{i2\pi ft} dt, \tag{2.43}$$

which has an analytic solution taking the form

$$\hat{u}(f) = \frac{\sin(\pi f t_p)}{\pi f}. \tag{2.44}$$

The amount of energy transferred through a bandwidth Δf is given by

$$E_f = \int_{-\frac{\Delta f}{2}}^{\frac{\Delta f}{2}} \hat{u}(f)^2 df. \tag{2.45}$$

Now looking at a bandwidth given by $\Delta f \cdot t_p = N$, the Energy transmitted through the frequencies in that bandwidth (for some N) is

$$E_f = t_p \cdot \begin{cases} 0.774 & \text{for } N = 0.5, \\ 0.903 & \text{for } N = 1, \\ 0.931 & \text{for } N = 1.5, \\ 0.950 & \text{for } N = 2, \\ \vdots & \vdots \end{cases} \tag{2.46}$$

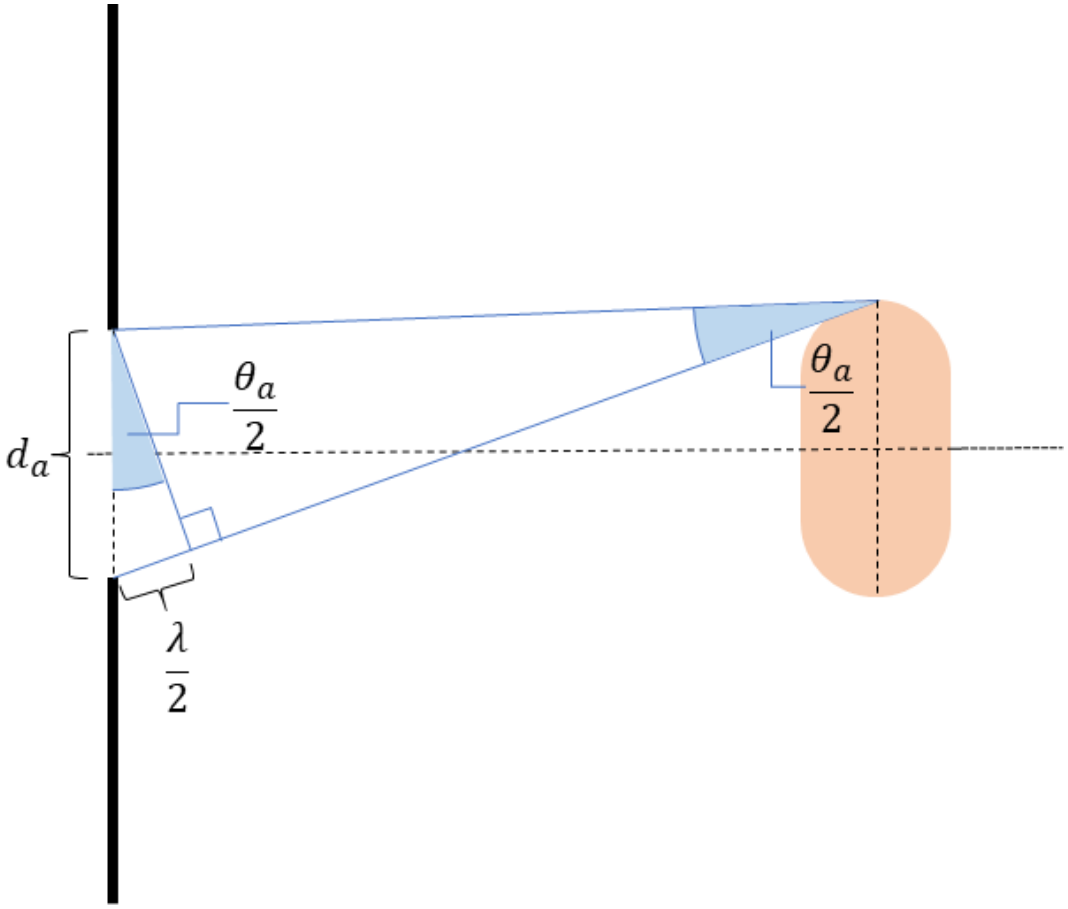


Figure 2.10: Relation of antenna opening angle to antenna length d_a. Note that this displays half the opening angle, as it points from the center to the first minimum and the full opening angle lies between the first minima. The shaded orange area represents the illuminated footprint.

So now the choice is made to set $N = 1$, meaning the bandwidth describes the range of frequencies through which 90.3% of the energy is transmitted. Hence one arrives at the previously given relation

$$\Delta f = \frac{1}{t_p}. \tag{2.47}$$

Inserting this into the above formula 2.37, the range resolution can be described by,

$$\delta_r = \frac{c}{2\ \Delta_f}. \tag{2.48}$$

Also of particular interest is the *angular resolution* δ_a of a radar system. Principally, this is equal to the angle at which two objects in space echo responses of the radar can still be separated in the returning signal. If one imagines a classic radar, it is thus exactly the opening angle of the beam θ_a. When originating from an emitting antenna, the width of the antenna beam is understood as the angular distance between the first interference minima on each side of the central maximum in the antenna pattern. The angular distance between them can easily be computed geometrically: A minimum occurs when the beam length from the far end of the antenna is half a wavelength longer than from the closer end

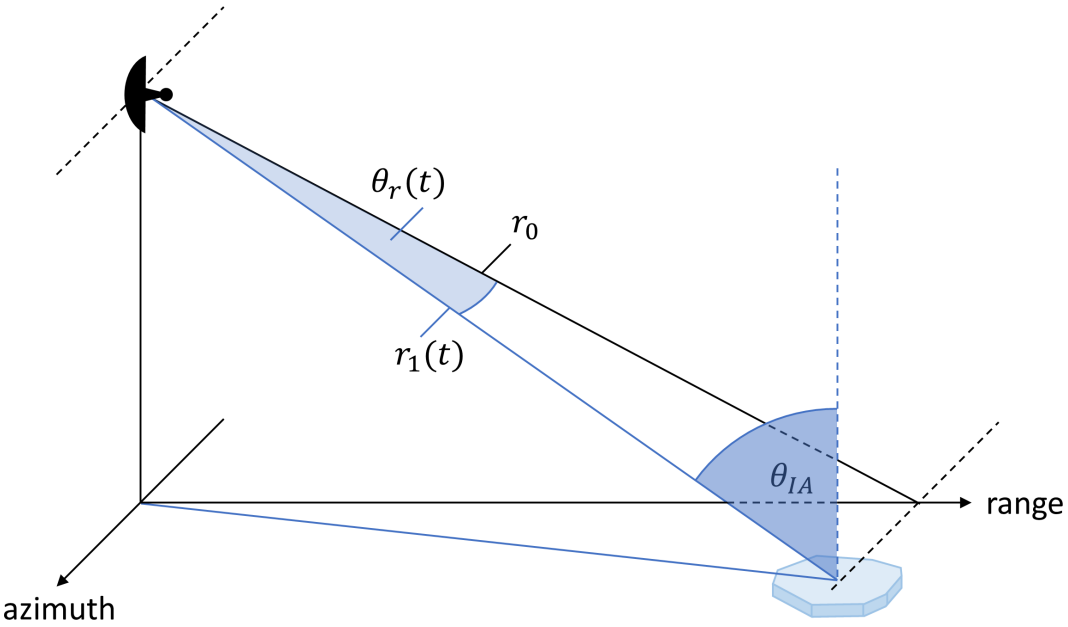


Figure 2.11: The geometry of a synthetic aperture radar (SAR) system.

of the antenna 2.10. As we are considering far field approaches, the triangle of small opening angle θ_a, opposite side $\lambda/2$ and the antenna length d_a as the hypotenuse, gives the relevant relation

$$\sin(\frac{\theta_a}{2}) = \frac{\lambda}{2d_a}. \tag{2.49}$$

For small opening angles $x \ll 1$ the Taylor series expansion of $\sin(x) \approx x$ and thus we get the common formula for the angular antenna resolution.

$$\delta_a = \theta_a = \frac{\lambda}{d_a}. \tag{2.50}$$

Now it is straightforward to compute the resolution in meters δ_{az}, for a target at distance r_0. We use a subscript az to refer to the azimuth dimension. For a synthetic aperture radar this is the dimension along the flight direction. Again given by the geometry

$$\tan(\theta_a/2) = \frac{\delta_{az}}{2r_0}. \tag{2.51}$$

Considering only small opening angles, the Taylor approximation yields

$$\delta_{az} = \theta_a r_0 = \frac{\lambda}{d_a} r_0 = \frac{c}{f}\frac{r_0}{d_a}. \tag{2.52}$$

2.6.2 Synthetic Aperture Radar

The principle design feature of radar with *synthetic aperture* (SAR) is that the sensor is moving relative to the observed object. Thus the target is illuminated for some part of the trajectory of the sensor. The length of the path of the sensor, where the object is illuminated is a synthetic aperture l_{sa} - similar to a very long antenna. Note that the length of the path of the sensor is equivalent to the beamwidth at the distance of the target r_0. With the

opening angle of the real antenna θ_a, we use the same formula (eq. 2.52) as for classic radar, as it describes the same geometry.

$$l_{sa} = \delta_a r_0 = \frac{\lambda}{d_a} r_0 \tag{2.53}$$

Now we want to compute the opening angle θ_{sa} of the synthetic antenna. With an effective antenna length l_{sa} and equation 2.50, one obtains

$$\theta_{sa} = \frac{\lambda/2}{l_{sa}}. \tag{2.54}$$

The extra factor of one-half of the wavelength might be unexpected here, as it contrasts the regular radar antenna formulation. It arises from the fact that interference happens twice in a SAR system, once on the way to the target and once back to the sensor, because the trajectory of the target relative to the sensor forms an aperture of its own. In a basic radar system, the target is stationary; thus, this term does not arise. If we substitute the length l_{sa} into the established formulation for θ_{sa}, we get

$$\theta_{sa} = \frac{d_a}{2r_0}. \tag{2.55}$$

Armed with that knowledge, it is possible to compute the azimuth resolution δ_{az} with help of equation 2.52., as

$$\delta_{az} = \theta_{sa} r_0 = \frac{d_a}{2}. \tag{2.56}$$

The crucial element of this relation is that the azimuth resolution is independent of the distance to the target r_0. This is the key concept that enables the acquisition of scenes with synthetic aperture radar, where (barring incidence angle dependent scattering) an objects response does not depend on the location in the scene. Optical images generally have a similar translationally invariant property, which is why the topics of image processing and SAR processing can be successfully related.

2.6.3 Decoding the Signal

In this section an analysis of the returning signal will be conducted. The central problem herein lies in the fact that the radar instrument measures a one dimensional signal (wave amplitude over time), which has to be converted into a two dimensional target space. Such a mapping means that

$$\psi : E(t) \mapsto E(a, r), \tag{2.57}$$

where we chose the coordinates azimut a and range r as the axes in two dimensional target space. It is clear that this mapping could only be surjective if its image is dense. This could only be realized with an infinitely long pulse (or one with infinite frequency), which is not feasable in the real world.

Having established that the mapping is not going to lie dense in the target space, is a first sign that discretisation of target space will be necessary to evaluate the signal. For now a closer look at the information contained the return signal $E(t)$ is in order. There are three

components that will prove helpful decoding the return signal: the amplitude, the time delay and the frequency.

The amplitude $E(t)$ at a given time (t) gives the strength of the returning signal. This will be mapped as brighter spots in the SAR scene and relates to the radar cross section σ_0 of the object in target space. The time delay $t_{in} - t_{out}$ is directly proportional to the distance of the target r_1, related via the speed of light

$$t_{in} - t_{out} = 2\frac{r_1}{c}. \tag{2.58}$$

Realise that the distance to the target r_1 is not identical to the range coordinate r_0. Knowing the distance only allows one to place the target on a curve within the radars lightcone. The frequencies f of a signal are shifted by the Doppler effect, as the radar is operating on a moving platform. Because signals are send and received in the same location, one can place oneself in the inertial system of the moving platform and continue with the classical Doppler effect, not the relativistic one (relativistic effects are results of time dilation between two different inertial systems). The Doppler effect is easy to compute. A shift of the wavelength occurring when an observer moving at speed v sends or receives a signal is given by the function

$$D : \lambda, v \to \lambda - \frac{\lambda}{c}v, \tag{2.59}$$

where λ is the wavelength of the emitted signal and v is the observer's velocity relative to the signal's source. In the case of radar, there will be two shifts occurring. Assuming the speed of the platform is much smaller than the speed of light, one can simplify and say both shifts occur at the same speed v_θ, which is the component of the speed of the platform that is parallel to the vector pointing at the target

$$v_\theta = \sin(\theta_r)v, \tag{2.60}$$

with the absolute value of the velcoity v and the angle θ between the target vector $\vec{r}_1$ and vector $\vec{r}_0$, which points along the shortest distance between the flight path (parametrised by $\vec{v} \cdot t$) and the target (see figure 2.11).
The final received wavelength λ_{in}after two doppler shifts (once when emmitting, once when receiving) is thus related to the outgoing wavelength λ_{out} and v_θ by

$$\lambda_{in} = D(v_\theta, D(v_\theta, \lambda_{out})) = \lambda_{out} - 2\lambda_{out}\frac{v_\theta}{c} + \mathcal{O}\left(\left(\frac{v_\theta}{c}\right)^2\right). \tag{2.61}$$

In the following, we will ignore second and higher order terms in $\frac{v}{c}$. Before we get to calculating the frequency from this, there is an additional effect that occurs. Namely, the phase of the wavefunction changes relative to the absolute distance r_1 to the target. The phase dependent on the distance r_1 via

$$\varphi(r1) = -\frac{4\pi}{\lambda}r_1. \tag{2.62}$$

To interpret the signal, it is paramount to understand how the frequencies of the signal behave for target objects at different ranges r_0. Let's choose a time coordinate t, such that

$t = 0$ at the point of closest approach to the target. Under the assumption that the angle θ_r between $\vec{r}_1$ and $\vec{r}_2$ is small, we will be satisfied with a first order Taylor approximation of the distance:

$$r_1 = r_0 + \frac{v^2 t^2}{2r_0}, \tag{2.63}$$

which we plug into the relation of the phase, which is now dependent on r_0 and t, as

$$\varphi(r_0, t) = -\frac{4\pi}{\lambda}\left(r_0 + \frac{v^2 t^2}{2r_0}\right). \tag{2.64}$$

Frequency is the derivative of the phase over time (up to a factor of 2π) and thus the frequency will receive another modulation (on top of the Doppler shift) from the phase. Finally, the relationship between incoming and outgoing frequencies in its entirety is

$$f_{in}(t) = \frac{c}{\lambda_{in}} + \frac{1}{2\pi}\frac{d\varphi}{dt}. \tag{2.65}$$

Computing this to first order in $\frac{v}{c}$ and using the relation from the Doppler shift (eq. 2.61) gives

$$f_{in}(t) = f_{out} + 2\frac{v_\theta}{\lambda} - 2\frac{v^2 t}{\lambda r_0}. \tag{2.66}$$

With $v_\theta = \sin(\theta_r)v$ and the geometric relation $\sin(\theta_r) = \frac{vt}{r_1}$ this can be simplified to

$$f_{in}(t) = f_{out} + 2\frac{v^2 t}{\lambda}\left(\frac{1}{r_1(t)} - \frac{1}{r_0}\right). \tag{2.67}$$

finally one can taylor expand $r_1(t)$ as we have already done previously, yielding the final expression

$$f_{in}(t) = f_{out} - \frac{v^4 t^3}{\lambda r_0^2}. \tag{2.68}$$

Thus, by measuring the time delay and the frequency of the return signal, we can, in theory, invert the signal to find the position of the scatterer with strength given by the amplitude of the returned signal (something we have not made use of so far). Now one might think: Wow, we have just decoded a 1-dimensional signal and got full information about an object in 2 dimensions. Have we broken some fundamental mathematics? The answer of course is no; this is not witchcraft. This inversion is only possible for a single scatterer. As soon as multiple scattering signals overlap, the inversion will be under-determined. In reality, we do not want to make an assumption about the number of scatterers, thus the techniques used need to be different.

2.6.4 SAR Processing

In realistic SAR processing, a discrete matrix that has azimuth in one and range in the other axis is filled. This is basically done by discretising the return signal into fixed bins (organised in rows) and for every pulse a new column is started. Then filter functions are computed that show, what a point target scatterer's response would look like in this regime in range and azimuth direction. The image is then convolved with these filter functions, first along range, then azimuth, to reconstruct the image. This discretisation can lead to

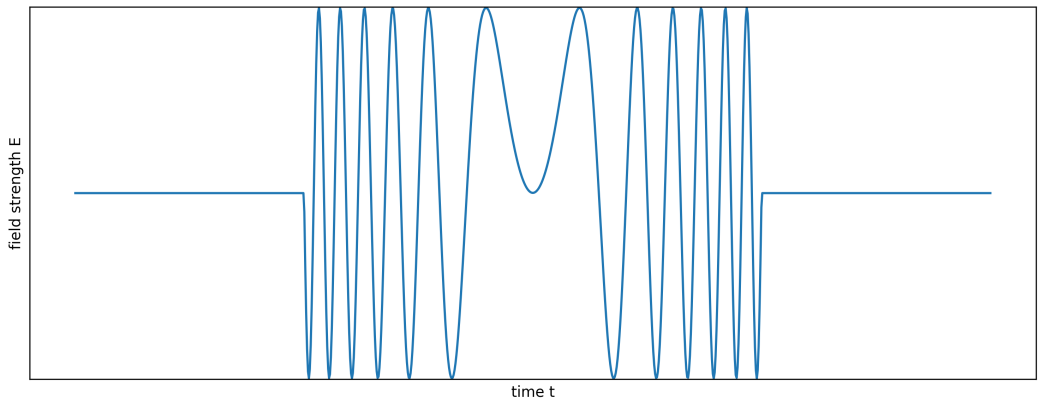
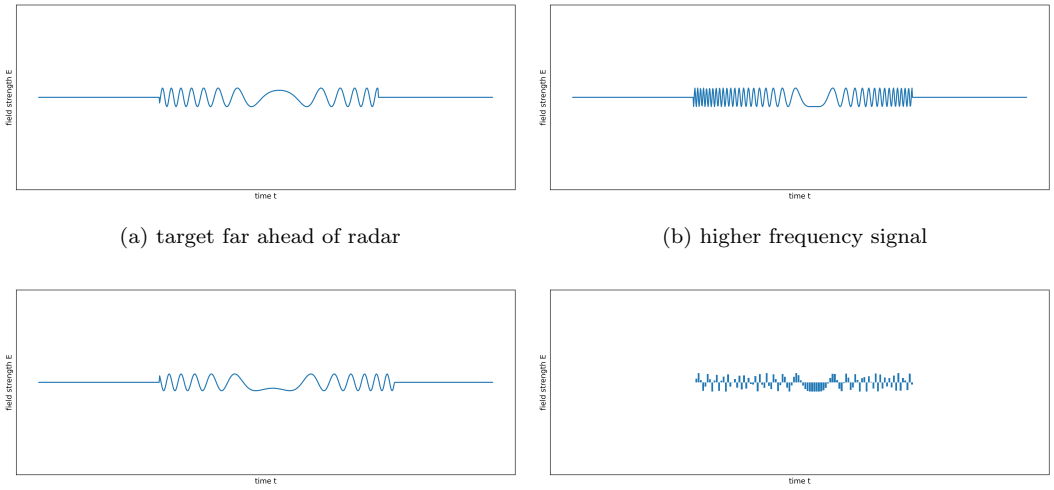


Figure 2.12: An example for a chirped signal as is used by synthetic aperture radars.



(a) target far ahead of radar

(b) higher frequency signal

(c) target far behind radar

(d) discretised higher frequency signal

Figure 2.13: (a-c) Examples of distorted return signals, at different times of observation. (d) Discretisation of signal (b) at a sampling rate lower than the maximum frequency of the signal. The modulation in all signals is due to Doppler shift. When the target is ahead of the satellite, the velocity component in range direction is positive and the signal is shortened (a). When the radar moves away from the target, the signal is lengthened (c).

the presence of some artefacts from strong scatterers. Luckily, this is generally not the case for sea ice acquisitions, as the backscatter is not high enough.

To explain in a little more detail, we will start by looking at the signal produced by the instrument. The signal emitted by a SAR sensor is chirped, meaning the frequency linearly increases with time. A signal of such nature can be seen in figure 2.12. The range of frequencies of this chirp is known as the bandwidth of the radar. As we can see from our description in frequency space, equation 2.68, upon return the signal is modulated by the Doppler effect and the phase shift. This incoming signal is thus distorted depending on the location of the scatterer. At the edge of the forward lightcone of the radar, the frequency gets increased by the Doppler shift and the signal becomes shorter. When the radar has passed the target and looks backwards, the frequencies are lowered and the signal is lengthened. An example for these phenomena can be seen in figure 2.13. At the point of closest approach, the signal will remain unmodulated, because the relative velocity between the target and the radar in range direction is zero.

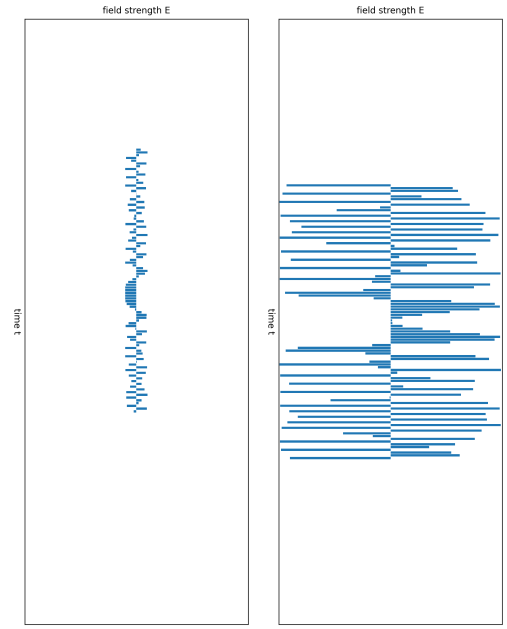
To now write the received signal intensities to a matrix, we need to discretise them. To give a more realistic example, we will choose a sampling frequency that is slightly lower than the maximum signal frequency. This will result in some image artefacts down the road. For an example of this discretisation, regard figure 2.13 d.

Finally, these discrete signals are written to the target matrix as columns. For an illustration of this procedure, see figure 2.14. The ambiguities introduced by undersampling the frequency are concentric formations above and below the signal centre.

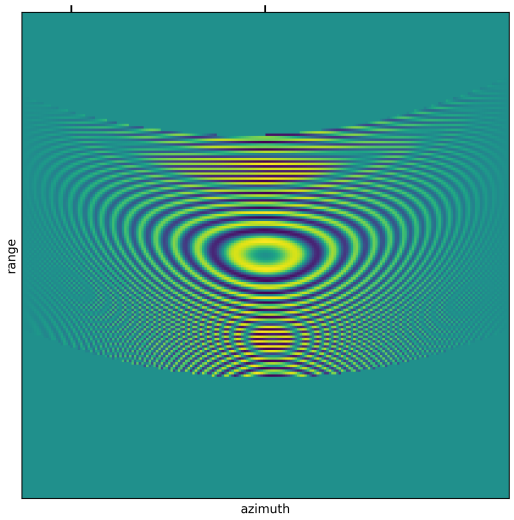
The next steps are to focus the SAR imagery. The range focussing is achieved by performing a convolution with the range reference function, which in this case is simply the signal itself (fig. 2.12). Next, the range migration has to be fixed: As the radar moves, the distance to the target is changing and thus the center of our signal migrates. In reality this is performed with some rather complicated filters. In our simulation we simply shift the signal by requiring symmetry. Finally the azimuth compression is performed. As with range compression, a theoretic signal is computed and convolved line by line. An example of all 4 processing stages of a SAR image can be seen in figure 2.15. The focussing we simulated here in our toy example is not quite perfect, as we did not put the time and effort into this example that a real SAR systems engineer would have, to get the convolution kernels and the sampling rates right. The example shown also has very strong Doppler effect influence, which would not be seen in nature, but improves visualisation of the asymmetric nature of the signal. With a very keen eye, some artefacts can bee seen in the final azimuth compressed image, figure 2.15 d.

2.7 Remote Sensing of Sea Ice with SAR

SAR is typically operated on wavelengths between 1 - 50 cm. A central reason for the range of the wavelengths is that there are little to atmospheric absorption lines at these energies.

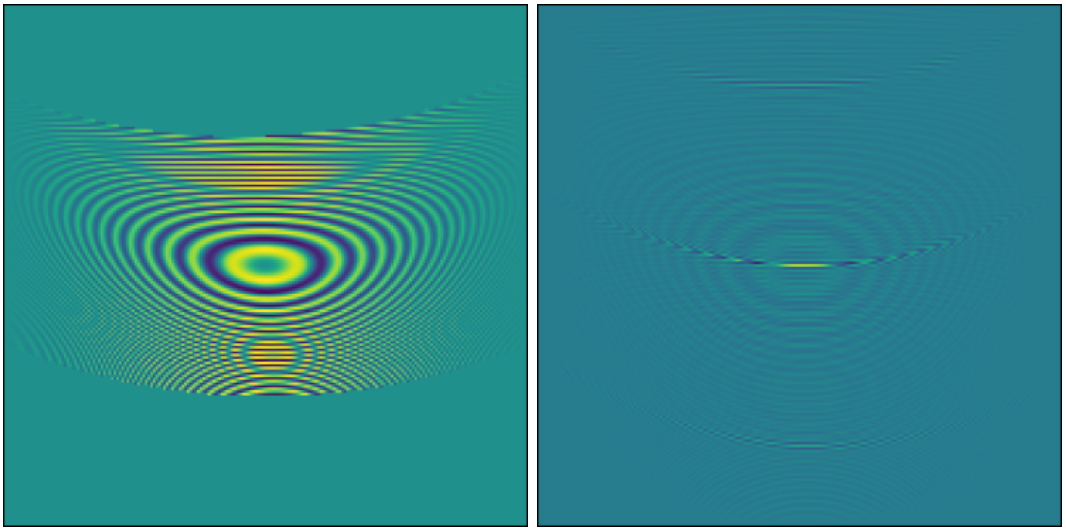


(a) column 10 (b) column 100



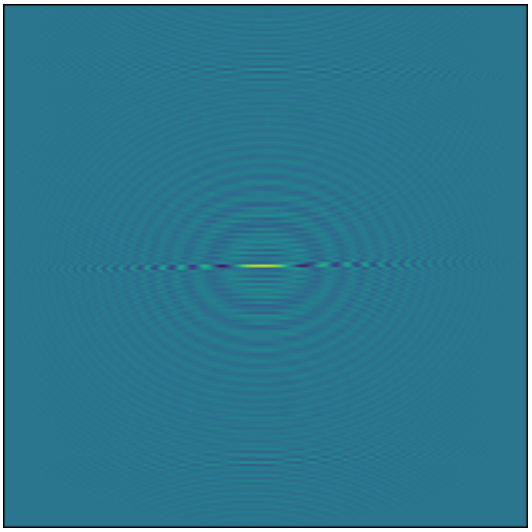
(c) raw SAR signal image

Figure 2.14: Formation of the raw SAR signal, made up of 200 columns. Two columns are displayed above and their location in the raw image is marked with a tick on top of its frame. Ambiguities from under-sampling can be seen most notably in the concentric signature below the real signal in the center.

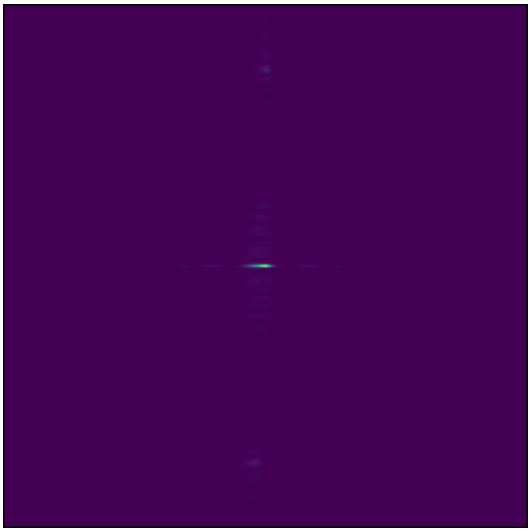


(a) raw signal

(b) range compression



(c) range cell migration undone



(d) azimuth compression

Figure 2.15: The various stages of SAR focussing, from the raw data to a fully focussed image. Focussing is achieved by convoution with the expected signal as computed from theory. For this example an unrealistically high velocity of the radar was chosen to accentuate the asymmetry due to the doppler shift. Additional structures, apart from the skewed concentric circles from the main target, are the result of under-sampling the return data.

Thus the earth's surface can be monitored without the absorption effects of clouds, rain or other atmospheric features. Combined with the fact that SAR is an active sensor and thus does not rely on sunlight (which is only available for half of the year in polar regions), makes this technology a prime candidate for monitoring the arctic ice sheets. As discussed in the introductory chapter, the electromagnetic field is sensitive to various features in the ice, resulting in diverse radar responses to the various states of sea ice cover. With a high resolution on the order of metres to tens of metres, SAR is a natural candidate for monitoring the sea ice masses of our planet. The resolution of freely available SAR products like the Sentinel-1 mission at 40 metres in extra-wide (EW) mode, which is used to monitor the Arctic Ocean, is principally high enough to extract valuable information about physical properties and processes in the ice that can further our understanding of dynamic processes and climate feedback mechanisms in the Arctic. A result from such observations, one hopes, would be the improvement of modelling efforts and thus, predictability of the Arctic and global climate in the future.

To this day, operational ice charting is largely carried out by human analysts that manually segment SAR scenes into various sea ice classes that are generally based on the ice concentration or stage of development. Whilst this data is very valuable for marine traffic, it does not have the level of detail one aspires to reach with autonomous classification. Instead, these products are rather coarse and do not fully use the sensor's high resolution.

An automated solution would not suffer from the time constraints of operational ice charting and is in principle able to make full use of the detail of the SAR measurement. Research towards such a retrieval method has been ongoing since the first satellite SAR missions. Reliable observations at the native resolution of the sensor have the opportunity to not only give more detailed information about the evolution of the sea ice cover but also to inform about sea ice dynamics processes such as lead and ice ridge formation, floe size distribution changes, and more. As the properties of a classifier are derivative of the training data, naively learning from operational charts brings severe limitations.

The retrieval of sea ice properties from SAR is manifold and layered (as are most unsolved problems in characterising large non-linear systems). At the core of the complications for sea ice property retrieval from SAR lie two concrete challenges:

- **The inverse problem:** The SAR backscatter depends on many sea ice properties and their distributions in the radar footprint/pixel. The inversion problem from one or two or even four channels (effectively three because $S^{HV} = S^{HH}$) to bulk sea ice properties is thus significantly underdetermined. Retrieving measurable quantities (other than the radar response itself) is therefore largely impossible. Part of the complication herein lies in the fact that modelling efforts have not yet yielded sensible bulk electromagnetic properties that could be inverted. With the highly nonlinear nature of the underlying interactions, there is no guarantee that such homogenised parameters will ever be found. This underdeterminedness of the inverse problem mapping radar backscatter to ice properties is the first major complication in the field.

- **The ground truth problem:** The polar regions are very hostile to the human organism, leading to only few and arduous campaigns that can measure the ice properties in situ. The additional complication that arises for sea ice in contrast to remote terrestrial regions, is that sea ice observations have only very limited spatiotemporal validity: The sea ice is constantly transforming in both space and time dimensions. Therefore, the already sparse data collected on the ground about the properties of the sea ice can only be related to satellite acquisitions from a very specific time. The available ground truth in the domain is therefore severely limited.

As the local backscatter values is not sufficient to retrieve ice properties, researchers try to leverage a priori knowledge, contextual data and generalisation to sea ice types to make meaningful retrievals a reality. As in most things, the retrievable ice classes depend on scale. In reality, there will mostly be a mixture of different ice in a single satellite SAR pixel. Because small scatterers can dominate entire pixels, unmixing the contributions from different classes is not feasible. Instead, one needs to settle for bulk classes at the scale of the SAR product. An increasingly popular approach in the sea ice community is to automate sea ice retrieval using machine learning approaches. The ideas here are natural insofar that neural network-based algorithms can relate the contextual data in the vicinity to overcome the first challenge of local information sparsity. At the core of these data-driven approaches stands the idea that the algorithm may use the large number of scenes captured without having to model the interaction of sea ice and the sensor explicitly. As is suggested by the term data-driven, what these approaches need to function proficiently, is high-quality data. Whilst we do have vast quantities of high-quality SAR data available, high-quality labels/ground measurements are severely sparse and the ground truth problem becomes increasingly severe when using data-driven approaches.

2.8 Neural Networks and Deep Learning

To solve the inversion problem of sea ice remote sensing with SAR, additional contextual data needs to be taken into account. However, it is not at all intuitive how spatial context can inform about the properties of an individual pixel. The human brain, for example, is very capable at using additional spatial context in an image to discern more detailed information about the nature of depictions. Neural network algorithms are data driven models that can learn similar non-linear relations between contextual image data and depicted objects. As the name suggests they are inspired by the organic counterparts. As they have been shown to perform well on image data, they are natural candidates to help overcome the inversion problem of SAR. Machine learning based algorithms have become increasingly involved in our day-to-day lives, seeing widespread use in commercial applications and in science. The number of machine learning publications has more than tripled from 2015 to

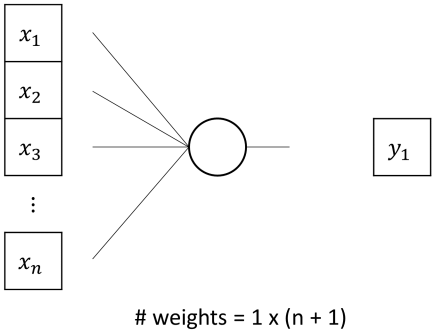


Figure 2.16: A graphic representation of a neuron mapping $N(x_i) = A(\sum a_i x_i + b)$, with $i = \{1, .., n\}$. Here A is some nonlinear function called the activation function. The number of trainable parameters/weights is thus $n + 1$ (a_i, b).

2020, as documented in Zhang et al., 2022. This section will briefly introduce the concept of neural networks and their application to image-like data. Later chapters in the thesis will build on the content presented here, mostly looking at higher level connections between inputs and outputs, rather than the individual functions mapping them. To be able to make sense of model behaviour, it is vital to understand the underlying mathematics. Luckily, they are not overly complicated.

2.8.1 Origins and Building Blocks

The inspiration for neural networks lies, as is suggested by the name, in the approximation of the workings of the brain. Although a complete understanding of how the brain learns and operates is yet to be achieved, we do know that it is made up of individual neurons capable of transmitting electrical signals to other connected neurons. A simple, linear mathematical formulation of synthesising a single outgoing signal from a range of incoming signals is a multivariate linear function

$$F(x_i) = \sum a_i x_i + b; \quad i = \{1, .., n\}, \tag{2.69}$$

where the value y, depending on a vector of incoming values x_i is given by a linear function with weights a and bias b. A typical graphic representation of such a network node is shown in figure 2.16. A neural network consisting of exactly one of these nodes

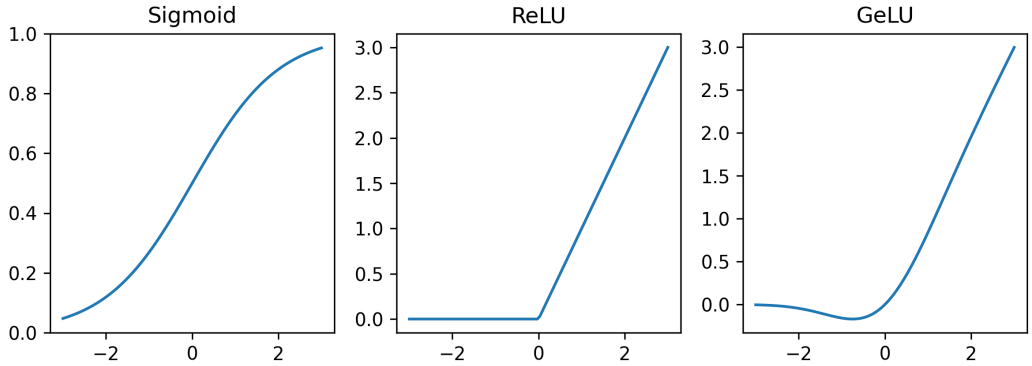


Figure 2.17: Some typical activation functions used to introduce non-linearity to neural networks.

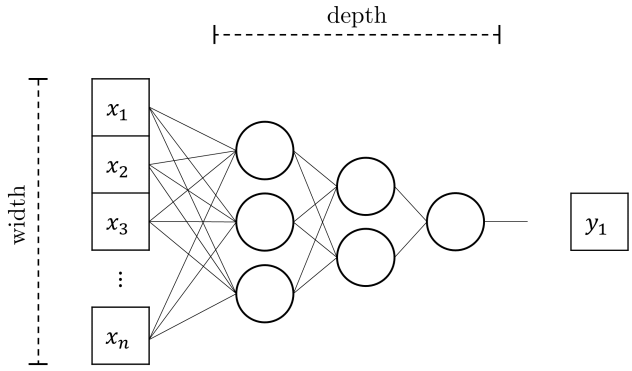
is equivalent to linear approximation. However, to achieve a more general approximator, multiple of these linear transformations can be chained together. Chains of linear operations are still linear and thus to be able to model more complex relationships, nonlinear operations need to be introduced. This is typically done by using an activation function A after the linear computation F, so that an entire neuron $N(x_i)$ is given by

$$N(x_i) = A(F(x_i)). \tag{2.70}$$

Typical activation functions are, for example, sigmoid, ReLU or GeLU, which can be seen in figure 2.17. Commonly a network using such neurons as primary building blocks are made up of layers of multiple neurons. The layers are fully connected, such that every neuron in a layer receives the output from every neuron in a previous layer and passes it to every output in the subsequent layer. If chaining together multiple such layers to construct a network of greater depth (number of layers) and width (number of neurons per layer), it has been shown that one obtains a universal approximator for well-behaved functions Cybenko, 1989. An example of a visual representation of a dense neural network with two hidden layers is shown in figure 2.18.

2.8.2 Image Data

When looking at image data (a matrix of dimensions height × width × channels), the simplest way would be to flatten the image to a vector and then continue with multiple fully connected layers. However, as images can be quite large, the number of parameters and thus computational cost rises rapidly. Therefore one is looking to exploit some fundamental properties of the data to make the computation more efficient. A simple yet powerful approach is to assume that the analysis of an image should have a property relating to translational invariance. For example, detecting an edge in one part of the image should work the same as detecting it in a different part of the image. Furthermore, pixels close to one another are probably related more directly than pixels far away from one another. These ideas are implemented by using sliding windows in the analysis, which compute local



weights = 3 x (n + 1) + 2 x (3 + 1) + 1 x (2 + 1) = 3n + 3 + 8 + 3 = 3n + 14

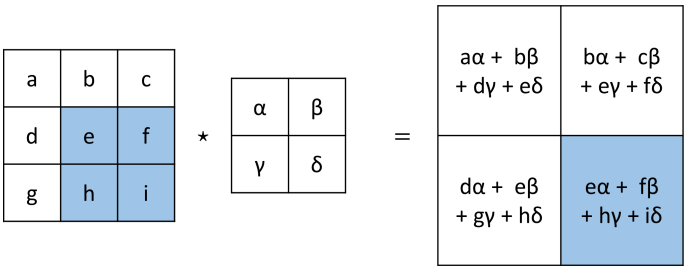
Figure 2.18: A graphic representation of a multi-layered dense network. The nodes are understood as in figure 2.16.

image properties and are translationally invariant. This drastically reduces the number of weights in the network. These sliding windows can also be interpreted as convolutions over a discrete space and are thus called convolutional layers. A two dimensional convolutional layer (Conv2D)is made up of a sliding window of weights, where each pixel in each channel of the window is assigned a weight and the output of the window is computed as the sum of all weights multiplied by the value in the input array (see fig. 2.19 for a visualisation). An additional constant (also a learnable weight) is sometimes added to the result. This constant is commonly called a bias. Typically a convolution function is also followed by an activation function, for the same reason of introducing non-linearity as in dense networks described earlier.

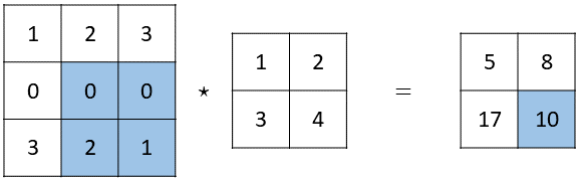
Thus, each pixel is primarily related to its neighbouring pixels and connections between pixels far from each other are explored only further down the network. For classifiers, where the output is a scalar or a vector, it is common practice to reduce the height and width dimensions of the resulting matrices, by using strides greater than one or pooling the results in between convolutional layers. A graphic explanation of a max pooling layer is shown in figure 2.20. That way, the layers' outputs are becoming narrower and deeper. Eventually, the outputs are flattened (turned into a vector) and some fully connected layers are added onto the end, to achieve the desired output shape.

An intuitive way of thinking about the levels of abstraction in a convolutional neural network is to view the detected features increasing in complexity and size with the depth of the network. Simple textures and lines are detected in the initial convolutions (sometimes called the stem). Later in the network, these are combined to form larger features, like floe edges or water lines. In the final stages of a classifier the network will then detect even higher-scale composite features, such as leads, floes or other larger features of the images.

In a network set-up where the output is also a matrix, a general approach is to first make the network narrower and deeper (as is also done for classifiers that output only a number or a vector). Then the inverse operations are done to achieve the desired output



(a) 2x2 Conv2D



(b) Example

Figure 2.19: A graphic representation of a discrete convolution operation commonly used in digital image processing and computer vision tasks. The example here uses a filter of size two, a stride of one and no additive constant. The input array values are multiplied with the weights of the filter and then the summed over to obtain the output array. The filter strides are variable and can be used to achieve a reduction of array dimensions.

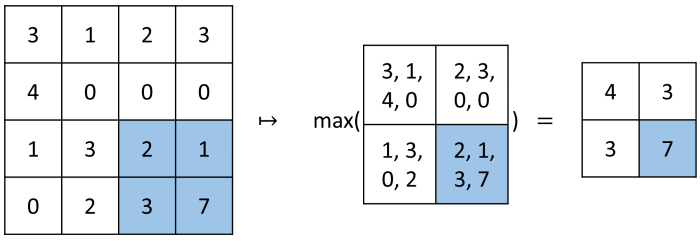
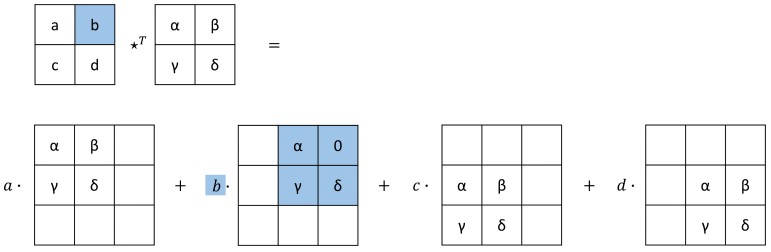
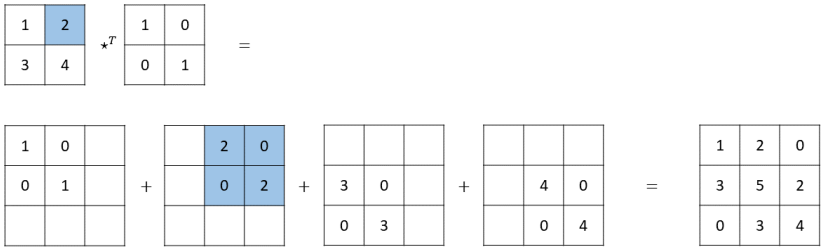


Figure 2.20: A graphic representation of a 2x2 max pooling operation with stride 2, as is used commonly in digital image processing and computer vision tasks. Only the maximum value in the filter window is kept. As with convolutional layers, the stride of the filter can be varied for different applications.



(a) 2x2 TConv2D



(b) Example

Figure 2.21: A graphic representation of a 2x2 discrete transpose convolution operation. The stride is 1. Larger strides lead to larger output dimensions.

dimensions. Having a lower parameter count in the middle of the network forces higher-level parametrisations and abstraction of the input data. As with convolutional layers, we can only decrease the width of the network, one needs an additional yet related operation: the transposed convolution. A transpose convolution is learning a filter that is multiplied with the input channels and then summed on the output matrix. If the stride of the filter is greater than one, it will result in a wider output. A visual explanation is shown in figure 2.21.

2.8.3 Learning

We have so far discussed the building blocks neural networks are commonly made up of but have not given detail on the learning aspect of a network. The overarching term to describe the configuration of blocks used to map input to output is termed the network architecture. Next to the architecture, we also need some way to describe how close the network is to its desired outcome. This is handled by the loss function L, which maps the output data (and possible auxiliary data, like ground truth) to a scalar. In case of supervised learning, where we are trying to approximate some ground truth, the loss function could simply be the squared difference between the prediction and the truth. To train the network then means to reduce the value of the loss function on the training data.

Network training consists of two steps, the forward and the backward propagation. During the forward propagation (or inference), the network is evaluated on given data. The learning aspect of the training then happens in the backwards propagation. This second step consists of calculating the gradients of the loss function in regard to every parameter in

the network and using these to adjust the parameters. In its simplest form, this is gradient descent:

Given a dataset of inputs X and desired outputs Y, a training loop of network N with learnable parameters $\{a_i\}$ and loss function $L(N(X), Y)$ can be described as follows: In the forward propagation step using a datapoint $(x, y) \in (X, Y)$

$$N : x \mapsto N(x), \tag{2.71}$$

In the backwards propagation step, the network weights are updated with gradient descend, such that

$$a_i \mapsto a_i - \alpha \cdot \frac{\partial L(N(x), y)}{\partial a_i}, \tag{2.72}$$

with α as the learning rate, parameterising the strength of the update.

How reliably the network converges to the best solution (minimal loss) is often decided not only by the network architecture but also by the preparation of the data. Additionally, for more complex tasks, the formulation of the loss function and interplay between different parts of the network is imperative. How well a network trains is determined by the landscape of the loss function. By landscape, we mean the value of the loss function on the entire dataset for every parameter in the network. This landscape is naturally very high dimensional, as the dimension is equal to the number of parameters in the network. Typically, this is on the order of tens of thousands to millions of parameters. A preferred landscape offers as few and shallow as possible local minima and has a deep and well-localised global minimum to converge to. Because of the high dimensionality of the space, this is reliant on intuition and heuristics and cannot be tested quantitatively with ease.

To improve the way the network moves through this landscape during training, some variations are made on the idea of gradient descent. A successful optimisation was achieved with the Adam algorithm by Kingma and Ba, 2014. The major features it incorporates are an internal momentum and learning rate decay. The momentum allows the optimizer not to be overly affected by smaller local features of the landscape, and the learning rate decay decreases the strength of adjustments later in the training run so that the minima are not being overshot.

One thing to remember is that the loss must be differentiable in regards to all trainable variables, so loss terms cannot involve non-differentiable operations like counting, histograms, thresholding etc.

2.8.4 The Relationship of Data and Trained Model

So far neural networks look like perfect candidates for retrieval models. Due to the universal approximation theorem, any relationships between inputs and outputs can in principle be mapped and advanced versions of stochastic gradient descent give a clear way to train the networks. The big caveat of these algorithms is that to find the desired mappings, a lot of data is required that needs to completely span the distributions of inputs and outputs. The networks are highly nonlinear models which have no directive on how to behave in between

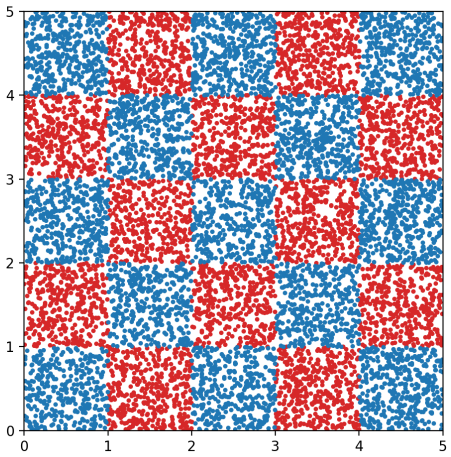


Figure 2.22: The training data used to train the demo network. Size=10000.

data points or even more crucially outside of them. Because of their non-linear nature, the behaviour outside of the training data domain is typically quite chaotic.

Input 2
Dense x 10
Dense x 10
Dense x 10
Dense x 10
Dense x 10
Dense x 1
Output 1

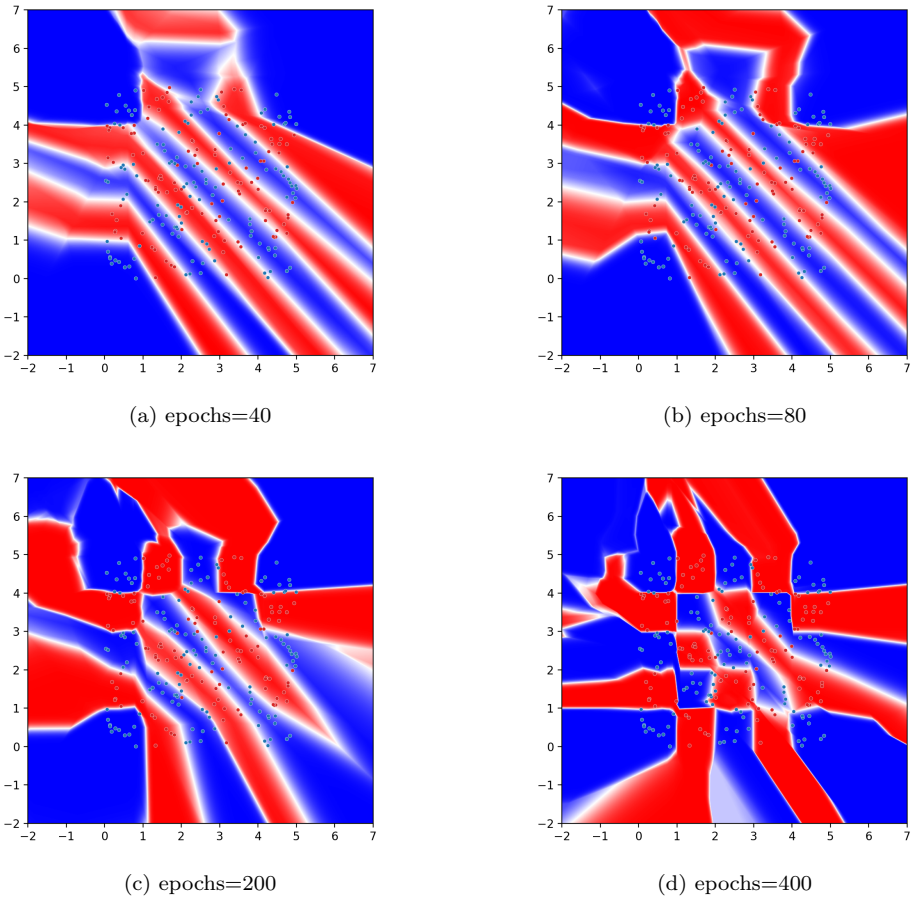
Table 2.1: The demonstration network. The Relu activation function is used except for the final layer, where sigmoid is applied.

To illustrate this behaviour, a simple dense network N with 5 hidden layers of 10 neurons each is trained to fit a five by five checker board pattern. The relu function is used as an activation until the final layer, which uses a sigmoid function. The loss is a cross-entropy loss, also known as log loss

$$L_{\text{CE}}(N(x), y) = -y \log(N(x)) - (1 - y) \log(1 - N(x)), \tag{2.73}$$

for an input x with label y. The training data is shown in figure 2.22 and consists of 10,000 samples.

In figure 2.23 the progress of training is shown. It is immediately obvious that the algorithm does not extrapolate well on data outside of the training data domain, showing rather chaotic behaviour outside the 5x5 square. This serves as a useful demonstration, that networks cannot be expected outside of the training data distribution. For the same reason, sea ice retrieval algorithms using neural networks should not be expected to work outside



(a) epochs=40

(b) epochs=80

(c) epochs=200

(d) epochs=400

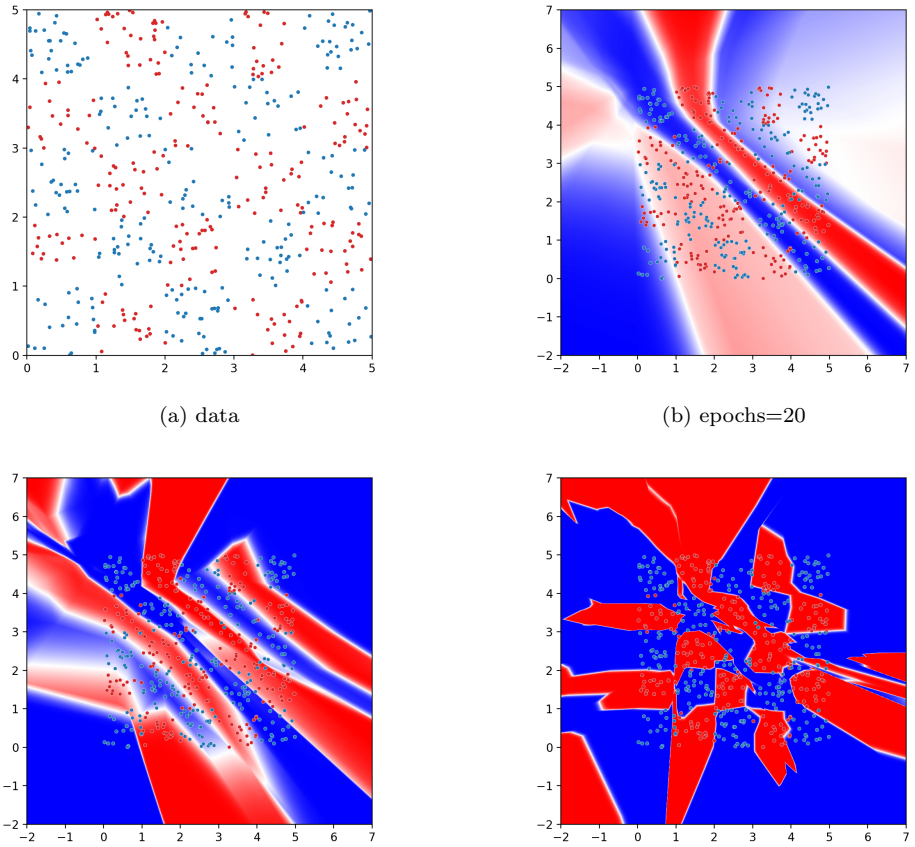
Figure 2.23: Training progress of the demo network after a given amount of epochs (number of times iterated over all training samples).

of conditions present in the training data set. As the possible ice and snow conditions span a large space, it is immediately obvious, that the sparse amount of ground truth in the polar regions have a particularly pronounced effect on retrieval methods using deep learning methods.

Above, the issues of out-of-domain classification were illustrated. Another common problem with limited data is overfitting. This describes the network fitting the training data so closely that it loses generality. To illustrate this with the above checker board example, we fit a network on only 500 data points, rather than 10000 and monitor the progress during training (figure 2.24). In contrast to the previous example in figure 2.23, the network refines the boundaries to a degree where they do not generalise as well to unseen examples, leading to jagged edges inside of the checkerboard pattern.

2.8.5 Regularisation

To avoid overfitting the training data, some techniques have been developed that ease this problem in a data-agnostic way. These techniques are commonly bundled under the term



(a) data

(b) epochs=20

(c) epochs=200

(d) epochs=6000

Figure 2.24: Training progress of the demo network on 500 samples after a given amount of epochs (number of times iterated over all training samples).

'regularisation'. One of the simplest and yet most effective regularisation techniques is called dropout. Dropout layers simply set a random subset of input features to zero, with the probability of being 'dropped out' given by a manually adjusted hyperparameter. Heuristically, this forces the network not to become overly reliant on certain features in the network, as they are not always available. The resulting diversification is useful to constrain the amount of overfitting. When the network is predicting samples after being fully trained, the dropout layers are typically switched off and dropout no longer occurs, so that the full wealth of learned features can be used for prediction.

2.9 Literature Overview

To set into context the work that is done in this thesis, we will briefly discuss existing published work, focussing on the deep learning contributions that were developed previously or in parallel with the work shown here.

The European Remote-Sensing Satellite ERS-1 was launched in 1991 with a SAR instrument on board, measuring at C-band. So also started the first efforts of automatic sea ice retrievals from SAR. Within the next year, some of the first research into autonomous classification of sea ice types was published in Kwok, Rignot, et al., 1992, using clustering of the normalized radar cross sections and look-up tables to distinguish classes for two scenes. The automation techniques moved to more sophisticated methods centred around extracting valuable features from the spatial neighbourhood. One of the main approaches used gray level co-occurrence matrices (GLCMs), developed by Haralick et al., 1973. They were first utilised in sea ice classification by Barber and LeDrew, 1991 and further spearheaded by Soh and Tsatsoulis, 1999. Clausi, 2002 continued developing the methodology of co-occurrence texture features, honing in on the effect of the quantisation of grey levels. In parallel, textures were also computed using Gabor filters, introduced by Jain and Farrokhnia, 1991. After some further investigation by Clausi and Ed Jernigan, 2000, they were also analysed for sea ice classification in Clausi, 2001 in comparison with GLCMs and Markov-Random-Fields (MRFs). Here the best results were achieved for GLCM and MRF-derived features. The authors could show that, in fact, these two features were independent of one another. This work was expanded upon in Clausi and Yue, 2004. Markov random fields were first introduced to the image segmentation worlds by Geman and Geman, 1984 and used in several papers for sea ice classification purposes. Some early works of note include unsupervised segmentation proposed by Deng and Clausi, 2005 and efforts towards operationalised ice charting by Maillard et al., 2005. Other texture feature-based approaches being considered in that time were autocorrelation features applied by Karvonen, Simila, et al., 2005 and wavelet-based textures by Yu et al., 2002. The first approaches using neural networks to classify sea ice were made by Hara et al., 1995, directly using local backscatter values as inputs. Later networks were then used on segmented data by Karvonen, 2004 and on features from GLCMs, autocorrelation or image region growing approaches as inputs by Bogdanov et

al., 2005. From this point onwards, machine learning developments grew with new processor and memory technologies. Whereas the studies presented so far were usually constrained to a handful of acquisitions, big data technologies were emerging into existence and fusion with additional instruments was explored further. Studies by Karvonen, 2014; Karvonen, 2017 used multi-layer perceptrons (MLP), also known as dense networks, to classify sea ice using both passive microwave and SAR radar using tens of scenes for training. Similar efforts were made again by Malmgren-Hansen et al., 2021, this time using convolutional neural networks (CNNs) and also by Radhakrishnan et al., 2021 using a semantic model in the Unet, that was developed by Ronneberger et al., 2015. Different GLCM features were also used for X-Band satellites in Ressel, Frost, et al., 2015 and appended with an investigation into a range of polarimetry features in Ressel, Singha, et al., 2016 and Aldenhoff et al., 2018. These methodologies were then applied to compare with freeboard measurements in Singha et al., 2018, measured with an airborne laser scanner. Recently, researchers have been moving away from manually selected texture features and towards automatic feature distillation with convolutional neural networks (CNNs). In general, there are two approaches here: to classify one pixel at a time or to segment the image patch-wise. The former has been explored by Wang, Scott, et al., 2017 for sea ice concentration, Karvonen, Rinne, et al., 2022 focussing on sea ice concentration in the Baltic Sea and Boulze et al., 2020 for sea ice classes from Sentinel-1. Lyu et al., 2022 investigated the use of CNNs for sea ice classification in the Davis straight from the Radarsat Constellation Mission using a norm-free variant of ResNet, which was developed in He et al., 2015. Jiang et al., 2022 used the original ResNet variant for Sea ice classification in the Beaufort Sea. Zhang et al., 2022 classified sea ice from Gaofen-3 SAR data with Mobilenetv3 developed byHoward et al., 2017. Khaleghian, Ullah, Kræmer, Hughes, et al., 2021 published work using the VGG16 classifier developed by Simonyan and Zisserman, 2015 and Ullah et al., 2021 did the same with the inclusion of the noise profile as additional input. Both of them work on the same Sentinel-1 dataset. The patchwise segmentation was investigated for ice and open water separation using variants of the Unet by Ren, Xu, et al., 2020; Ren, Li, et al., 2022, the aforementioned work by Radhakrishnan et al., 2021 combined this with passive microwave imagery and Wang and Li, 2021 used a stack of models for better performance. Gélis et al., 2021 also used a derivative Unet architecture for sea ice concentration mappings. In Murashkin and Frost, 2021, the authors used a Unet++, developed in Zhou, Siddiquee, et al., 2018; Zhou, Siddiquee, et al., 2019, to perform sea ice classification from Sentinel-1 images.

From the short recap above, it is clear that deep learning approaches have gained increasing traction in recent years. With some of these works, like Park, Korosov, Babiker, Won, et al., 2020; Malmgren-Hansen et al., 2021; Wang and Li, 2021; Gélis et al., 2021; Karvonen, Rinne, et al., 2022 using over a hundred scenes in their studies, it seems that the field has arrived in the age of big data. These authors use ice charts provided by various national ice charting agencies to facilitate this volume of labels. In fact, except for some labelling using passive microwave radiometers, all of the works above lean on manual human

interpretation of ice charts to generate training data. As mentioned previously, these ice charts do not have the level of detail desired for advanced observations valuable for modelling and a deeper understanding of the change in ice type distribution over the past decade. For example, all ice far enough from the ice edge is usually considered old ice or of the highest stage of development. Consequentially no meaningful separation exists in the heart of the ice pack. This includes no labelling of leads or heavily ridged areas. This property, unfortunately, translates to the resulting classifiers. As a result, we are still quite a ways away from the optimal classifier that robustly gives sea ice classifications over multiple seasons at high resolution. In fact, the robustness of the algorithms to changing environmental conditions is never rigorously tested and quantified in the existing literature.

Chapter 3

Robust Deep Learning of Ice Types - X-Band SAR Observations around the MOSAiC floe.

3.1 Overview

Whilst neural network based approaches to sea ice retrieval have shown promise for retrieving sea ice information from SAR imagery in the past, the possibilities of exploring long distance connections in the SAR image, by including larger contextual windows has not yet been tested. Also, the robustness of neural network classifiers over a larger time span has never been studied. To do so, especially at high resolution, one needs to track sea ice across time and space. To address these two challenges, a sea ice classification algorithm based on convolutional neural networks is developed to classify a time series of SAR scenes acquired during the Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC). The footprints of the scenes are shown in figure 3.1. The aim is to establish strategies for extrapolating manually labelled ice regions across a large span of possible ice conditions while monitoring the algorithm's robustness. Finally, the developed model is compared and contrasted with the well-established image classification architecture VGG16, developed by Simonyan and Zisserman, 2015. When orienting oneself regarding the two big problems of sea ice remote sensing outlined in the previous section 2.7, this chapter concerns itself with using deep learning methods to solve the inversion problem by maximising the use of spatial contextual image data to perform accurate classification. Manually labelled ice classes are used as training data. Therefore, the ground truth problem is not further considered in this chapter. The body of this work has been published in Kortum, Singha, and Spreen, 2022.

3.2 Introduction

During freeze-up and melt periods, sea ice classification becomes increasingly difficult. The main challenges are wet snow lowering radar penetration depth, snow metamorphism and increased ice dynamics as surmised acutely by Mahmud, Nandan, et al., 2020. A general downward trend in radar response can be observed after freeze-up, especially for level ice (fig. 3.2), showing that there exists a correlation with backscatter and the seasonality of sea ice. Similarly, there is also an effect of the incidence angle (fig. 5.1 in a later chapter addresses this in more detail). Although the difficulty of classifying due to high incidence angles is derivative of the acquisition method, not the sea ice itself, autonomous classification will have to perform robustly across the range of these effects.

Due to the decreased penetration depth in warmer seasons, the SAR texture features, essential to most autonomous classification, become decreasingly reliable as the backscatter signal becomes more uniform across the different ice types. A possible approach to tackling this is including more contextual image data. For example, with larger sliding windows around the ice to be classified. Then, using automated feature extraction and classification with a convolutional neural network (CNN) is especially helpful because the neural network can learn to relate all the information in the window to only the centre pixel one is trying to classify. Thus, it handles large contextual windows better than texture feature-based

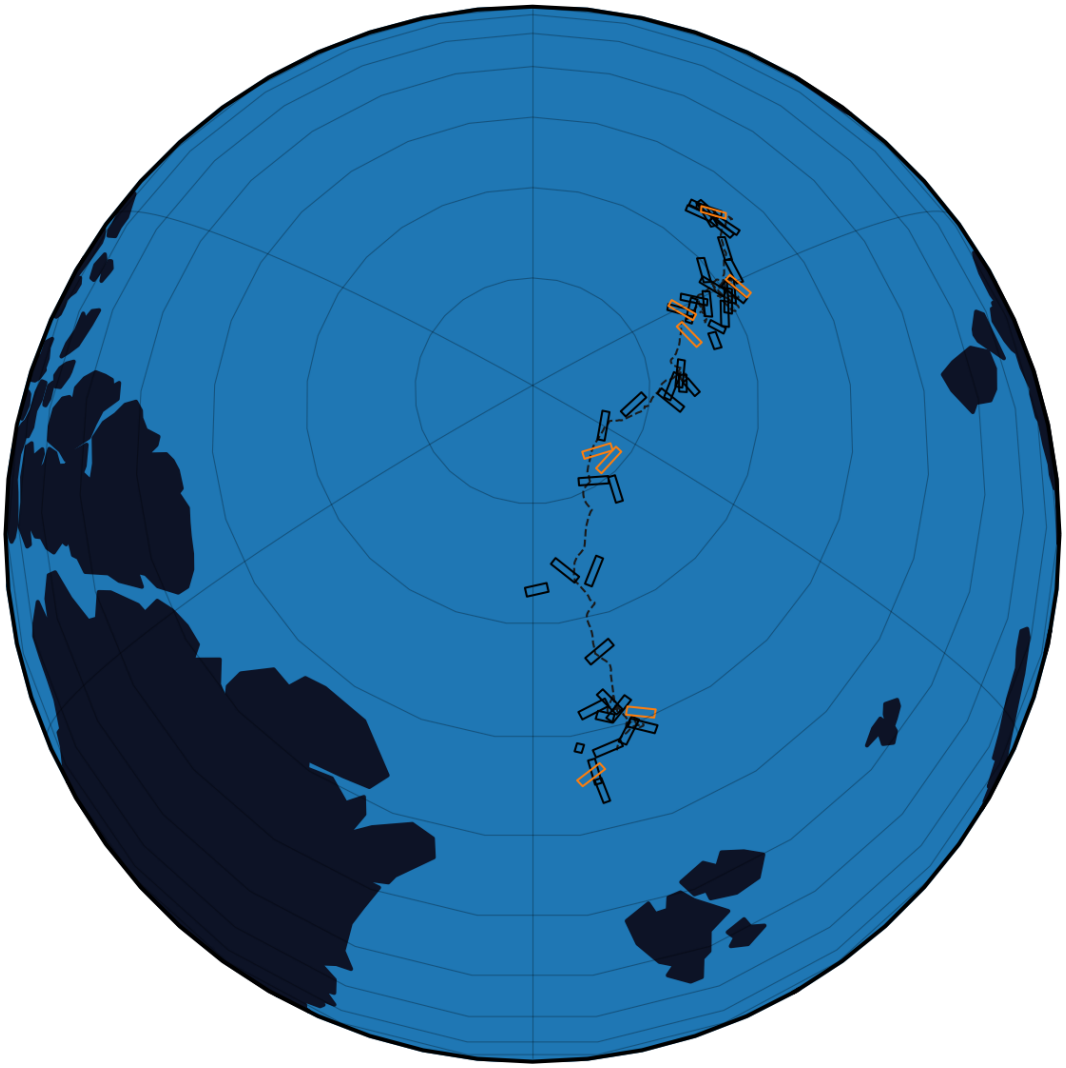


Figure 3.1: All TerraSAR-X scenes that were manually labelled for this study, along the MOSAiC drift track between October and May. Test set scenes are picked out in orange.

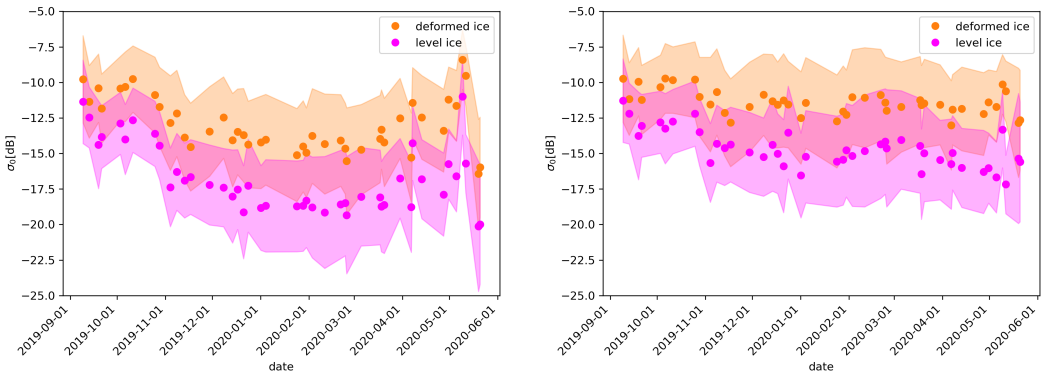


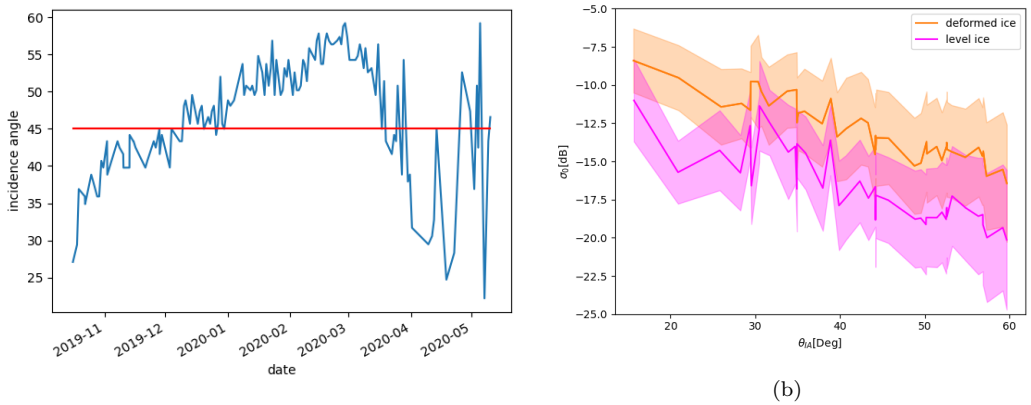
Figure 3.2: Charts showing the mean sigma nought calibrated backscatter for the HH band, for level ice (LI) and deformed ice (DI) from October through May in all labelled scenes. The data is extracted from 52 manually labelled TerraSAR-X scenes. The right hand side shows the chart after correcting for incidence angle dependence using values from Guo et al., 2022. Note that, because this is only data that was selected manually for training, these values are affected by a selection bias: Only areas are labelled, where the ice type could be determined with high confidence. In reality the contrast between the two classes is most probably smaller.

classification, where the contribution of a pixel to the output is independent of location.

The analysis in this chapter is based on a large time series of TerraSAR-X acquisitions at X-Band. C- and L-Band SAR have historically been preferred for sea ice classification. Not only is there greater coverage, with large satellite missions such as Sentinel-1 and Radarsat, longer wavelengths also offer bigger penetration depths, as detailed in Johansson, Brekke, et al., 2018. This makes it easier to discriminate between ice classes from backscatter and texture features alone at the longer wavelengths. The classification from X-Band SAR consequently has more to gain from including additional contextual image data.

A core idea of the classification approach discussed in this chapter is using contextual windows at various zoom levels, allowing the classifier to exploit long range dependencies and larger image features. A dense conditional random field (DCRF) is used in a post-processing step. This concept of adding spatially aware boundary refinement has been implemented in image segmentation as early as 2014 by Chen et al., 2014. Random fields have been used successfully in the past for automatic ice charting by Ochilov and Clausi, 2012 and have shown promise as a post-processing step with sparse labels in image processing in Hua et al., 2021. The combination of CNNs and conditional random fields has also recently been shown to be successful for ice concentration mapping by Cooke and Scott, 2019.

To truly obtain a measure of the robustness of a classifier, one has to show continuity in the classification of overlapping and near coincident SAR scenes, which demonstrate that a patch of pack ice is predicted to be of the same class across different scenes. This is feasible for large-resolution scenes, but even here, most research (e.g. Park, Korosov, Babiker, Won, et al., 2020) has focused only on a few such overlapping scenes, and robustness across a



(a) (b)

Figure 3.3: Average incidence angle the scenes were acquired at (a) and average brightness of different classes dependent on incidence angle (b). The red horizontal line at $45°$ in (a) indicates the limit of the full performance range of the satellite sensor. Note that as there was also a correlation between incidence angle and time, direct estimation of the incidence angle dependent slopes is not possible from plot (b). These acquisitions are up to 8 Months apart.

greater range of conditions is still a challenge. At high resolution, coverage is small. Thus, imaging a small pack ice region for an extended time is even more complex. Not only does the drift of the ice have to be tracked, it also needs to be predicted due to the delay of the ordering and the capturing of a scene. Over the course of MOSAiC, this task has been tackled by various spaceborne SAR sensors. Such a dataset captured by the TerraSAR-X satellite in Dual-pol StripMap (HH, VV) mode is used in this investigation. It presents the opportunity to validate the robustness of a classifier over an extended time period and a large number of scenes.

3.3 Data

Test, Training and Robustness Analysis Data Sets. This analysis' training and test data sets comprise 44 and 8 TerraSAR-X Dual-pol StripMap scenes, respectively. The scenes were acquired between October 2019 and May 2020 over or near the Polarstern vessel during its drift with the Arctic pack ice. The two channels acquired are the HH and VV polarisations, respectively. The images have a row (azimuth) and column (range) spacing of 3.5 m, and are typically around 16000 pixels by 4000 pixels in size. This corresponds to an area of 56 km by 14 km. The test scenes contain one randomly chosen scene from every month of the drift. The data points extracted from the 44 training scenes are split into two disjoint training and validation sets, with a size ratio of 9:1. The classifier is trained on the training set, whilst performance on the validation data set is used to stop the training in time to prevent overfitting. The data used for robustness analysis is made up of 162 scenes. All of these entirely contain the immediate area around Polarstern (a 3 km by 3

	training data	test data	robustness eval. data
# scenes	44	8	162
feature	labelled	labelled	unlabelled, contains 3x3km study area

Table 3.1: Overview over the datasets used in this analysis.

km square). To keep similar time-spacing between scenes, no more than one scene was used daily. Henceforth, this data is referred to as the robustness evaluation dataset. An overview over the datsets is shown in table 3.1

Due to the drift with the Arctic pack ice, the RV Polarstern entered very high latitudes in the beginning of 2020. Figure 3.3a shows that this time, the SAR images were consistently taken outside of the full performance range (which is between 20° and 45° for TerraSAR-X StripMap images). The SAR measurements for such high incidence angles have significantly lower signal-to-noise ratios, making it increasingly difficult to differentiate ice types. Weather conditions varied throughout the mission, including events such as storms and warming periods. Their effects concerning this study are constrained to the contribution to increased ice dynamics, as the radar signal is not susceptible to atmospheric conditions at X-band.

label	meaning	colour
OW	open water	blue
TI	thin ice	green
LI	level ice	magenta
DI	deformed ice	yellow
HDI	heavily deformed ice	red

Table 3.2: Table showing class definitions and labels.

Labels. As mentioned in the opening paragraph of this chapter, this analysis relies on manually labelled data. Due to the time-intensive nature of the labelling process, there is naturally less data than when using ice charts. The ground truth problem thus manifests itself in this labelling approach in the shortcomings of the manual labels. Apart from not being validated, they are also quite coarse - not at the sensor's resolution - and, therefore, cannot capture small individual floes or deformation features. Also, the labels cannot fully span the diverse space of possible ice and acquisition conditions during the eight months of the dataset - especially as the freeze-up and early pre-melt seasons are included in this time period. Labelling was done based on the X-band SAR data for five classes. These were chosen to align with qualitative in situ observations made by members of the MOSAiC expedition. First, suitable classes and respective areas to be labelled were found, using the in-situ observations as a guide. Then, the established logic was manually extrapolated to the rest of the areas using the SAR data only. The five classes are shown in table 3.2. The colour coding used can be found there as well. In data-driven algorithms, the resulting classifier and the training data are intrinsically linked. The challenges of the manual labels need to be considered when designing the model's architecture. These mainly consist of inaccuracies,

selection bias towards larger and easy-to-identify regions away from class edges. An example of a scene with labels is shown in figure 3.4.

3.4 Classifier Design

In this section an advanced architecture is proposed that leverages additional contextual data and introduces some heuristics to increase the classifiers robustness and accuracy over a traditional VGG16 style architecture. Before the individual components of the classifier is described in detail, an overview over the deep learning approach is given. The core of the proposed classification scheme is a convolutional neural network (CNN). To increase robustness, a discriminator and a dense conditional random field (DCRF) are used for further processing. The algorithm assigns one of five classes (tab. 3.2) to a given 5x5 pixel patch of SAR imagery.

Figure 3.5 depicts the classification pipeline for the algorithm used. After pre-processing, features of varying scope and resolution (zoom levels) are processed by the network alongside each 5x5 image slice that is to be classified (tab. 3.3). A CNN is fed these features and makes an initial prediction for that patch (see figure 3.6 for details). The predictions are then checked by a second discriminating network (also trained with the same data as the classifier) that removes some labels deemed to be misclassifications. Finally, a dense conditional random field smooths over the labels by relating the spatial context of the labelled data and the underlying image. This also fills the missing values left by the discrimination step.

3.4.1 Pre-Processing

In the initial step of data pre-processing, the original dual polarised SAR scene is calibrated to the slant range (β_0) and a false colour composition of the data is constructed. The composition consists of four channels: HH, VV, HH-VV, HH/VV. The difference and ratio are common for manual ice charting and visualisation of SAR scenes, as they promote contrast across ice types and open water. Additionally, they have been shown to be useful for classification in the past by Geldsetzer and Yackel, 2009. The raw backscatter channels HH and VV are rescaled with a *tanh* function to achieve good contrast across the dynamic range of the image. This step also normalizes the input data between 0 and 1, which aids the numerical stability of the network. The composite features HH-VV and HH/VV are also scaled with a *tanh* function and an additional offset. The exact scaling parameters are manually selected to give strong contrast. Whilst the network is in principle able to learn the composite features, feeding them directly alleviates some of the workload of the network and gave slightly improved results in preliminary testing.

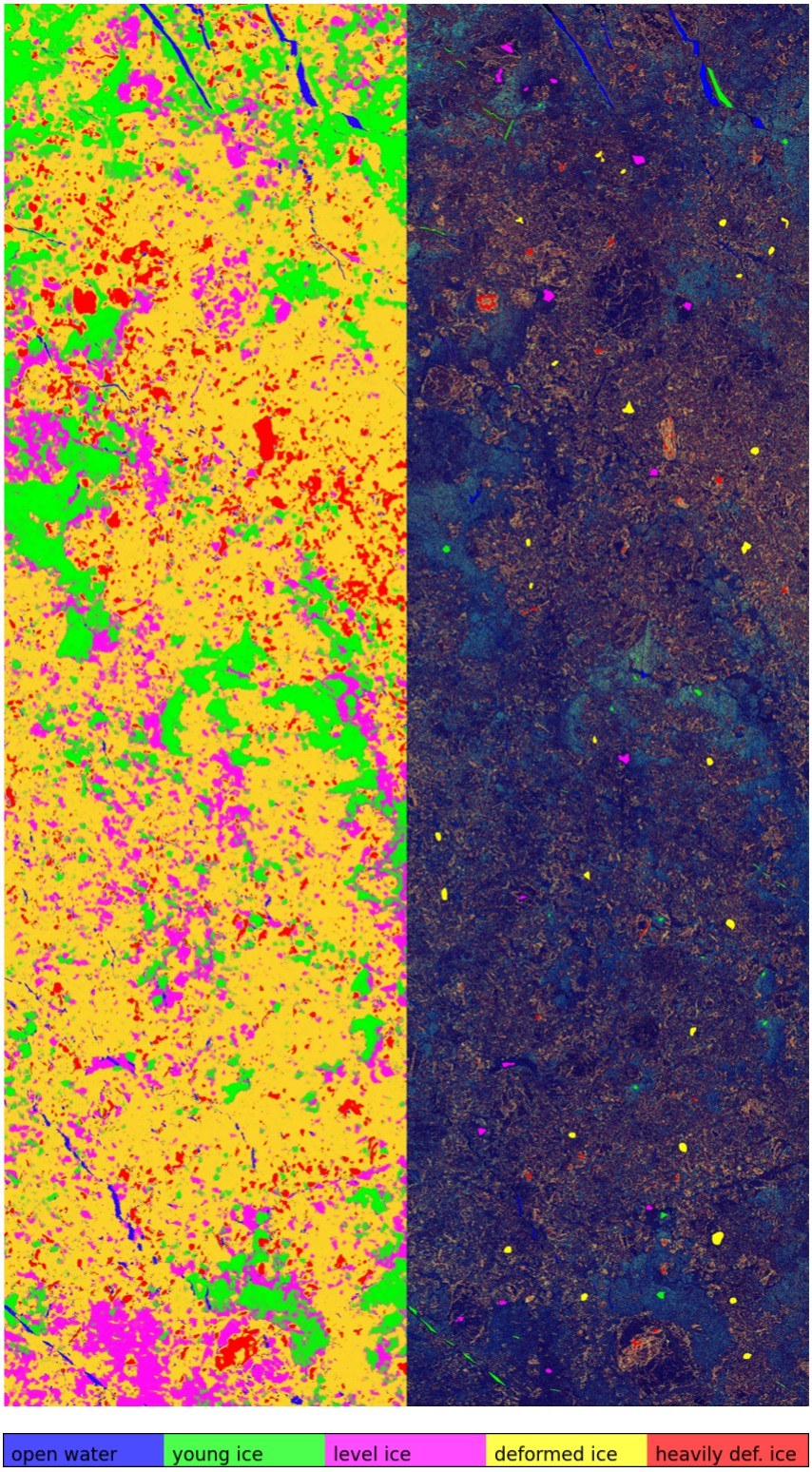


Figure 3.4: A TerraSAR-X Stripmap scene acquired during October 2019. A false colour composition using (VV, HH-VV, HH/VV) in the RGB channels is shown on the right, overlayed with manual annotations. On the left the classification map is shown.

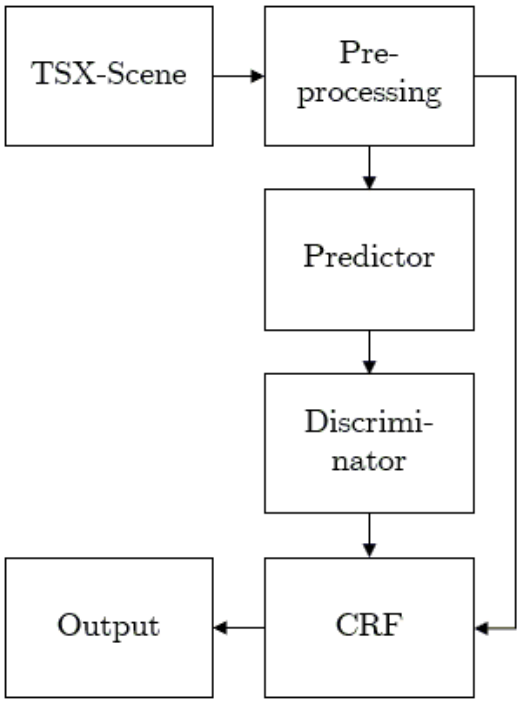


Figure 3.5: Flowchart showing the pipeline for the proposed ice classification algorithm.

3.4.2 Convolutional Neural Network

Architecture. The core of the classification approach is a Convolutional Neural Network (CNN) (see figure 3.6). It predicts one of five classes for each 5 x 5 pixel patch of the SAR scene. These patches ('local features') are appended by additional information of the surroundings. The first of these additional features is a 16 x 16 pixel patch ('superlocal feature') of the surrounding area that is taken from the SAR scene and downscaled by a factor of five. Thus, moving to the right one 5 x 5 patch in the original image moves one pixel to the right in the rescaled product the superlocal patch is taken from. This patch gives insight into the surrounding area, allowing the algorithm to take advantage of surface features nearby, such as ridges or leads, to gain spatial context. For example, the CNN might learn that heavily deformed ice is more likely to occur with well-defined edges in the surrounding area, like the edge of a multi-year ice floe. The patch sizes of 5 x 5 and 16 x 16 were established empirically. In general, both are compromises of resolution and accuracy: The larger the windows will get, the better the accuracy will become, as there is more information in the image to use. However, it becomes more difficult for the classifier to relate all this information to only the data in the centre of the patch. This leads to a lack of effective resolution in the classified product. To give a more complete picture, the entire scene (or the largest possible near quadratic slice of it) is additionally resized to 64x64 pixels and input to the model ('global feature'). The StripMap data used here is captured in rectangular strips, typically around four times longer than wide. In such a case, the scene is split into four near quadratic slices along the azimuth axis (the long axis). The global input

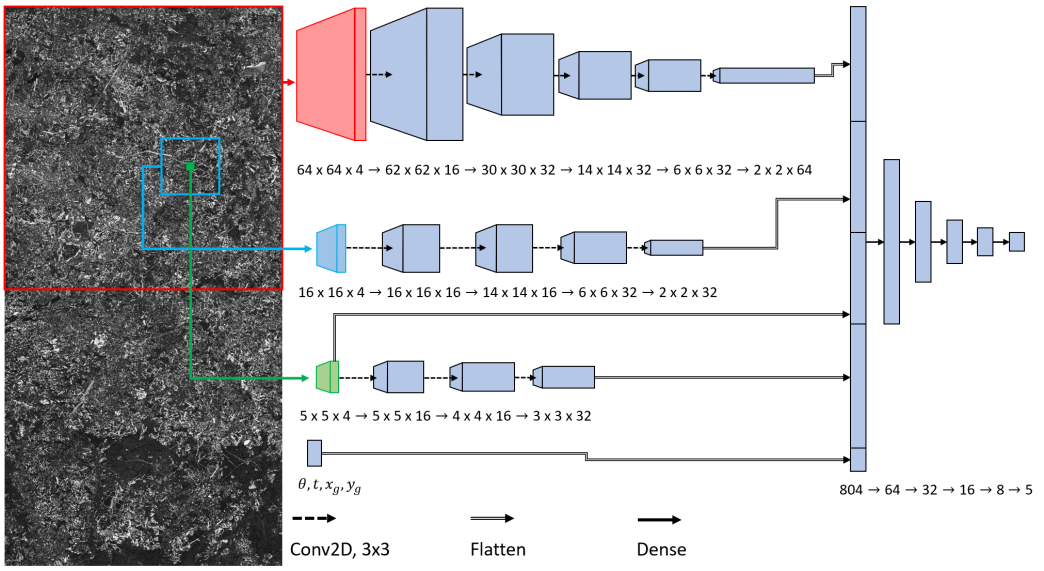
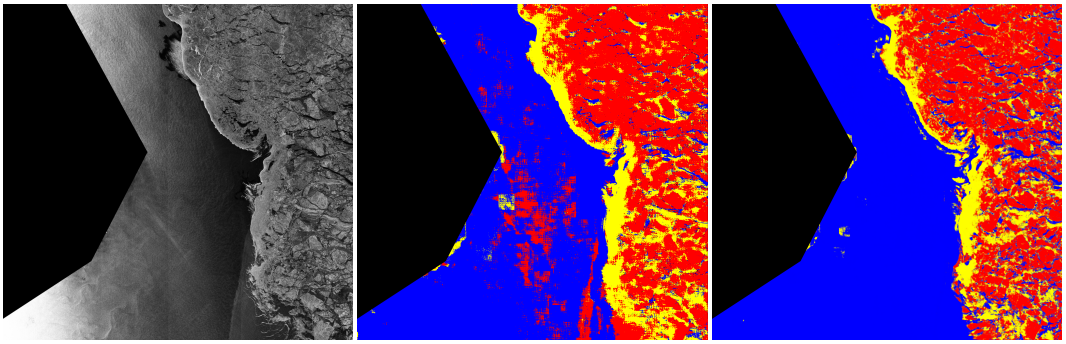


Figure 3.6: Illustration of the CNN architecture used in the proposed classifier. Wherever the spatial dimensions in the convolutional bloacks are downsampled (decrease by more than a factor of .5), a stride of 2 was used. Not included in the Image are the batch normalisation layers after the first convolutional layers for each input and the dropout layers used for regularisation during training. The parameters x_g and y_g denote the coordinates of the location of the local patch in the global patch, θ is the incidence angle and t the acquisition time. Parameter count = 120421.

feature allows some insight into large-scale features, such as the general brightness of the scene, interfaces between ice masses or (not important for this data set) the ice-water edge. An example using Sentinel-1 SAR imagery can be seen in figure 3.7. The global features can then be related to the high-resolution features and helped classify scenes with locally very high backscatter or low radar response (e.g. high incidence angle). As the entire range domain of the scene is parsed here, ensuring that the region to be classified lies in the centre of the image is no longer possible. Consequently, a fourth input ('extra feature') consisting of four parameters is provided, containing the region's position to be classified in the larger 64x64 input. It also contains the incidence angle of the patch and the time at which the product was acquired. A summary of these input features is included in table 3.3.

Data Augmentation. The training data set consists of 44 scenes. The scenes are often split into four near quadratic slices for the global feature. Thus, the number of different inputs for that feature is only $\approx 44 \times 4 = 176$. Herein lies a substantial risk that the algorithm overfits to the training data. It might memorise where in each scene which ice class is located, rather than deduce the ice class from a combination of the inputs. To combat this potential problem, some data augmentation techniques are applied. Work by Shorten and Khoshgoftaar, 2019 is recommended for an overview of data augmentation techniques. Specifically, random crops, rotations and flips of the global feature input are used. These are easy to accomplish, significantly multiplying the number of possible samples. Because



(a) S1 HH channel (b) without global features (c) with global features

Figure 3.7: Demonstration of the effect that global features can have for ice classification. Open water is depicted in blue, multi-year ice in red and first-year ice in yellow. The SAR scene is a Sentinel-1 Extra Wide mode acquisition from the first of December 2019.

the classifier needs to work throughout multiple seasons, which correlate with the ice type distribution and the radar response, the acquisition time proved to be a helpful parameter for the model. With only 44 different acquisition times, strong artificial noise is added to reduce the risk of overfitting. The random noise is sampled from a normal distribution with a standard deviation of one week.

Incidence Angle Information. The effect of the incidence angle on radar backscatter is well-researched. The correlation for labelled data in this study is shown in figure 3.3b. However, one does not have access to the local incidence angle slopes. Clustering techniques, as developed by Cristea et al., 2020, need a greater range of incidence angles than present in the StripMap scene, which spans only 2-3 degrees. Incidence angle normalisation to σ_0 is useful, but this does not account for different gradients across ice classes, which are reported, for example, by Mahmud, Geldsetzer, et al., 2018 and Lohse et al., 2021. In fact even after normalisation, almost all ice types still have negative gradients. As a compromise, the incidence angle is directly included as an input to the classifier. This allows the model to learn these differing incidence angle dependencies of the sea ice backscatter, similar to the classifier in Lohse et al., 2021, provided the range of incidence angles is covered well enough by the training data. However, the network is not forced to use the incidence angle information in a predefined manner - this is only done implicitly.

Layer Configuration. The exact details of the network hyperparameters are largely based on heuristics and experiments. The 3 x 3 kernel size in the convolutional layers has proven most useful, as the input features are not that large themselves. More success was found in downsampling with convolutions with step size 2, instead of maxpooling layers - in tests, it seems the network lost a little information in the maxpool layer that was still useful for classification. The LeakyRelu activation function introduces some necessary non-linearity and does not suffer from the problem of vanishing gradients. The latter property is especially useful for deeper networks. Strong regularisation is achieved with multiple

input name	dimensions	contents
local	5x5x4	4 channels, $r_{rel} = 1$
superlocal	16x16x4	4 channels, $r_{rel} = 1/5$
global	64x64x4	4 channels, $r_{rel} \approx 1/64$
extra	4	x_{global}, y_{global}, θ_{inc}, t

Table 3.3: Table showing input features and their content. The resolution r_{rel} is is given relative to the full resolution product. The four channels used are (HH, VV, HH-VV, HH/VV). The coordinates x_{global}, y_{global} are the normalised pixel coordinates of the region to be classified in the rescaled global patch.

dropout layers with a dropout rate of 0.3. This is necessary because the training data is still quite sparse in contrast to the scope of possible backscatter signatures from sea ice, and thus overfitting remains a concern. Additionally, small spatial dimensions after convolutions (before flattening) and fewer neurons are used to force the network to parametrise the input features. This leads to better extrapolation to unseen data. Batch normalisation in the early layers slightly sped up the convergence of the network. The Adam optimizer developed by Kingma and Ba, 2014 was used to update weights during training.

Loss Function. The network uses a categorical cross-entropy loss appended by an additional term from the FESTA loss from Hua et al., 2021, specifically the distance of the softmax outputs. The additional term encourages the separation of predictions independent of correct classification, leading to a more decisive output and increased robustness.

Labels. An important step in the label preparation is to smooth the labels. Instead of feeding a one-hot vector as a label - where the correct label is denoted as 1, all others as 0 -, uncertainties are integrated into the labelling in a rudimentary way. The idea is to treat the label vector as a set of probabilities rather than as a boolean vector. This is particularly useful for the ice classes where the manual labelling is most error-prone. In this case, distinguishing deformed and level ice benefitted most of this treatment because it is partially non-local property. Explicitly, ice has varying deformations across larger regions so that individual pixel-sized areas might be smooth, but it is apparent from the surrounding ice that the area is deformed. Including uncertainties only minimally lowered the accuracy. However, it leads to significantly increased robustness, which is preferred in this case. This is in line with observations made in Asadi et al., 2021. To smooth the labels, a random sample from a uniform distribution is taken from a given interval that parametrises the uncertainties. The intervals used across the different classes are listed in table 3.4 and were chosen qualitatively in line with the experience of which areas are difficult to label. Note that after random sampling, each output vector is normalised. A labeled scene from the test set is shown in figure 3.4, together with the model's prediction.

Implementation. The network was implemented using the tensorflow library for python (whitepaper by Martıń Abadi et al., 2015). On a an Intel i7-9850H, a commercially available mid-range CPU, inference for an entire scene consisting of $\approx 2.5 * 10^6$ classifications takes

	ow	ti	li	di	hdi
ow	$U_{[0.7,1]}$	0	$U_{[0,0.3]}$	0	0
ti	0	1	0	0	0
li	0	0	$U_{[0.9,1]}$	$U_{[0,0.1]}$	0
di	0	0	$U_{[0,0.3]}$	$U_{[0.7,1]}$	0
hdi	0	0	0	0	1

Table 3.4: Table shows the label smoothing parameters used. $U_{[a,b]}$ denotes a random sample from a uniform distribution from the interval $[a, b]$. The label vectors were normalised after random sampling.

around 8.5 minutes.

3.4.3 Discriminator

The discriminator model has a nearly identical structure to the classifier (fig. 3.6), except for the additional input layer containing a proposed label and the output being one-dimensional. Its task is to check whether the proposed label is correct. This binary classification is fundamentally easier than predicting one of five classes and can correct for some systematic errors the classifier makes. The discriminator is trained on randomly mislabelled data as ground truth for mislabelled patches, which performed better than a discriminator trained specifically on the correctly labelled and mislabelled data of the classifier: The data suggests specific training promotes an overfit to training data, and the whole paradigm of this architecture is to promote generalisation to a wider range of conditions. This step particularly helps with mitigating open water and thin ice misclassifications.

3.4.4 Conditional Random Field

The pixels deemed to be wrongly classified by the discriminator are removed from the classified product. Then, a dense conditional random field is applied that has a bilateral kernel next to the unary potential. This fills the missing values and clears up some noise-like mislabels, like single pixels classified differently than all their surrounding pixels. The implementation of the dense conditional random field (DCRF) is straightforward, using a Python implementation by Beyer, 2015 for an algorithm published in Krähenbühl and Koltun, 2012. A bilateral approach is used, with the energy function E given by a unary ϕ_u and a bilateral term ϕ_b, such that for N feature vectors y_i and labels x_i

$$E = \sum_{i}^{N} \phi_u(x_i, y_i) + \sum_{i,j}^{N} \mu(x_i, x_j)\phi_b(y_i, y_j). \tag{3.1}$$

The label compatibility function μ is learned from the classified data and describes the relationship of how likely labels are to occur next to each other. Thus, the energy function penalises incompatible labels close to one another. In this case, the features consist of a vector of colour intensities $\vec{I} = (I_r, I_g, I_b)$ across the RGB channels of a colour composite

image as well as the position of the pixel $\vec{P}$. The RGB channels in the colour composite image are VV, HH-VV and HH/VV, appropriately scaled to capture the relevant dynamic range. The unary potential is given as the logarithm of the probability $p(x_i|y_i)$ of label x_i given feature y_i. It is modelled by the softmax output of the classifier. The bilateral term consists of weighted differences in position and colour. Thus, the energy function can be expressed as

$$E = \sum_{i}^{N} \log p(x_i|y_i) + \sum_{i \neq j}^{N} \left[\frac{\|\vec{P}_i - \vec{P}_j\|}{s_{xy}} + \frac{\|\vec{I}_i - \vec{I}_j\|}{s_c} \right], \tag{3.2}$$

where $\|\ldots\|$ denotes the euclidian norm. The weights s_{xy} and s_c were adjusted manually to balance smoothness and classification accuracy.

3.4.5 Robustness

As mentioned in the introduction, there are 162 scenes to test the robustness of the classifier across eight months of different conditions in the Arctic ice. The idea is to test the ice distribution of the same patch of sea ice over this entire time period and investigate how it changes over this timeframe. This should give insight into how stable the classifier performs at high resolution. Given the positioning of the RV Polarstern, no pack ice can be tracked more accurately than the ice around the research vessel itself. Thus, this is the region to be used. The first step is of course to classify all scenes in the robustness evaluation data set. Then it is possible to calculate the probability of pixels not changing class, which is a good measure of robustness. The window chosen is approximately 3 km by 3 km in size. Of course, one cannot expect the ice to stay static over the entire time period; ice dynamics and new ice growth will change the ice type distribution. Rapid change in ice type is, however, constrained to the open water and thin ice classes, which are ignored for this analysis. The other ice classes are typically stable over the period of 3 days which is used to calculate the robustness scores. Change due to shifting of the floe is easily spotted by looking at the individual images and thus can be taken into account qualitatively during the model evaluation. It should also be noted that care was taken not to label the area used for robustness analysis in the training set, so the classifier has not 'seen' these regions. Using the ship's GPS information, the drift of the ice can be corrected by using a coordinate transformation to local ship coordinates published by Hendricks, 2019. Then, it is trivial to identify the same area for each scene.

The robustness criterion mentioned in the previous paragraph is defined concretely to obtain a quantitative measurement of robustness. A (pixel-sized) ice area is deemed to be classified *robustly* in one scene if the same prediction for the same area is made for the previous and the following scene. This criterion should be defined in a way that parametrises robust classification in one scene, not three scenes. Note that the computed probability $P_i^3(c)$ of finding the same class c at the same spot for a scene i and its two nearest neighbours is a product of the probabilities $P_i(c)$ of having robustly classified in each of the scenes.

	OW	TI	LI	DI	HDI
OW	90.89%	1.85%	7.21%	0.04%	0.01%
TI	0.1%	78.84%	11.34%	6.25%	3.46%
LI	0.8%	6.48%	89.14%	3.5%	0.08%
DI	0.04%	6.82%	17.22%	75.1%	0.82%
HDI	0.01%	1.82%	0.24%	0.62%	97.85%

Table 3.5: Confusion matrix showing the percentage of the VGG16's predicted classes (cols) for all ground truth labels (rows) on the training data. Classes are open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (HDI).

$$P_i^3(c) = P_{i-1}(c)P_i(c)P_{i+1}(c) \tag{3.3}$$

With the assumption that $P_i(c) \approx P_{i+1}(c)$, one can thus approximate the probability $P_i(c)$ of having classified robustly for scene i and class c as

$$P_i(c) = \left(P_i^3(c)\right)^{\frac{1}{3}} \tag{3.4}$$

The regions of ice used to test this are the pixels in the stabilised images, such as seen in figure (3.11). In the following analysis, any statements derived must assume that the same pixel over three scenes actually maps to the same physical area of ice for three consecutive scenes. As the stabilisation is imperfect and ice dynamics are neglected in this assumption, this is not necessarily the case. Treating these phenomena as some underlying noise in the analysis means that when robustness according to this criterion is calculated, it is only a lower bound of the real robustness - given that random noise is entirely not robust.

3.5 Results

The classifier was trained on 44 scenes and tested on 8, which do not contribute to training data. The eight scenes that make up the test set are randomly selected from October through May. The performance of the classifier is analysed across both datasets. As a comparison, a simple VGG16 (original publication in Simonyan and Zisserman, 2015) inspired architecture is an alternative classifier (see figure 3.13 in the appendix for details). This model only uses the superlocal 16 x 16 input data. The classifiers' performances are shown in tables 3.5, 3.6 for the training set and tables 3.7, 3.8 for the test set.

The two classes for which both classifiers perform worst are open water and thin ice. The classification across the other three ice classes are more accurate, particularly for the proposed classifier. Heavily deformed ice stands out as being especially easy to classify accurately. Open water and young ice examples are the scarcest in the training and test scenes. Despite balancing the amount of samples per class upon training, the data set is much less diverse for these samples. In other words, the number of regions with open water or thin ice is significantly lower than that of the other classes. This can be seen (albeit at a smaller scale) in figure 3.9.

	OW	TI	LI	DI	HDI
OW	99.29%	0.11%	0.38%	0.21%	0%
TI	0.28%	93.72%	1.36%	4.11%	0.53%
LI	0.16%	0.54%	97.38%	1.89%	0.02%
DI	0%	0.58%	1.47%	97.84%	0.1%
HDI	0%	0.11%	0%	0.62%	99.26%

Table 3.6: Confusion matrix showing the percentage of the developed network's predicted classes (columns) for all ground truth labels (rows) on the training data. For example, 0.11% of data points of OW were incorrectly predicted to be TI. Classes are open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (HDI).

	OW	TI	LI	DI	HDI
OW	80.89%	4.6%	14.23%	0.27%	0.01%
TI	2.4%	54.48%	17.42%	15.36%	10.34%
LI	0.54%	11.15%	81.85%	6.44%	0.02%
DI	0.17%	6.95%	13.64%	78.78%	0.46%
HDI	.010%	1.36%	0.21%	0.4%	98.02%

Table 3.7: Confusion matrix showing the percentage of the (used for comparison) VGG16 style network's predicted classes (cols) for all ground truth labels (rows) on the test data. Classes are open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (HDI).

	OW	TI	LI	DI	HDI
OW	74.86%	8.12%	16.4%	0.61%	0%
TI	5.05%	71.07%	7.03%	14.32%	2.52%
LI	0.49%	0.62%	95.51%	3.38%	0%
DI	0.02%	0.55%	2%	97.37%	0.05%
HDI	0%	0.02%	0%	0.92%	99.06%

Table 3.8: Confusion matrix showing the percentage of the developed networks' predicted classes (cols) for all ground truth labels (rows) on the test data. For example, 8.12% of data points of OW were incorrectly predicted to be TI. Classes are open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (HDI).

To illustrate the entire process of classifier assessment, an example with two consecutive scenes from the 13th and the 14th of December 2019 is given in figure 3.8. In the first step, the StripMap scene is cropped along the longer range axis to a near quadratic slice, which is needed for the global input feature (3.3). This slice is then labelled using the classifier and checked by the discriminator. The results of this step are shown in figure 3.8a and 3.8b. The next step is to apply the DCRF to refine labels and fill in missing values left by the discriminator. The results of the DCRF for the two example scenes are shown in figure 3.8c and 3.8d. Finally, the image is rotated and cropped to the surroundings of the RV Polarstern, allowing us to image the same ice region continuously. For the two scenes from December used as an example, the cropped images are shown in figure 3.8e and 3.8f.

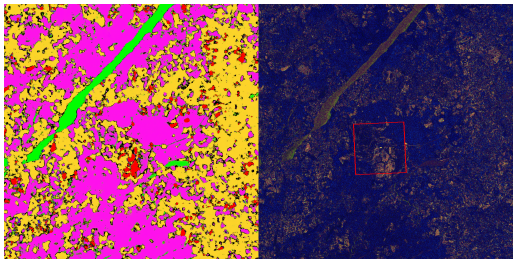
By executing the procedure illustrated in figure 3.8 for all scenes from October through May, the data necessary to perform quantitative robustness analysis is generated. Figure 3.9 shows the predicted ice type distribution evolution over the investigated time span in the three by three kilometre box around Polarstern.

The distribution chart allows insight into the classifier's performance over large timescales and shows lower stability, especially in the discrimination of level and deformed ice. Spikes of open water and thin ice are generally tied to some ice dynamics. To gain additional insight into these classes' variance, the relative standard deviation of the ice type fraction for every scene and its four nearest neighbours (fig. 3.10) is calculated. When interpreting this as a deviation of the classification, again, it is implicitly assumed the real ice type distribution is stable over five neighbouring scenes, which neglects physical changes of the surface. Specifically, one cannot include the OW and TI classes, which are rare and only present in the case of strong sea ice dynamics such as leads forming. With lifetimes shorter than five days, they cannot be analysed in this way.

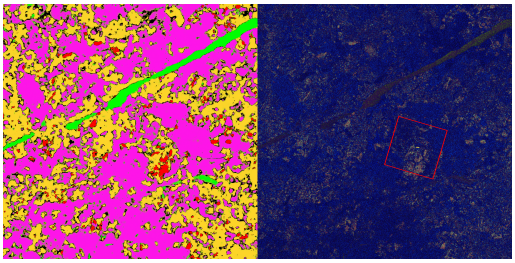
The chart of the standard errors reveals three time periods with heightened error. In fact, the one stable period around the beginning of January stands out. Here, conditions are optimal, as ice dynamics are minimal and the incidence angle is inside the full performance range. Whilst the early and later periods of increased variance can likely be explained by snow metamorphism, wet snow or increased ice dynamics in pre-melt and freeze-up seasons, the increased uncertainties from mid-January to early March can be rationalised with the increased incidence angle during this time period (see fig.3.3a). It is also apparent that this cause of error plays a role in the increased uncertainties observed during the melt season.

To qualitatively illustrate the performance of the classifier in these three periods with increased error, a highlight of three pairs of scenes is shown in figure 3.11 from those time spans, specifically late November, late February and finally, early May.

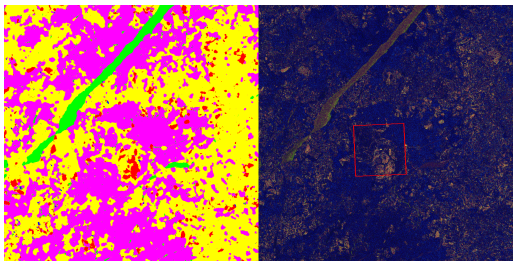
The scenes from the end of November are only three days apart yet drastically different due to ice dynamics. In the earlier scene (3.11a), one can see some freshly frozen-over leads with thin ice cover and in the later scene (3.11b), the leads have closed up again, and all signs of young ice have disappeared. These pictures document the most drastic of these events, where the central floe split; however, most of the strong ice-type deviations in early



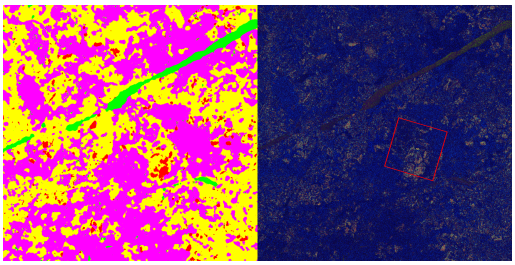
(a) 2019/12/13, $\theta_{inc} = 45.69°$



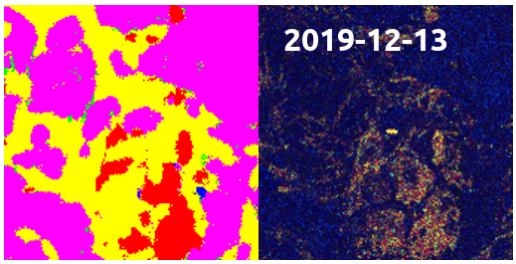
(b) 2019/12/14, $\theta_{inc} = 49.55°$



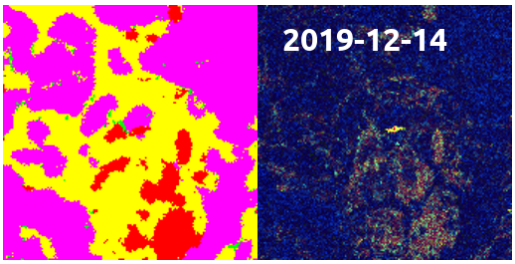
(c) 2019/12/13, $\theta_{inc} = 45.69°$



(d) 2019/12/14, $\theta_{inc} = 49.55°$



(e) 2019/12/13, $\theta_{inc} = 45.69°$



(f) 2019/12/14, $\theta_{inc} = 49.55°$

Figure 3.8: Illustration of robustness analysis for two consecutive days in December 2019. Pictures (a, b) show the sub-scenes after classification and discrimination. In (c, d) the dense conditional random field has been applied, and (e, f) shows the scenes cropped and rotated to the region of interest. The false colour compositions (VV, HH-VV, HH/VV) are used in the (R, G, B) channels, respectively. The top two rows additionally show the area used for robustness analysis, which is cropped to in the final row.

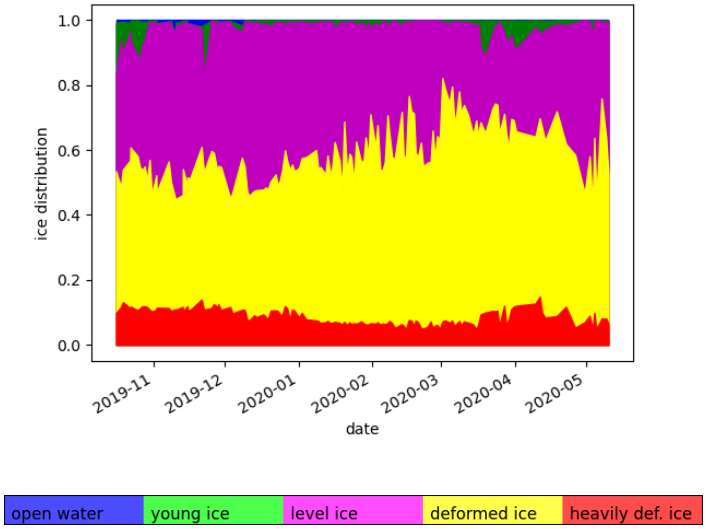


Figure 3.9: Chart showing the predicted ice distribution in the 3 km by 3 km area surrounding the Polarstern vessel from October through May.

winter can be attributed to such events and are thus real surface changes. The pair of scenes from late February (3.11c, 3.11d) are both taken at very high incidence angles, far out of the full performance range of $20 - 45$ degrees. It is evident from the images that the signal is significantly weaker. At this angle, an overestimation of the deformed ice class in the second image can be seen. This is especially notable in the top left quarter of the patch, which contained large areas of level ice surfaces but is classified as almost entirely deformed ice in the scene from the first of March. At such high incidence angles, the classification seems to become more volatile. Two scenes from early May (3.11e, 3.11f) give insight into how both ice dynamics and incidence angle changes are at least partially responsible for high variance in the scenes from early May.

Most of the high variances in ice class distribution change can be attributed to ice dynamics or struggles with high incidence angles. The classifier seems robust in the discrimination of classes with larger areas, but the transitional areas between classes are seemingly classified less robustly (see the extend of the level ice on the left of the image 3.11e versus 3.11f)). This effect is particularly evident at high incidence angles.

In figure 3.12, robustness development is shown across the dataset. This was smoothed over by a moving average, weighted with a quadratic function and averaging over five scenes. Note that the dip in the beginning is due to strong ice dynamics. Figure 3.14 shows a comparison of the proposed model with the VGG16 (fig.3.13) inspired classifier for two months. Additionally, an average robust classification probability over the entire time span is shown in table 3.9.

Consider a probability $P_{rc}(c)$ of robust and correct classification as the product of the

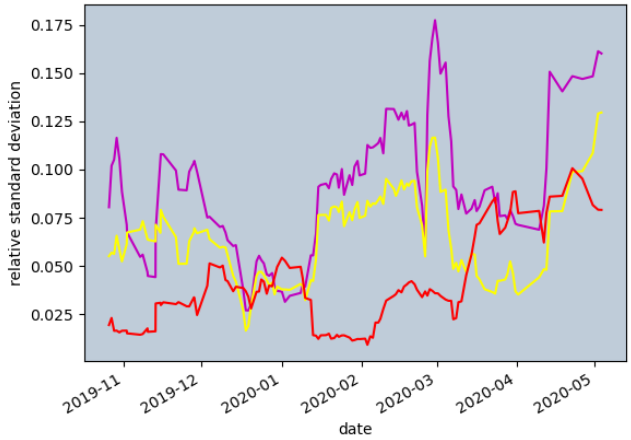


Figure 3.10: Relative standard deviation of the level deformed and heavily deformed ice distribution at each date and the four nearest scenes.

	LI	DI	HDI
P_r	88.34%	88.62%	85.34%

Table 3.9: Table showing the average probability $P_r(c)$ of robust classification for LI, DI and HDI across the entire dataset.

two, i. e.

$$P_{rc}(c) = P_r(c)P_c(c). \tag{3.5}$$

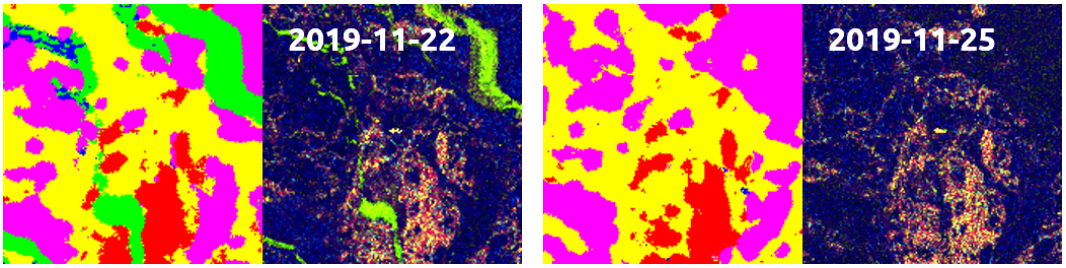
With classification probabilities P_c from table 3.8 and the lower bounds of P_r from 3.9, one can compute the bounds of probabilities of robust and correct classification P_{rc} for the three solid ice classes. Results are shown in table 3.10.

	$P_{rc}(c)$
LI	$84.09\% < P_{rc}(LI) < 95.19\%$
DI	$86.71\% < P_{rc}(DI) < 97.84\%$
HDI	$84.71\% < P_{rc}(HDI) < 99.26\%$

Table 3.10: Table showing the bounds of robust and correct classification for the classifier's LI, DI and HDI predictions.

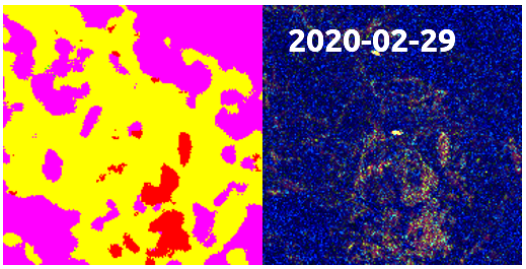
3.6 Discussion

Before turning to the approach's advantages, some limitations and challenges are discussed. First, let us discuss the data itself – the foundation of any machine-learning approach. The training data set of 44 scenes is not comprehensive enough to capture all the intricacies of different backscatter from varying ice types, making it difficult to classify robustly. Leads

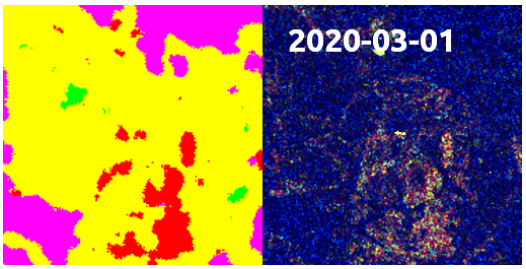


(a) 2019/11/22, $\theta_{inc} = 40.71°$

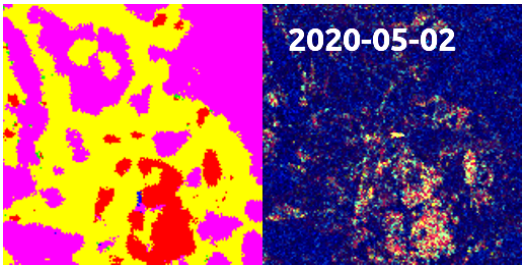
(b) 2019/11/25, $\theta_{inc} = 43.33°$



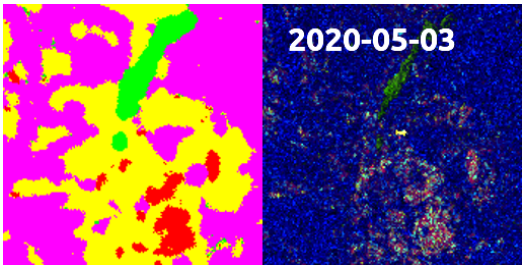
(c) 2020/02/29, $\theta_{inc} = 57.36°$



(d) 2020/03/01, $\theta_{inc} = 54.27°$



(e) 2020/05/02, $\theta_{inc} = 36.90°$



(f) 2019/05/03, $\theta_{inc} = 50.82°$

Figure 3.11: Pairs of classified scenes from time intervals with low robustness. The left parts show the cropped classified image and the right parts are false colour compositions.

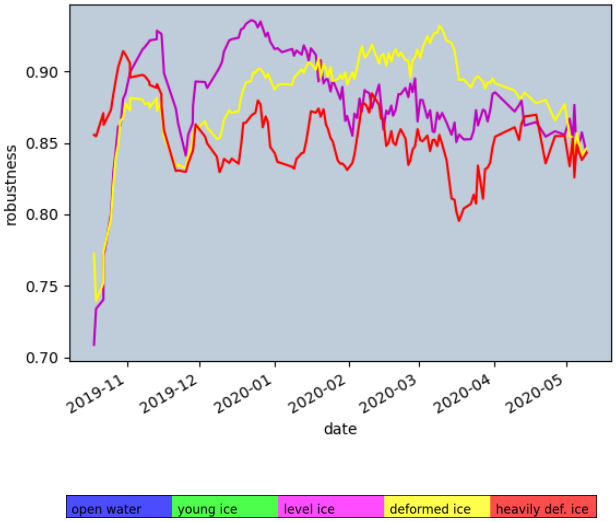


Figure 3.12: Chart showing a moving average of the probability $P_r(c)$ of robust classification for three ice classes: level ice, deformed ice, heavily deformed ice. $P_r(c)$ is the percentage of robustly classified pixels of class c per scene.

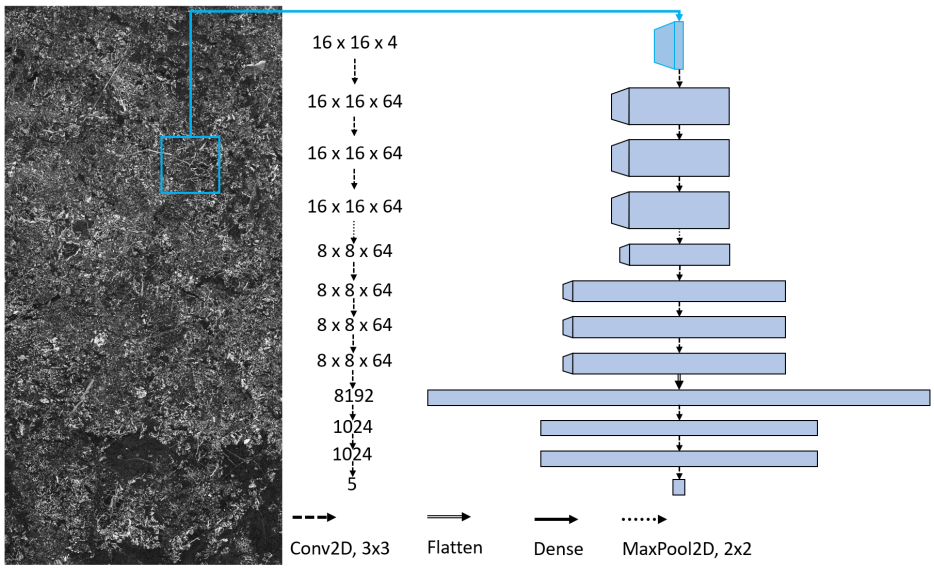
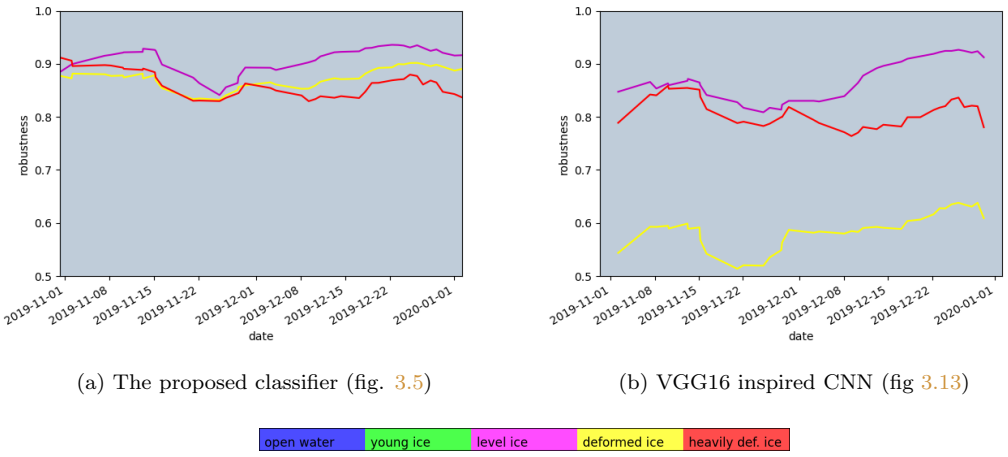


Figure 3.13: Illustration of the VGG16 inspired network architecture used as a comparison. Parameter count = 9889605.



(a) The proposed classifier (fig. 3.5)

(b) VGG16 inspired CNN (fig 3.13)

Figure 3.14: Charts showing a moving average of the probability $P_r(c)$ of robust classification for three ice classes LI, DI, HDI, for the proposed classifier and a VGG16-inspired model.

freezing over is a good example of one of such phenomena. Not only is their occurrence sparse in the data set, but the dynamics during initial freeze-over have a strong effect on radar response and are fast relative to the revisit time of the satellite. This makes it difficult to capture enough samples in the training set for the classifier to interpret the entire space of possible radar backscatter correctly. One can observe this struggle in some scenes where the radar response of a frozen lead is so bright it becomes very similar to heavily deformed ice's backscatter. This can, for example, occur when frost flowers form atop the lead, leading to high volume scattering. Here, the classifier struggles to differentiate the two classes.

The open water classification also proved challenging for this data set. Traditionally, the polarisation ratio proves very useful in distinguishing this class. At high incidence angles, the radar response becomes very similar to that of young smooth ice, and the discrimination between the two suffers.

Manual labelling is definitely the greatest source of underlying error and bias. Despite having mitigated the effect of errors with the use of smooth labels, there are some biases arising from manual labels that smooth labelling cannot compensate for. This bias is not merely a case of being more likely to mislabel a certain class - this can be kept minimal by only labelling classes that are discernible with certainty - it is rather that a human selection process already filters the choice of labelled regions. For example, there is a tendency not to label a region with a small area as it would make the labelling process very tedious. This translates to the classifier, which struggles with smaller regions of one class, often wrongfully labelling them to be the same as the surrounding ice class. Additionally, when manually selecting polygons, labels at the boundaries between classes are naturally much sparser than labels in the centre of ice classes, leading to increased classification difficulty in these transitional areas between ice classes. When viewing the classified robustness analysis data, this effect was obvious as the boundaries between classes shifted, whilst the classification of

pixels in the centre of same-ice regions appeared robust. This bias could be eliminated by deriving ground truth data from in-situ measurements.

Deformed and level ice discrimination relies on non-local features and hence suffers most from the above-mentioned boundary problem. Deformed ice is not always identified by a higher brightness and lower polarisation ratio for each individual pixel but also by the density of brighter pixels in the surrounding area. Here, it is especially difficult to define hard boundaries between classes, as the transitional areas between level and deformed ice are not boundaries but a continuum. Hence, defining a hard boundary when manually labelling data is difficult. Generally, the rule when labelling manually is only to label areas where one is confident in the label. Therefore, these transitional areas are not only difficult to classify but also sparse in the training data set, which culminates in misclassification in the transitional areas of deformed and level ice classes, especially at high incidence angles, where the signal-to-noise ratio suffers (fig. 3.11).

The post-processing step using a Markov Random Field relates spatial information of classes but uses the backscatter intensities as a guide for class distributions. With no ground truth data available, it is not at all clear that the spatial distribution of backscatter is strongly related to the spatial distribution of ice types, as very similarly bright regions can stem from very different ice surfaces. Therefore it is not an optimal way to introduce spatial information to the labels. The core problem here is the lack of ground truth, meaning the spatial ice type distributions are not available to be learned from.

The success of the algorithm is evident in the discrimination of ice classes at high accuracy in multiple seasons and becomes increasingly apparent in contrast to the VGG16-inspired network (see fig. 3.14, tab. 3.5 - 3.8). Furthermore, the areas of lower robustness that can be seen to occur at high incidence angles are outside of the full performance range of the radar instrument.

An analysis of approximately 500 scenes around the MOSAiC experiment shown in figure 3.15 demonstrates that whilst deformed ice increases due to dynamics as expected, the heavily deformed ice does not increase. This suggests, that the processes that lead to the transition to from the deformed ice to heavily deformed ice are not happening in the study region over the observed timescales. This leaves two plausible reasons: The increase in backscatter associated with this class could be due to advanced desalination of the ice, that takes place over longer timescales. Alternatively, the forces needed to deform the ice to that degree are not present in this part of the Arctic. This means that the heavily deformed ice we did detect is probably either significantly older or originated in a different part of the Arctic, where deformation processes are stronger. These results are corroborated by surface roughnesses extracted from two airborne laser scanning flights over the study area from the 30th of November and the 27th of February, as can be seen in figure 3.16 At this point we cannot compare the deformation rates of first and second year ice, because the classifier does not perform at pixel level resolution. Thus additional deformation in already deformed second year ice areas cannot be detected.

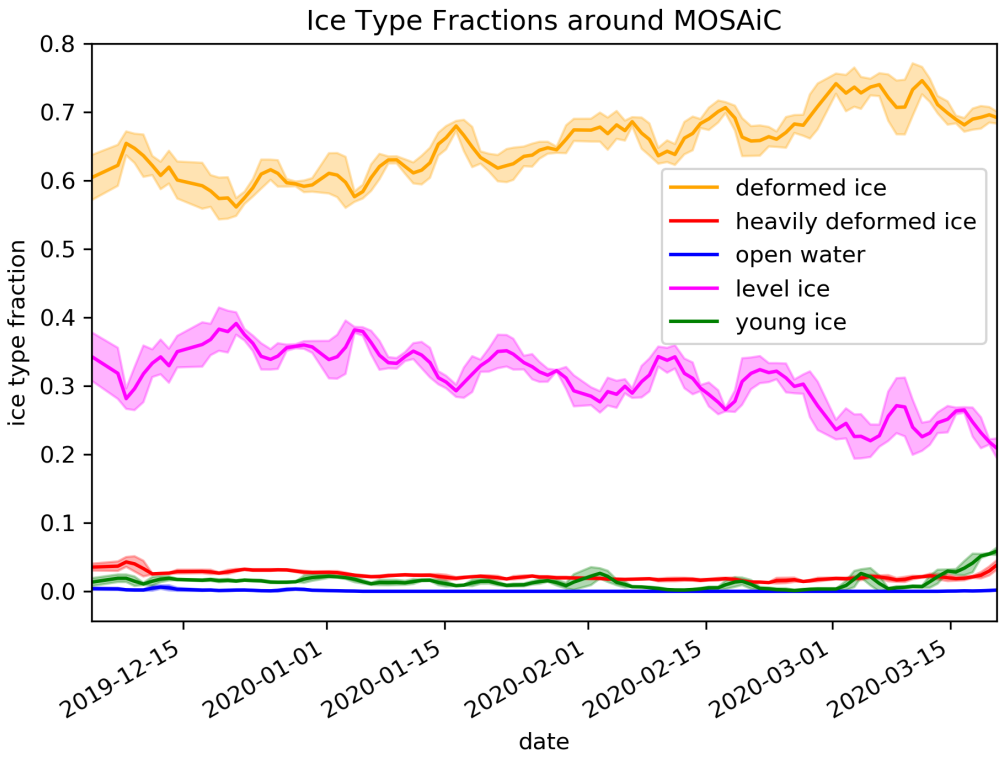


Figure 3.15: Ice type fractions from approximately 500 StripMap scenes near the MOSAiC floe from the most stable period of classification between December and April. A steady increase in deformed ice can be observed, whilst the heavily deformed ice is constant until increased dynamics in March. The increase here is most certainly due to classification of young ice in open leads. The small scale variance in the fractions is due to uncertainty of the classifier and the fact that the area imaged is not always exactly the same, but only a subsample of the same region.

Weather effects contribute significantly to snow wetness, metamorphism and increased ice dynamics. The most notable of these is the seasonal warming and cooling, which leads to decreased robustness in the analysis (fig. 3.12). However, the robustness criterion also fails to take into account these weather-induced changes, and there is less training data available in these shorter time periods, especially as they are at the very beginning and end of the study period. Thus, it is difficult to isolate and make statements about the effect of weather events on classification performance.

This classification approach was also tested on Sentinel-1 scenes and obtained comparable results. It was found that the most important parameters to tune when applying these ideas to different sensors are the sizes of the contextual windows ('local' and 'superlocal' features). On large-scale images, including the 'global' feature was particularly successful in ice and open water discrimination in the marginal ice zone, where the ice water edge could be detected (see figure 3.7.

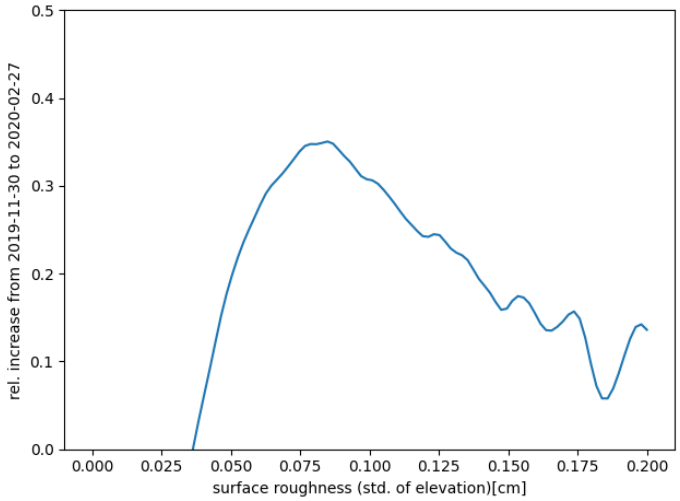


Figure 3.16: Increase of surface roughness (std. of surface elevation) distribution between 30th of November and the 27th of February relative to its abundance on the earlier date. Measured with a helicopter borne airborne laser scanning device that measured in area approximately 5 by 10km centered over the MOSAiC floe in both cases. This shows, that higher roughness areas increased disproportionately little over the time period, suggesting that the forces involved in their creation are not at play here. The variance increases for the very high roughness regions, because the sample size is very small there. Also, the area flown over was not exactly the same. Continuing the downward trend see in the plot would mean that very little to no new very rough ice surfaces were formed between the two measurements, despite some of those the existing at the outset.

3.7 Conclusion

Accurate geolocation and drift correction were used to construct a dataset that enabled quantitative testing for the robustness of SAR ice-type classification. The proposed classification method performs accurately and robustly for three surface ice classes: level, deformed and heavily deformed ice. Open water and thin ice classes have proven harder to classify. However, it must be noted that these classes are also sparser in the dataset and have been more difficult to identify in some scenes, especially at higher incidence angles. Due to their dynamic nature, one cannot perform a robustness analysis for these two ice types. It was shown that regions of increased classification inaccuracy and lack of robustness coincide with shortcomings of a manual labelling process.

The work presented in this chapter marks a first foray into sea ice classification using advanced neural network configuration to incorporate additional information to further network accuracy and robustness, specifically for sea ice retrieval problems. It also presents a systematic way to test a network's robustness using a large time series of observations. The network architecture designed for ice classification from manual labels in this work spearheads three major design principles: The first is to limit the number of parameters in the

network. Secondly, to include a large range of contextual information and thridly to incorporate the uncertainties of the labels and classification by using discriminators and label smoothing. The heuristics behind limiting the number of parameters is that the training set cannot accurately represent the full spread of ice conditions in nature. Limiting the number of parameters makes it impossible to fit specific data features overly, forcing the network to generalise to more universal features. The increase of contextual data is managed by including differently sized contextual windows with diverse resolutions. A central concept here is that features that are further away from the pixel to be classified need to be larger to impact the class significantly. Thus, it is sensible for the contextual windows, including a larger part of the scene, to be of lower resolution. The advances presented here help solve the inversion problem of sea ice classification with advanced neural network configurations tailored to the manual labels used in this study and underline that the label properties transfer to the classification properties.

Chapter 4

Extrapolating Measured Classes - Connecting Sea Ice Topography and SAR Signal during MOSAiC

4.1 Overview

In the previous chapter strategies for robust learning for high resolution imagery was developed, based on rough human annotations. In this chapter, machine learning approaches are reviewed for high resolution labels, that are extracted from a co-located measurement of the topography, made by an Airborne Laser Scanner. The aim of the following work is to find meaningful directions for further convolutional neural network based algorithms, that aim to extrapolate labels at the resolution of the sensor and to give a realistic perspective on the performance of existing architectures that have thus far not been tested only on manual annotations not on much more representative measured labels. The contents of this work are currently under open review and published as a pre-print for discussion in Kortum, Singha, Spreen, Hutter, et al., 2023, which is first-authored by the author of this thesis, who conducted the research presented below.

4.2 Introduction

Sea ice classification from remote sensing and especially SAR instruments have been used for monitoring the Arctic sea ice for multiple decades, with automation being proposed as early as the mid eighties by Fily and Rothrock, 1986. However, even with the inception of advanced machine learning methods and modern data analysis, there does not yet exist a universally reliable classifier to retrieve sea ice classes from radar imagery. At this point, with many different classifiers having been proposed and developed (e.g. Kwok, Rignot, et al., 1992; Soh and Tsatsoulis, 1999; Hara et al., 1995; Karvonen, 2004; Ressel, Frost, et al., 2015; Doulgeris, 2015; Johansson, Malnes, et al., 2020; Lohse et al., 2021) one must ask the question why no meaningful direction has yet established itself in the ongoing research. The answer to the question - aside from the complexity of the subject - is twofold. Firstly and most important is the state of the data. Although there exists a great wealth of satellite SAR acquisitions of the sea ice in diverse states and conditions, the corresponding ground truth information is lacking. Secondly, the constantly varying and difficult-to-predict drift and deformation of sea ice makes it nearly impossible to image the same area of sea ice over longer time series to evaluate any proposed classifiers' robustness. The latter is particularly true for high-resolution imagery. These two shortcomings open this topic up to a plethora of different challenges because one has almost no way to test, iterate and improve sea ice retrieval algorithms in a structured manner. This stifles the rate at which progress in the field can be made or even recognised.

On a mission to fill gaps in our knowledge about the Arctic sea ice and its climatology, the MOSAiC expedition launched in the autumn of 2019 and the ship Polarstern spent a year adrift with the ice pack. Aboard, interdisciplinary teams of scientists worked to collect as many data as possible, which will help to further our understanding of one of Earth's most remote regions. With the mission came the unique opportunity to collect exactly the type of ground truth over a long time period, that is needed to test sea ice retrieval algorithms,

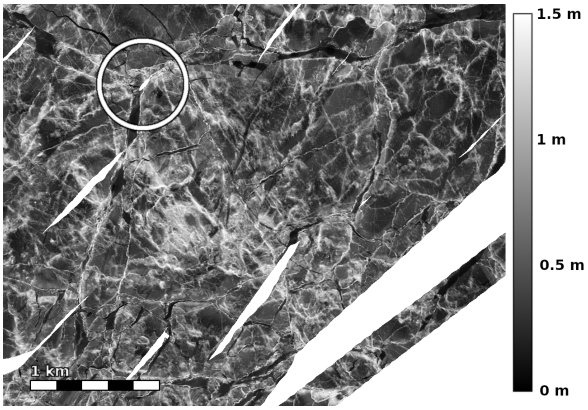


Figure 4.1: Section of ALS measured freeboard over the MOSAiC floe on April 8, 2020. RV Polarstern can be seen in the centre of the white circle. Brighter values correspond to higher freeboard values whereas white areas indicate no data. The displayed freeboard range is 0 to 1.5 metres.

with satellite-borne SAR data being acquired at the same time. An overview of the snow and ice related activities is given in Nicolaus et al., 2022.

Ice and snow transects from Itkin et al., 2021 or drilling hold the most detailed information of the underlying ice. Unfortunately, the spatial extents of these measurements are too sparse to be used for comparison with the satellite acquisitions. Aerial measurements taken from helicopters, such as the Airborne Laser Scanner (ALS) data products by Hutter, Hendricks, et al., 2022b; Hutter, Hendricks, et al., 2022a being used in this approach (Fig. 4.1) provide information about the height of the snow and/or ice surface above the local sea level, i.e. freeboard, and surface reflectance at scales of kilometres to tens of kilometres. These data are therefore a prime candidate to extract ground truth information for ice classification based on roughness and thickness.

One prominent emerging method of segmenting image data are machine learning based approaches based on convolutional neural networks, such as published in Simonyan and Zisserman, 2015; He et al., 2015; Liu, Mao, et al., 2022; Ronneberger et al., 2015; Zhou, Siddiquee, et al., 2018; Zhou, Siddiquee, et al., 2019. Advancements in the field of machine vision are being made at a rapid pace, able to leverage the improvements in chip design and the increasing amount of data that are being generated. The image-like properties of SAR acquisitions mean that this knowledge is transferable to the ice classification domain (e.g. Boulze et al., 2020; Ullah et al., 2021; Wang and Li, 2021; Kortum, Singha, Spreen, and Hendricks, 2021; Kortum, Singha, and Spreen, 2022). Historically, this has been done with texture extraction and subsequent dense neural networks as in Ressel, Singha, et al., 2016; Singha et al., 2018; Murashkin, Spreen, et al., 2018, pixel-wise classification using image classifiers based on convolutional neural networks as by Boulze et al., 2020; Ullah et al., 2021 and segmentation models that are able to segment an entire patch simultaneously like in Wang and Li, 2021.

In this study, the unique opportunity provided by 20 instances of near-coincident ALS and SAR data over a period of 8 months is used to compare a variety of machine learning-based classification approaches in terms of classification accuracy and robustness on classes delineated directly from measurements. The time difference between measurements is seven hours on average. For the first time, accurate, high resolution sea ice topography measurements of freeboard and surface reflectance have been taken, which exhibit high spatial overlap and low time differences with SAR data. This enables us to truly test the capability of retrieving freeboard and (above snow) surface roughness based sea ice classes from SAR data. In contrast to existing ALS and SAR datasets, such as produced in Singha et al., 2018, the MOSAiC experiment provides the opportunity to monitor the same ice across a large temporal time span at high resolution. The amount of colocations achieved here is significantly greater than in previous studies, which enables the training of deep learning models requiring large datasets.

4.3 Methodology

4.3.1 The Data

The SAR component of the analysis is made up of TerraSAR-X X-band acquisitions in StripMap (SM) mode. The intensity scenes are normalised to σ_0 and calibration is performed as per the product specifications in Fritz et al., 2007. The resulting scenes have a pixel spacing of 3.5 metres and a native radiometric resolution of 16 bit. Both HH and VV bands are acquired by the satellite simultaneously. This configuration of polarisations has been shown to yield valuable information for ice classification in Ressel, Singha, et al., 2016Geldsetzer and Yackel, 2009. As only 2 bands can be acquired simultaeneously, the cross-pol band is not present in the data. Each combination of two channels will have some shortcomings, however, so this needs to be accepted. The footprint of a single scenes is typically around 50x15km.

The ALS data from Hutter, Hendricks, et al., 2022a; Hutter, Hendricks, et al., 2022b from 20 scenes (appendix 4.7) between October 2019 and May 2020 are used to delineate sea ice classes. The data were acquired by flying a mow-the-lawn pattern over the ice near the MOSAiC central observatory. The resulting ALS grid has a geospatial resolution of 0.5 metres. For midwinter flights in high latitudes of >85°N, the post-processing of the helicopter INS/GPS data failed and ALS data processing was performed using a lower frequency real-time navigation solution with metre-scale undulations in GPS altitude that propagated to the surface elevation retrieved from the ALS. The undulations in the computed freeboard could be minimised using a correction calculated from swath-to-swath overlap. It should be noted that the local standard deviation of the freeboard is left intact by these processing artefacts and can still be used to derive a parametrisation of the local surface roughness, where these undulations are present. An additional measurement aside from freeboard is the surface reflectance at the wavelength of the laser (1064 nm), which is useful to identify

regions of young ice that have not yet been covered by snow. For the acquisitions with unphysical undulations in the freeboard measurement, freeboard was not used to delineate class labels. Instead, only classes which could be inferred from the surface roughness and reflectivity were used. The footpring of a single flight is typically around 5x10km.

Colocation: For each ALS grid, the first step for co-locating with SAR data is to find the SAR acquisition that is closest to the ALS measurement time, whilst still having substantial spatial overlap. Then, by using the Polarstern ship to determine a common coordinate system, the two measurements are fused by assigning each ALS data point to the closest TSX pixel (see. Kortum, Singha, Spreen, and Hendricks, 2021; Hendricks, 2019.) In the common coordinate system, this means that the two measurements are in the same TerraSAR-X grid cell relative to the ship. Because of the difference in resolutions ($0.5m$ ALS and $3.5m$ SAR), one obtains approximately 49 points of ALS measurements per SAR pixel. The freeboard and roughness are then computed as the respective mean and standard deviation of these points. Investigation showed that the median and mean of the local distributions where on average within less than a percent of the span of the distribution. This lends confidence, that the distribution is roughly symmetrical and thus the mean and standard deviation describe the statistical nature adequately. Using the Polarstern as an origin of the common coordinate system is sensible, as accurate GPS positioning and heading account for ice drift and rotation. The matching of the two products using this method was accurate to a couple of metres. To further improve the accuracy of colocation, a final translation and rotation was then determined manually. Afterwards, the features overlapped perfectly at (TerraSAR-X) pixel resolution. The accuracy of co-location is made possible by more than daily TerraSAR-X SAR acquisitions of the MOSAiC floe, which helps keep the time differences between satellite and helicopter measurements small.

Determining labels: We have categorised the measured sea ice into three classes. A label is given for each SAR pixel, for which ALS information is available. For ease of reference, they are given names which are easier to contextualise. However, the exact definitions of the classes is given here. They are fully given by the ALS measurement. The three classes are: Open water and young ice (OW/YI), level first-year ice (LFYI) and deformed first-year and multiyear ice (DFYI/MYI). These classes are defined as follows (see Fig. 4.2 for a visual aid):

- OW/YI: Ice whose reflectance (range corrected target echo amplitude) is significantly lower than that of the surrounding snow covered ice. Typically values around -7dB were used as a threshold value and adjusted manually if needed. Note, that finer separation here is not possible from the data alone, but from reports of scientists on the expedition it is known that most ice in this class will have already formed a thin ice layer and entirely open water was very rare during the flights.

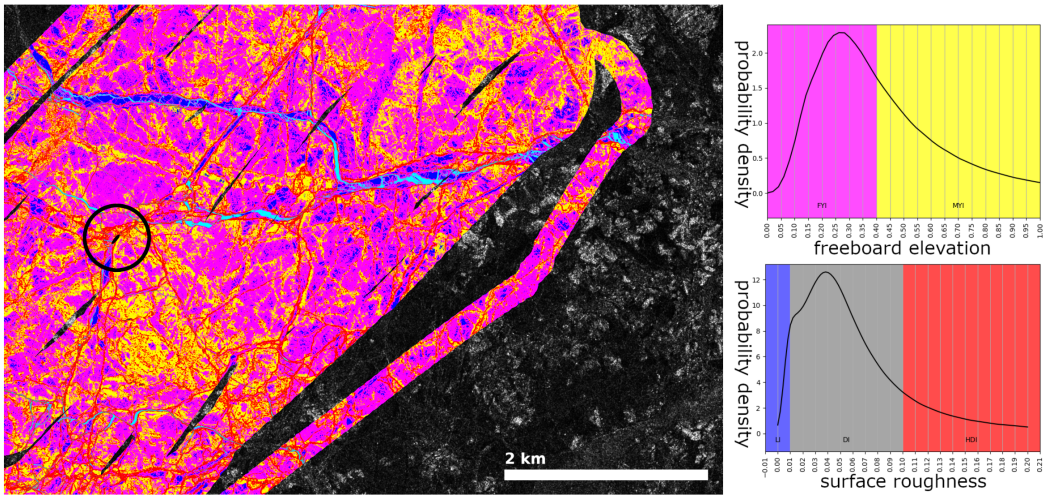


Figure 4.2: Derived labels from the ALS acquisition on the April 8, 2020 overlayed on the HH channel of the near-coincident SAR measurement (left) and estimated probability density functions from the distributions of freeboard and surface roughness (in this case this is the local standard deviation of the freeboard) (right). *Yellow* indicates ice with a higher freeboard than the high inflection point of the distribution. *Magenta* is ice with a lower free board than that. *Red* are areas with higher surface roughness than 10 cm. *Blue* areas are ice with surface roughness of less than 1 cm. *Cyan* areas have reflectivity indicating no snow cover (less than -7dB Echo Amplitude). For this study, yellow and red, as well as magenta and blue classes are combined. The grey background of the surface roughness distribution denotes the region that was not used to identify ice classes, as there was considerable mixing in this parameter region.

- LFYI: Snow-covered ice with a surface roughness (standard deviation of freeboard measurements at scales of the ALS grid ($0.5m$) calculated over one TSX pixel ($3.5m^2$)) of less than 1 centimetre or a freeboard value lower than the higher inflection point in the freeboard distribution (typically around 40 centimetres).

- DFYI/MYI: Snow covered ice with a surface roughness of more than 10 centimetres or a freeboard greater than the higher inflection point in the freeboard distribution.

Because these labels are entirely defined by measurements of the ice surface (Fig. 4.2), one can also infer the probabilities of belonging to each class, by assuming a gaussian distribution of ALS freeboard and reflection measurements at each TSX pixel. From the 49 ALS measurements, the mean and standard deviation of the freeboard are computed and then the probabilities of lying below or above any freeboard thresholds are given by the error function. Explicitly, one can integrate the area under the curve of the gaussian distribution, above and below the threshold, to get the desired probabilities. Thus one obtains labels which give the probabilities of belonging to a certain class, rather than discrete classes. Assuming a gaussian distribution also allows to infer uncertainties of the surface roughness.

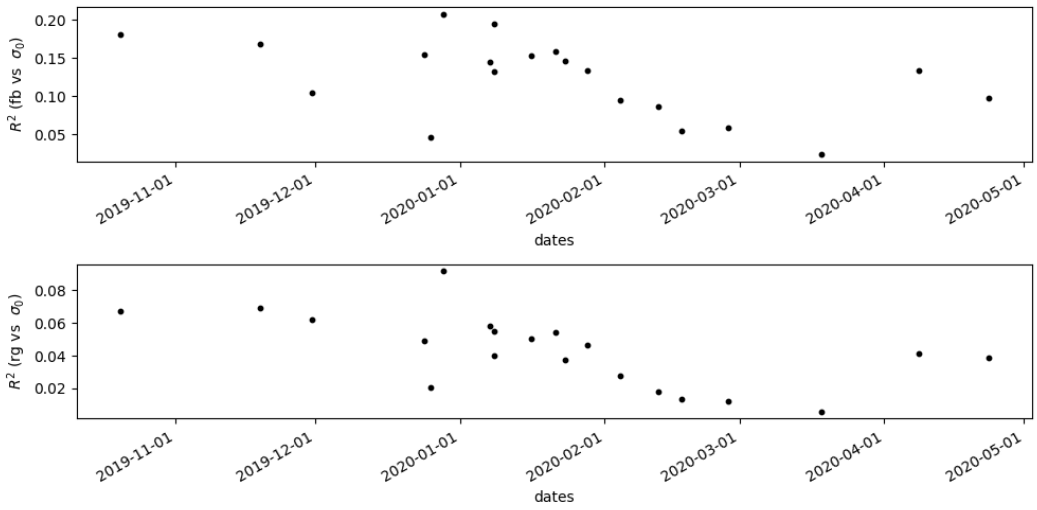


Figure 4.3: Evolution of correlations between freeboard [fb] or surface roughness [rg] and HH SAR backscatter [σ_0]over time. Note that the surface roughness is measured at the snow atmosphere interface and at correlation lenghts of 0.5 metres, whilst the SAR sensor is most sensitive to the ice snow interface and roughness at correlation lengths at the wavelength of the sensor, which is only 3.1 centimetres. The same analysis with VV channels gives very similar results.

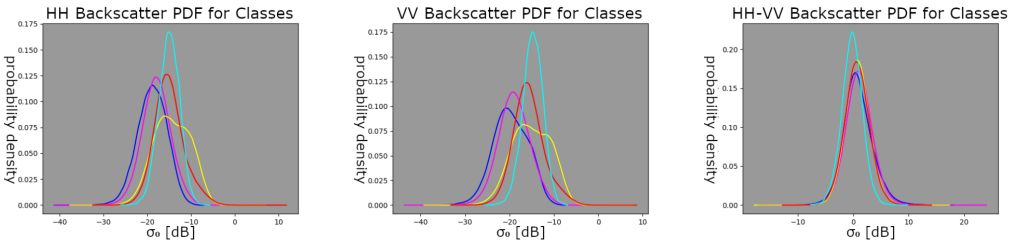


Figure 4.4: Approximate probability density functions (PDFs) for the sigma nought backscatter of each class across the different polarization configurations, for one flight on the 8th of April. Note that no two classes can be reliably separated using backscatter alone. Colours are as defined as in Fig. 4.2

The derived labels from each scene are split into two mutually exclusive connected subsets. By connected it is meant, that in all but edge cases pixels are neighbouring ones from the same subset. The training set is made up of 75% of labels whilst the test set consists of 20%. The remaining 5% of the data is used as a validation set. The validation data is used only to decide when to stop training. All subsets (test, training, validation) contain data from every scene. Imbalances of the classes were handled by balancing the dataset for pixel-wise classifiers and weighting the classes inversely to their frequency for the segmentation approaches, where an entire patch is segmented at once. Thus the training of the networks is set up so that performing equally well for each class yields the lowest loss. As the classes are not balanced in the labels, better performance would certainly be achieved on the training data set without balancing the classes, but it would hinder the generalisation of the classifier and make the results more difficult to interpret. As generalisation to a larger space of ice conditions is a property one would like to be reflected in the results as directly as possible, balancing was undertaken here.

In Figure 4.3 the correlation between backscatter and surface topography measurements is shown. It becomes evident immediately, that the backscatter characteristics alone are only very weakly correlated with the topography and thus separation using the backscatter alone would surely be futile. This is further underlined by looking at the backscatter distributions of the delineated classes from the flight on april 8th (fig. 4.4), where the correlations are relatively average in regards to all other flights. Here it is again obvious that the backscatter characteristics are not very valuable for class separation. Thus most of the information needed to classify accurately must come from contextual data.

4.3.2 Robustness

To test the robustness of each classifier, the same steps outlined in Kortum, Singha, and Spreen, 2022 are undertaken. In brief: Using the Polarstern as an origin, a 3km x 3km region around the ship is used as the robustness test set. This area has been identified in 162 TerraSAR-X SM scenes from different days. The robustness is then defined as the probability of each pixel being classified the same as in the previous and subsequent acquisitions (time between acquisitions is typically one day). Taking into account that the surface conditions are changing over time and that Polarstern was not perfectly stationary, this approximation of the robustness will serve only as a lower bound of the actual robustness of the classifier. To be more precise, these calculations are done under the assumption that in a time period of two days, the percentage of ice that has changed class (e.g. through deformation) is significantly smaller than the percentage of ice that has remained in the same class. Note that this test is only sensible for the two solid ice classes and not for the OW/YI class, which is too dynamic on a daily timescale to be analysed in this manner. The robustness is first computed for the two classes and their average is used as an indicator for the network's robustness.

4.3.3 The Network Architectures

In this paper, five different architectures are compared: two established image classifiers in the VGG16 developed by Simonyan and Zisserman, 2015 (ice classification in e.g. Khaleghian, Ullah, Kræmer, Hughes, et al., 2021) and the ConvNext network proposed by Liu, Mao, et al., 2022 (an improvement over ResNet, used for SAR sea ice classification in e.g. Song, Li, et al., 2021), a custom CNN (cCNN) pixel-wise classifier introduced in the previous chapter (and published in Kortum, Singha, and Spreen, 2022) specifically designed for ice classification and two established segmentation models in the Unet by Ronneberger et al., 2015 (SAR sea ice classification in e.g. Nagi et al., 2021; Ren, Li, et al., 2022) and Unet++ proposed in Zhou, Siddiquee, et al., 2018; Zhou, Siddiquee, et al., 2019 (used in e.g. Murashkin and Frost, 2021). These first three (VGG16, ConvNext, cCNN) and last two (Unet, Unet++) models have one fundamental difference: Classification approaches (VGG16 etc.) are given a patch and are then asked to predict the class of the centre of the image. Segmentation approaches (Unet etc.) are tasked to produce a label for every pixel in the patch at the same time. For the Unet++ the features of the multiple output layers are averaged over in the deep supervision part of the model. The exact specifications of all the models can be found in the appendix.

4.3.4 Training

During training, the networks are tasked with minimising the Kullback Leibler Divergence (KLD) between the output and the label distributions. This allows us to fit the probabilities of each class occurring at each pixel, which one can infer from the ALS measurements. As this serves as a benchmark and comparison of these models concerning their applicability for sea ice retrieval, no further optimisations have taken place. For each of the model architectures, ten separate instances are trained. Training is stopped using the small independent validation set (5% of data). The model population allows some additional insight into the reliability of each architecture. The ingested SAR data are pre-processed by converting each band to sigma nought and then applying a logarithm. The incidence angle is provided in a third channel. The size of each patch to be classified is chosen to be 256x256 pixels, except for the cCNN which receives input patches at various scales (a 5x5, a 16x16 and a 64x64 pixel patch).

4.4 Results

The performance of different network architectures can be seen in table 4.1. They paint a clear picture of segmentation models' (Unet, Unet++) improvement over centre-pixel classification approaches. Of the centre-pixel classification approaches, the custom CNN classifier performed best, yet it was still significantly inferior to the segmentation models. Part of the reason for this is probably the high spatial resolution of the labels, as a label

Model	Mean acc. [%]	Std. of acc. [%]	Mean KLD	Std. of KLD	Mean rb. [%]	Std. of rb. [%]
VGG16	40.52	7.83	0.8493	0.0458	79.95	5.82
ConvNext	45.12	3.17	0.872	0.0363	81.16	4.84
cCNN	47.89	3.74	0.7886	**0.0240**	68.52	18.81
Unet	**68.07**	**1.74**	**0.6032**	0.0406	**84.42**	1.78
Unet++	67.92	2.13	0.6249	0.0597	82.06	**1.36**

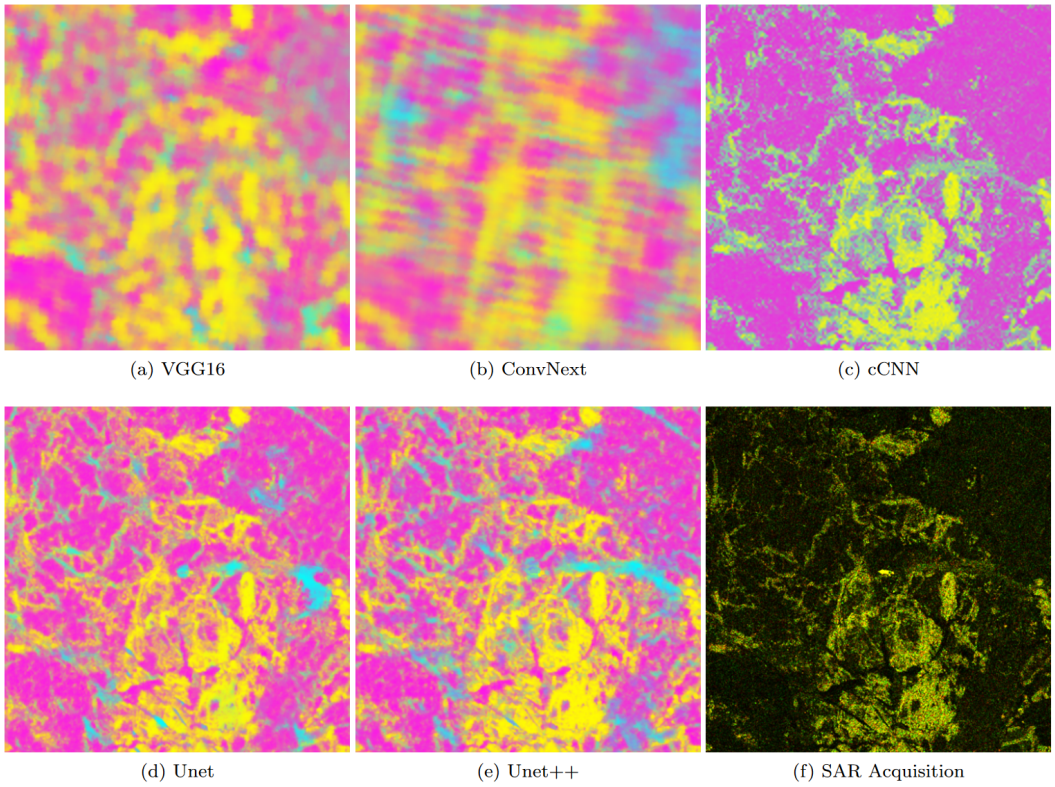
Table 4.1: Network performances on the independent test set after training. For brevity, accuracy is shortened to acc, standard deviation to std and robustness to rb. The means and standard deviations are computed from the 10 models in the population for every architecture. Best-in-category results are highlighted in bold font. Ten instances were trained for every model. The Unet and Unet++ architectures show significantly better performance than the rest.

is given for every pixel from most of the ALS measurements. The centre-pixel classifiers cannot make use of any relationships between labels, like shape, sparsity and correlations. This seems to be detrimental to their performance.

A more detailed analysis of the output of different models (Fig. 4.5) shows, how the VGG16 and ConvNext models struggle to relate all the information of the patch to only the classification of the central pixel, leading to a diffuse-looking classified scene. This seems most pronounced for the ConvNext model. A possible reason for this are the larger convolutional kernels (7x7 in contrast to 3x3) used in the architecture. The cCNN seems to struggle with using contextual data to separate rough ice and young ice. In general, the predicted probabilities at each pixel are higher in the non-dominant class, leading to a seemingly different colour palette in this visualisation. The Unet and Unet++ classifications are largely similar. Some difficulty in the separation of deformed and young ice signatures persists as can be seen in the mixing of yellow and cyan areas.

It is also worth pointing out that the very same cCNN and a VGG16 performed at accuracies around 85-95% on manual labels in the previous chapter, illustrating the difference between training and testing on quantitatively measured labels in contrast to human-generated annotations. In Ren, Li, et al., 2022, the Unet is reported to perform sea ice and open water separation on manual labels at 93-95 % accuracy. Wang and Li, 2021 report accuracies of 96 % for the same task, using ice charts as training data and test data and 94% accuraciy when comparing to an operational sea ice cover product (Interactive Multisensor Snow and Ice Mapping System, IMS). Murashkin, Spreen, et al., 2018 show classification accuracies of the Unet++ around 96% on manually labeled training and test data across 6 classes.

Whilst the mean KLD's are in accordance with the accuracies, the spread (std) of the KLD's across the model populations seems to be very similar across all models and there is no clear gap between segmentation and classification approaches. Overall, one cannot say that one model converges more reliably than another - as would be suggested by the standard deviation of the accuracies alone. It is also apparent, that the cCNN does not perform well in the robustness scores on this dataset. This model is considerably smaller than the others (in terms of parameter count) and was heavily optimised using a different



(a) VGG16 (b) ConvNext (c) cCNN

(d) Unet (e) Unet++ (f) SAR Acquisition

Figure 4.5: Comparison of classifications from different models, randomly selected from the ten instances trained. Colours are the same as the classes discussed above, but the intensity is given by the predicted probabilities, so mixed colours can occur. This can be seen most easily in the cCNN classification (c). The scene was acquired over the Polarstern (center of the images) on January 14^{th}. The false colour composition consists of HH, VV and HH/VV channels, normalised with a $tanh$ function. The area shown is a 6 by 6 kilometre square.

dataset with lower fidelity labels, which seems to have come at a cost of flexibilty/generality of the architecture. The spread of robustnesses of the segmentation models seems to be considerably smaller than those of the generative models - additionally indicating these approaches are more reliable for ice classification from SAR.

The classifications (e.g. in Fig. 4.6) show a very plausible set of results, that align with the observations of members on board the expedition. The fine labels at high resolution seem to have transcended into a similarly detailed classification map. The examples in Fig 4.7 also illustrate a general increase of deformation in the first year ice: The magenta FYI area close to Polarstern, marked by a black square in figure 4.6, is getting progressively more deformed as time progresses (detail in Fig. 4.7). The areas most prone to error seem to be the OW/YI classifications. This is to be expected as they are naturally the most sparse in the training data set. Additionally, they are very dynamic, which leads to extremely diverse backscatter properties that can be exhibited, in turn making them more difficult to classify.

We also observe decreasing correlation of backscatter and surface topography variables from the onset of the expedition until early April - particularly during January and February (4.3), where the MOSAiC expedition was met by numerous storms. Some of the decorrelation can be accounted for because of snow accumulation and redistribution, but it is difficult to quantify this phenomenon with the given data. As first-year ice thickness grows much faster than second-year ice, it makes sense that freeboard and backscatter decorrelate over time. The mechanism leading to increased backscatter in older ice were identified in chapter 3 to be processes on taking place over timescales of multiple seasons, therefore the freeboard increases faster than the change of backscatter response - this could also be the cause of this decorrelation. As was previously discussed, these processes are most likely linked to desalination in the upper layers and formation of ridges. However, since this trend is broken in April it seems that whatever drove this decorrelation is revertable to some extend and therefore changes in the snow or geometrical influences, like the incidence angle, are a more plausible explanation than ice deformation.

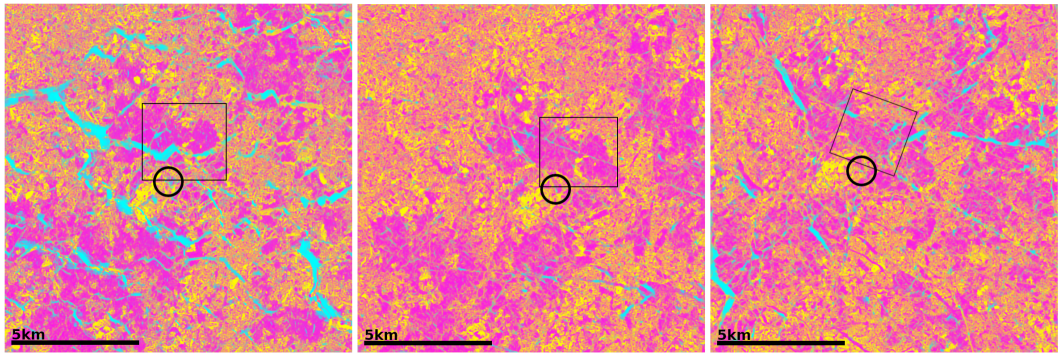
Additionally, a first-year and a second-year ice area were tracked from November to March, and the deformation increase as predicted by the top performing Unet model is measured. The results are shown in 4.8. The maximum and minimum measured area of deformed ice is compared. Here it is important that we take into account, that additional deformation cannot be resolved in areas of existing deformation. Therefore the percentage of newly gained deformation P_n can be related to the measured deformation P_m and the original existing deformation P_e via the relation

$$P_\mathrm{m} = P_\mathrm{n}(1 - P_\mathrm{e}). \tag{4.1}$$

This can be easily solved for the new deformation

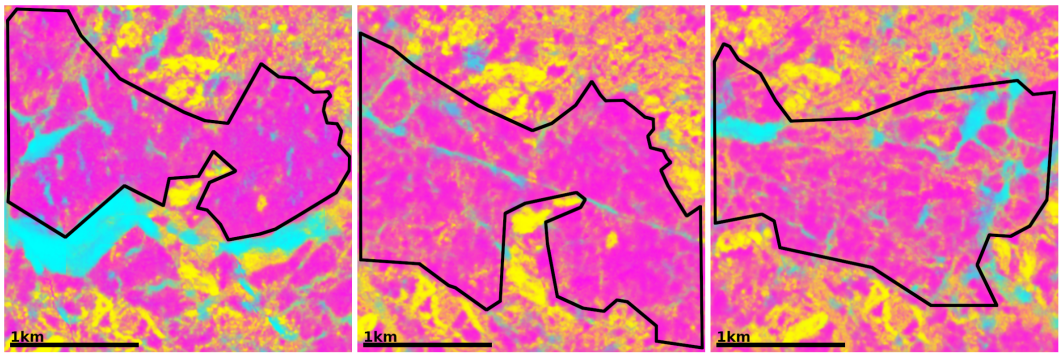
$$P_\mathrm{n} = \frac{P_\mathrm{n}}{1 - P_\mathrm{e}}. \tag{4.2}$$

Using this formula and the measured values for both observed areas, we see that the second-year ice area gained additional deformation equal to 0.186% of its area, whilst the first-year



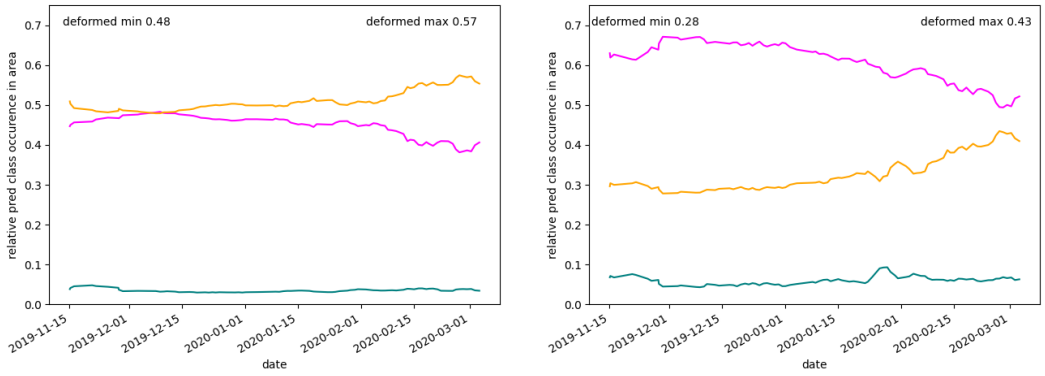
(a) 22nd of November 20219 (b) 13th of January 2020 (c) 25th of March 2020

Figure 4.6: Collection of classified subscenes (Unet, pixel spacing = 3.5m) including the MOSAiC floe, after a storm (a), in calm conditions with some shearing indications (b) and with some breakup of the ice cover visible (c). The Polarstern location is indicated by the black circle. The DFYI/MYI class probability is displayed in yellow, the LFYI probability in magenta and the OW/YI probability is cyan. The black square marks the area shown at full resolution below (Fig. 4.7)



(a) 22nd of November 2019 - DFYI: 3.1% (b) 13th of January 2020 - DFYI: 5.6% (c) 25th of March 2020 - DFYI: 13.5%

Figure 4.7: Full resolution excerpt from the scenes show in Fig. 4.6. The classified images reflect the increased deformation of the first year ice area over time accurately, as the DFYI occurence rises. The DFYI fraction is computed inside of the black border. In the first scene some misclassification of the open lead (cyan) as older, deformed ice (yellow) is seen (outside of the area the DFYI fraction is calculated) - this is a common issue in SAR sea ice retrieval as the backscatter can become very similar, as also reported by e.g. Guo et al., 2022.



(a) SYI (b) FYI

Figure 4.8: Measured deformation of two adjacent areas dominated by first- and second-year ice respectively.

ice area gained 0.244%. Assuming that the dynamic forces acting on this area were similar over the observed time period, which is reasonable, given they are adjacent to one another, suggests that first-year ice deforms faster than second-year ice.

4.5 Discussion

The top models in this investigation perform at around 68% accuracy on the test data set (Tab. 4.1). The segmentation models predictions are approximately 20% more accurate than the classification models. The only concrete difference between these models is that the segmentation approaches can learn from the distribution of labels, which appears to be highly important. Even the highest accuracies measured here are considerably lower than what many author's report for algorithms trained with human-made labels. To understand these discrepancies, the main differences between these measured labels and human annotations are discussed.

The measured labels used in this study have some underlying difficulties. Because the snow depths and densities are not known, one cannot discern how strong the correlation of freeboard and ice-thickness is and cannot eliminate this error. Also the reflectance used to disseminate young ice and open water is based purely on the coverage of the surface being snow free and thus not directly correlated with ice age: if thin ice has formed the atmospheric conditions will dictate whether or not snow has gathered on top or if the bare ice is visible to the sensor. Thus the quality of labels could still be improved on, if more information were available.

To assess the impact of the individual thresholds used to determine classes (e.g. the location of the inflection point in the freeboard distribution) the top-performing Unet architecture (accuracy ca 68%) was also evaluated on the same dataset, but excluding points near the thresholds. To do this labels where the certainty of the most probable class was

lower than 90% were not considered. For example in regions with local standard deviation of approximately 3cm, that means that points within 6cm of the thresholds are not considered and so uncertain labels do not influence the result. In case of the test dataset, these data points account for 24.1% of all data. Under these circumstances, the average accuracy of the Unet model is 72.5% which is an increase of less than 4% although 24.1% of the least certain labels where removed. Thus one can conclude that the exact location of the thresholds had only marginal impact on model performance, lending increased confidence that the model performances are representative of performance evaluated against ground truth.

For comparison with human annotation/ice charts, the resolution is a major factor. In this case, every individual pixel gets its own class and there is no semantic grouping of pixels into the same class based on proximity or likeness. This is a stark contrast to ice charts, where the labels are made up of only few polygons per scene. To test the influence that these coarser labels have on the classifier, another batch of classifiers was trained, where the labels were smoothed with a Gaussian kernel (sigma = 10px). The resulting classifier dropped to around 62.5% accuracy on average, showing that the coarseness (like that of manual labels) comes at a detriment of performance. As expected from the coarser labels, the resulting models also had a very hard time resolving smaller features like ridges and leads.

Even when not training from such ice charts, humans generating training data for algorithms at high resolution generally limit themselves to areas which they can confidently identify. Not much can be said about the correctness of these labels per se, but one should keep in mind that in these instances, the accuracy achieved by the classifier is constrained to those easy-to-identify regions and are therefore not representative of the classifier's performance on the whole. Because of the size of SAR acquistions obtaining labels at pixel resolution from human annotation is not feasible. The great advantage that labels from measurements have is that they are truly indicative of performance on the entire scope of ice conditions in the scene (every pixel is labelled, thus there is no selection bias). Only by holding the testing of high resolution retrieval algorithms to this standard can one show with certainty when an improved method of classification is developed, but of course to do so available data sets are lacking.

This study had only a small effective study region and a large temporal span to test the diverse conditions. Overall the constancy of the ice in the scenes should only improve the classifiers' performances. Unfortunately the 20 helicopter flights are not quite enough to make meaningful statements about temporal changes in performance, as the differences in performance will be outweighed by the local conditions in the scene. Additionally seasons in the data where one would expect the classification to be most difficult (freeze-up and pre early melt onset) are only very sparsely represented in the data. This means the contributions of the data sparsity, seasonality and spatial variability cannot be meaningfully separated.

In the summer season the ice surface is dominated by wet snow, bare ice and melt ponds and more open water is found between floes. The spatial distribution of classes is very

distinctive between the surface types, so one can expect the main result of the difference between centre-pixel classifiers and segmentation models to persist.

In most data-driven approaches to classification, the performance of the classifier is limited by the quality of the labels. Therefore, one should be careful when using manually labelled data, such as ice charts, as ground truth. These practices are common in the current research - as not many other sources of labels are available. However, the potential is much greater than that. The challenge of course remains, that high-resolution measurements are very sparse.

Because the MOSAiC mission provided us with an unmatched opportunity for training and testing algorithms with measured labels over a long time period, this study has made obvious that there is considerable room for improvement even with modern deep learning algorithms. It needs to be mentioned, that due to the spatial constraint to the area near the MOSAiC floe, the training dataset does not capture the full extent of possible winter ice conditions in the Arctic, thus one cannot expect the classifier to perform equally well on a pan-Arctic scale. Instances of OW/YI are very sparse and their entire span of possible conditions and consequent radar response is not covered well by data. Since a better in-situ dataset is probably not going to emerge in the near future, it is clear that measured labels alone are not enough to train a stable algorithm that can deal with the full span of ice conditions. It seems that to achieve this, one would need to leverage a great number of scenes without labels. Semi-supervised and self-supervised approaches come to mind. Some first examples of their development exist for optical data by HAN et al., 2019, ice and open water discrimination from SAR in Li et al., 2015; Khaleghian, Ullah, Kræmer, Eltoft, et al., 2021 and for sea ice classes from SAR in Imber, 2022.

4.6 Conclusion

The MOSAiC expedition enabled the generation of a large dataset (approx. 20 million data points) of SAR acquisitions and appropriate labels delineated from in-situ laser scanning measurements. It has become clear that both the freeboard and the above snow surface roughness (at lengths of 50 cm) are only weakly correlated with X-Band SAR backscatter, with average R^2 values of 0.124 and 0.043 respectively. It has been shown that deep-learning segmentation approaches such as the Unet can approximate these labels from the SAR measurement at accuracies around 68%. Thus, the performance of modern network architectures on a representative set of labels was measured for the first time. From the performances of the different models, one can conclude that the segmentation approaches advantage of being able to exploit intra-label relations is crucial (+20% accuracy) to the performance. It is notable that these label distributions at the scale of the measurement resolution are not contained in ice charts or human annotations, which suggests that classifying accurately at the resolution of the measurement when trained on human-annotated labels is improbable. As a more comprehensive dataset than created here is unlikely to be acquired in the near

future, newly developed classifiers aiming at classification at the resolution of the sensor will need to find some way to gain access to the spatial ice type distributions to be successful.

4.7 Appendix: Heli Flights

1	20191020_01_PS122-1_2-167
2	20191119_01_PS122-1_8-23
3	20191130_01_PS122-1_9-98
4	20191224_01_PS122-2_17-98
5	20191225_01_PS122-2_17-99
6	20191228_01_PS122-2_17-101
7	20200107_01_PS122-2_19-44
8	20200108_01_PS122-2_19-46
9	20200108_03_PS122-2_19-52
10	20200116_01_PS122-2_20-52
11	20200121_01_PS122-2_21-41
12	20200123_02_PS122-2_21-78
13	20200128_01_PS122-2_22-16
14	20200204_01_PS122-2_23-14
15	20200212_01_PS122-2_24-31
16	20200217_02_PS122-2_25-8
17	20200227_01_PS122-3_29-49
18	20200318_01_PS122-3_32-42
19	20200408_01_PS122-3_35-49
20	20200423_01_PS122-3_37-63

Table 4.2: List of the 20 helicopter flights used in this research. Data is published in Hutter, Hendricks, et al., 2022b.

4.8 Appendix: Architectures

Briefly the network architectures used in this investigation are presented.The following conventions are made use of to keep the figures concise. FCX is short for a fully connected layer with X neurons. ConvX x Y denotes a 2D convolutional layer with filter sizes X and number of filters Y. Unless otherwise specified the convolutional layers have stride 1. If a layer has multiple inputs, they are concatenated before being parsed to the layer.

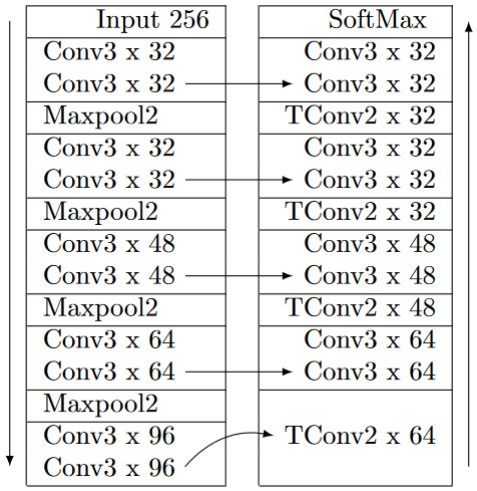


Table 4.4: The Unet architecture as used in this paper and published in Ronneberger et al., 2015. The ReLU activation is used throughout the network and the padding is set to 'same' where applicable.

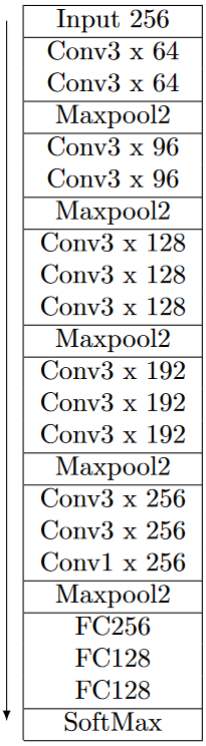


Table 4.3: VGG16 architecture as used in the paper. Published in Simonyan and Zisserman, 2015. The ReLU activation is used throughout the network. The padding is set to 'same'.

Input 256
Conv4 (stride 4) x 96
ConvNx x 96 ConvNx x 96 ConvNx x 96
Conv2 (st 2) x 96
ConvNx x 192 ConvNx x 192 ConvNx x 192
Conv2 (st 2) x 192
ConvNx x 384 ConvNx x 384 ConvNx x 384 ConvNx x 384 ConvNx x 384 ConvNx x 384 ConvNx x 384 ConvNx x 384 ConvNx x 384
Conv2 (st 2) x 384
ConvNx x 768 ConvNx x 768 ConvNx x 768
GlobalAvgPool2D LayerNorm
SoftMax

where:

ConvNx x n =

Layer	Activation
Input	
Conv7 x n	Layer Norm
Conv1 x (4 n)	GELU
Conv1 x n	
Output	

Table 4.5: The ConvNext-T architecture used in this paper. Developed in Liu, Mao, et al., 2022.

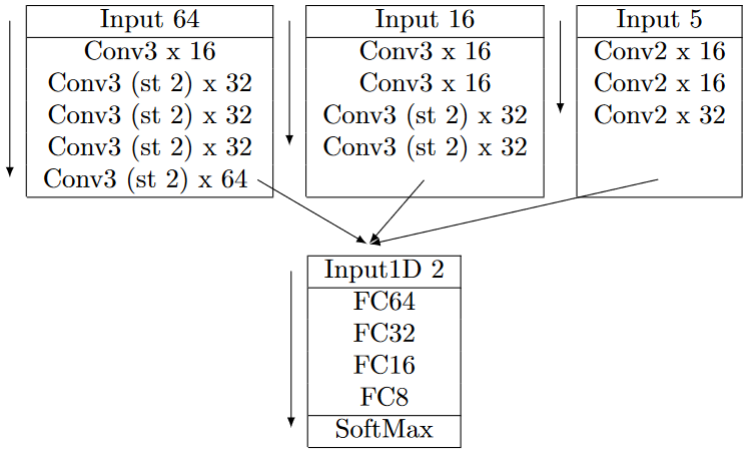


Table 4.6: The custom CNN architecture from Kortum, Singha, and Spreen, 2022 used in this paper. The inputs at different scales are flattened and concatenated before being output to the fully connected layers. Leaky ReLU is used for activation and padding is set to 'valid'. The 16x16 pixel input is downscaled from the original scene by factor 5 and the 64x64 pixel input is a square cutout that is rescaled so that the width of the entire scene is 64 pixels. The 1D input contains the relative coordinates of the pixel in the 64x64 pixel input.

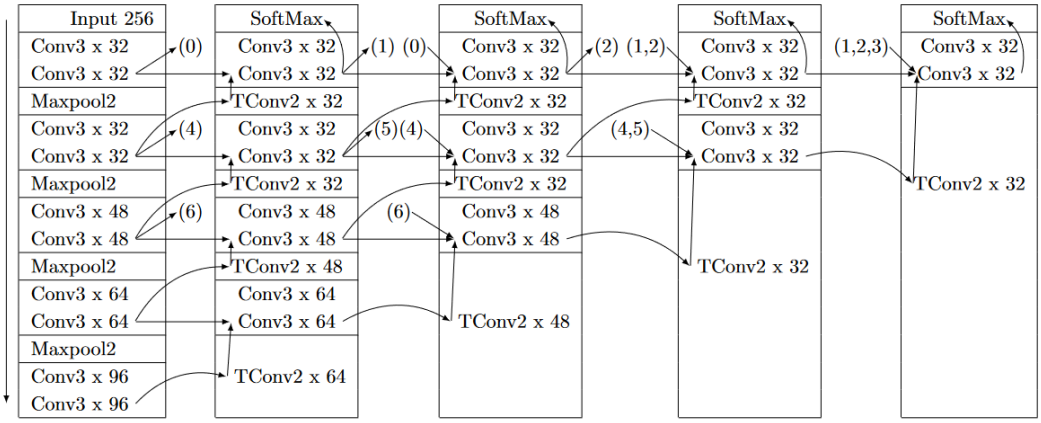


Table 4.7: Unet++ architecture used in this paper, published in Zhou, Siddiquee, et al., 2018; Zhou, Siddiquee, et al., 2019. Note that the left column is identical to the downwards convolution side of the regular Unet and the lowest rows from left to right form the upwards side of the Unet. The Unet++ then uses extra layers in between to extend the architecture. All layers within a cell are considered to be a block, so they are all executed before parsing the output to the next block. All layers marked 'Softmax' are averaged before the final linear layer and the softmax are applied. ReLU is used as the activation function throughout and the padding is set to 'same'.

Chapter 5

Learning from Unlabelled Scenes - A Novel GAN Architecture for Unlabelled Feature Extraction

5.1 Overview

The space of possible sea ice conditions is so large that labelling enough images at full resolution to fully span this distribution is seemingly impossible. Mainly, this is due to the fact that the ice conditions are often not known and cannot be discerned accurately, even by an expert human observer. Even if judging human interpretation to be good enough, the sheer amount of time it would take to label every pixel of a single 10000 x 10000 pixel scene is a challenge. Realistically, one would need to label more than 100 scenes to capture the span of Arctic ice conditions. In the previous chapter, it was shown that the capability of a network to exploit relationships between labels is critical for higher accuracy classifications. Coarser labels do not contain the label distributions and were thus show to be less informative. One idea that is explored in the following is to use the incidence angle dependent backscatter slopes as a proxy for ice type labels. Given the two properties are correlated, the spatial distribution of the two should be very similar. Under this assumption, the aim of this chapter is to derive features from SAR sea ice imagery that in part parametrise the sea ice class distribution without using any labels at all. The innovative crux of the neural network setup developed and discussed in this chapter, is that sea ice information can be extracted by predicting slopes without any prior knowledge of them. If the extracted features are of high quality, a large part of the ground truth problem could be circumvented by training the majority of a networks parameters without any ground truth at all.

5.2 Introduction

Sea ice's backscatter has been observed to exhibit different incidence angle dependence across different ice types. At C-band, for example, first-year ice (FYI) backscatter decreases faster than multiyear ice (MYI) backscatter as the incidence angle increases, whilst the inverse is true for L-Band (see work by Mahmud, Geldsetzer, et al., 2018; Zakhvatkina et al., 2013; Gill et al., 2015; Makynen et al., 2002; Liu, Guo, et al., 2015). This diverse relation of backscatter to viewing angle geometry is a result of oriented structures like surface deformation, air bubbles or brine enclosures that correlate with different stages of ice development.

As such ice types relate to a plethora of physical, chemical and biological processes of the Arctic; any property which allows us to separate them using remote sensing measurements is valuable. An existing classifier by Lohse et al., 2021 (also used in Guo et al., 2022) successfully makes use of this property by first extracting incidence angle dependence of backscatter and texture features from data labelled by humans and then relating those observations to different ice types in the classification step. In logarithmic scaling for the backscatter, the relation between σ_0 and the incidence angle is approximately linear, as was shown in Onstott, 1992. Thus, this property of incidence angle dependence of backscatter can be expressed as a single number, the slope, which is constant across the incidence angle range of Sentinel-1 scenes.

The ground truth problem in sea ice remote sensing manifests itself in suboptimal

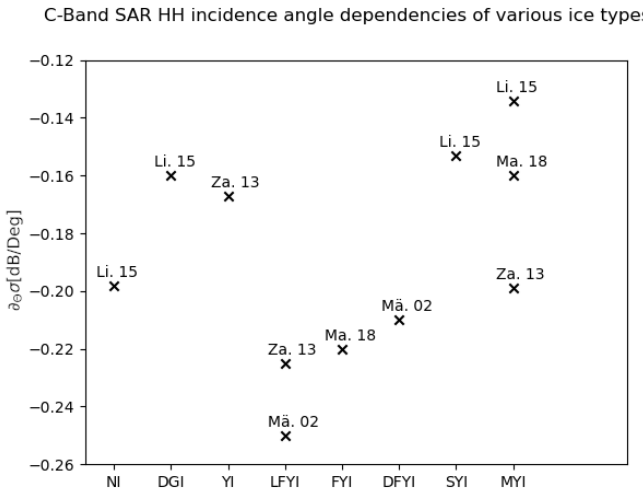


Figure 5.1: Measurements of sea ice backscatter (HH) slopes relative to incidence angle change during freeze-up or winter from C-Band satellite SAR up to 2018. The y-axis corresponds to the backscatter slope and the x-axis denotes the ice types as given in the publications: NI - Nilas, DGI - Deformed Grey Ice, YI - Young Ice, LFYI - Level First Year Ice, FYI -First Year Ice, DFYI - Deformed First Year Ice, SYI - Second Year Ice, MYI - Multi-Year Ice. The labels are the first two letters of the name of the first author of the paper where the results were originally published Makynen et al., 2002; Zakhvatkina et al., 2013; Liu, Guo, et al., 2015; Mahmud, Geldsetzer, et al., 2018 and the final two digits of the year of publishing (all in the 21st century). For a review of the observations up to 2017, see Mäkynen and Karvonen, 2017.

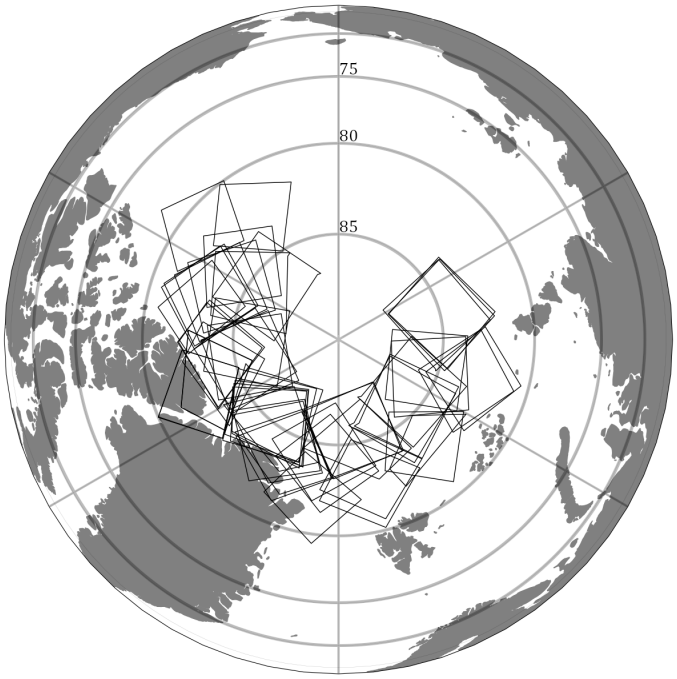


Figure 5.2: Outlines of the Sentinel-1 EW (Extended Wide) Scenes used for this research. The zero meridian is marked by the vertical line. The map is projected using a vertical near-side perspective of an observer at an altitude of 1500 kilometres directly over the north pole.

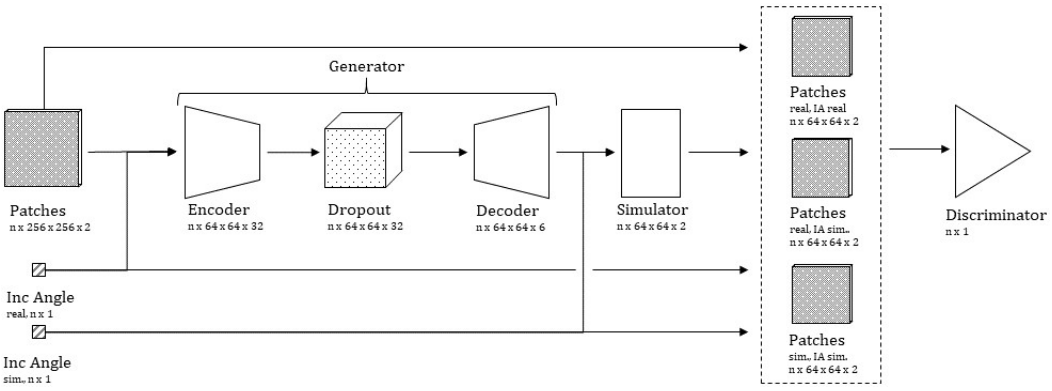


Figure 5.3: The devised network architecture. The arrows denote a forward pass.

datasets, which translate their shortcomings to data driven algorithms trained on them: Having a dataset that does not cover the entire span of possible ice conditions results in an algorithm that does not perform well on scenes with conditions different from those trained or modelled on (like the checkerboard example shown in the theory section). Similarly, low-resolution training labels will, in turn, translate to a classifier that struggles to produce high-resolution classifications. The previous two chapters showed that only because an algorithm performs well at extrapolating coarse manual labels, does not mean the same architecture is particularly good for predicting high resolution labels. Thus, it is very difficult to train a classifier or retrieval algorithm that performs both for a large span of ice conditions and at high resolution. Localised incidence angle dependence near the resolution of the instrument can help bridge this challenge, as it correlates with different ice classes (fig. 5.1).

The issue with localised incidence angle dependence is that it also can't be easily obtained from observations. To do so, one historically had to make repeated observations of the same ice at different incidence angles. In work by Mahmud, Nandan, et al., 2020, this had to be done by hand - identifying the same patches of ice in different acquisitions or multiple occurrences of the same type across different incidence angles. The alternative would be to match observations using geolocation. At Sentinel-1 resolution, these observations would need to be made within minutes due to the speed of ice drift. Such pairs of scenes are rare, even from the Sentinel-1 constellation Mäkynen and Karvonen, 2017. A frequent coverage of such pairs would require a very special constellation of satellites, which, at least until now, does not exist.

One approach to predict the incidence angle dependent slopes developed by Cristea et al., 2020 leverages the linear evolution of the brightnesses with incidence angle of similar ice using a Gaussian mixture clustering that varies the cluster's means with incidence angle. This technique uses global image information of brightness evolution without any spatial or contextual awareness. The underlying assumption is that the same type of ice is present across a range of different incidence angles, which is not always the case. Also the results of the algorithm are discrete clusters, whilst the real dependence comes from a continuous space. For the same reason, mixing of different ice types in one pixel cannot be considered either. Additionally, noise from the antenna pattern also is a function of the range and cannot be fully corrected for - which can influence the clustering in an unwanted way. For high-resolution observations, the incidence angle span is usually very small, meaning this approach would have trouble in these situations. These troubles are probably not insurmountable, but there could be merit in predicting the incidence angle dependence from local data only.

This chapter deals with an approach to predict the incidence angle dependencies from only local image patches. As these correlate with ice development, this forces a neural network to implicitly learn about ice development without additional labels. The hope is to use the incidence angle prediction task to produce features valuable for ice type classification

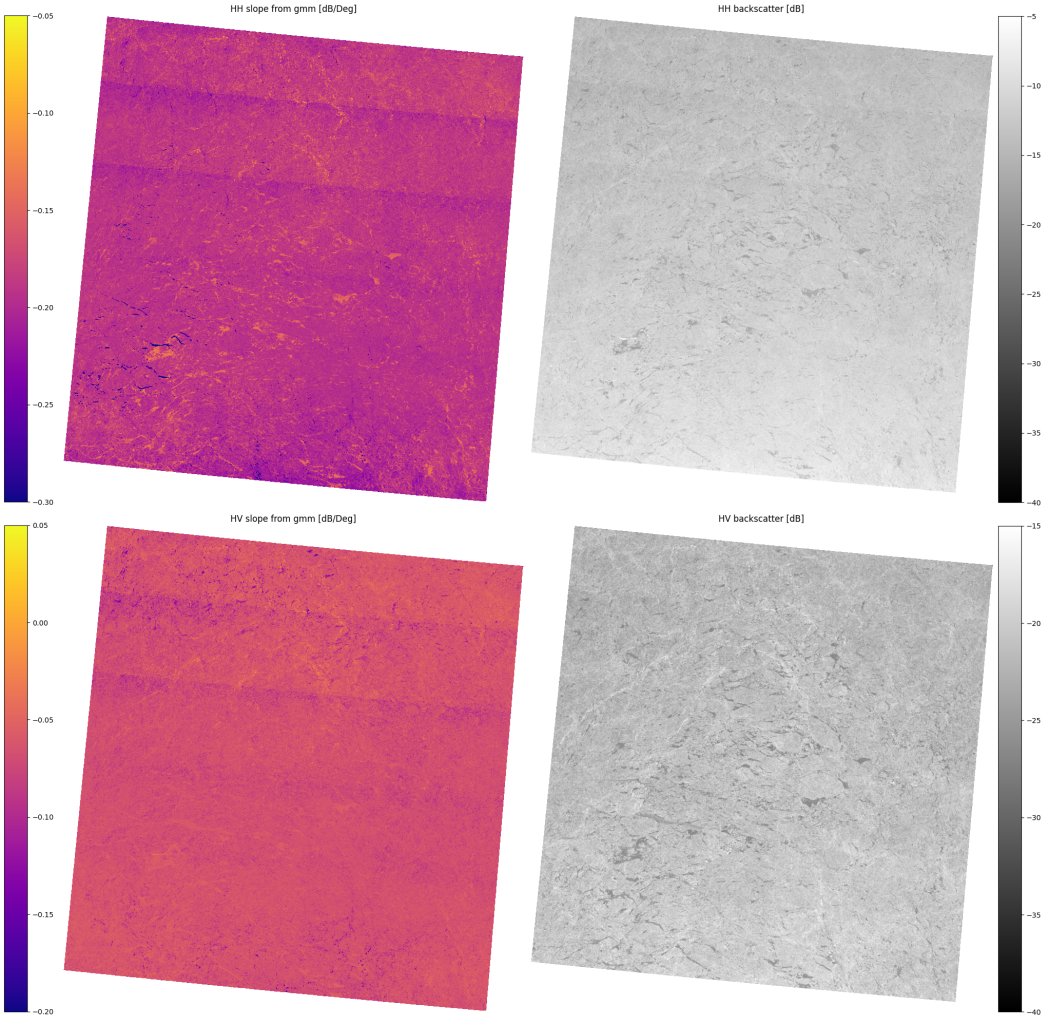


Figure 5.4: Incidence angle dependent slopes from a Sentinel-1 EW scene of the central Arctic, calculated using a Gaussian mixture model according to work published by Cristea et al., 2020. The scene is from the third of November 2018.

in an intermediate encoding layer (output of the encoder in figure 5.3) . This has the potential to open up a big data approach to ice retrieval, where one can continuously improve ice retrieval models with nothing but SAR data. Also, it offers more insight into the spatial variability due to the slopes' continuous parameter space.

To achieve this vision, a machine learning setup (see figure 5.3) is introduced to predict the incidence angle behaviour from a single patch of a SAR scene, incorporating spatial information compression with an encoding and decoding scheme. The network is best described as a physics-informed sparse conditional Generative Adversarial Network (GAN) (for related work seeGoodfellow, Pouget-Abadie, et al., 2014; Mirza and Osindero, 2014; Zhou, Gao, et al., 2019). The network is trained using only the SAR data with its included incidence angle information to predict the incidence angle dependence of each pixel in a SAR scene without additional measurement or prior knowledge of the slopes. To evaluate the predictions of the network 59 near-coincident laser altimeter tracks from the ICESat-2 mission are used. The derived freeboard elevation of ICESat-2's Level 3A data product ATL10, explained in Kwok, Markus, et al., 2019 and Kwok, Petty, et al., 2022, is used to differentiate first-year ice (FYI) and multiyear ice (MYI) using a dynamic threshold. Using these labels, one can then measure the average incidence angle dependence of that ice type and compare those to the predictions.

In configurations where algorithms are trained on human-labelled data, the algorithm can never be expected to significantly outperform humans. Whilst some label noise can be compensated for, systematic biases arising from human annotation (like tendencies not to label small features etc.) are very difficult to overcome. Simply put, a data driven algorithm cannot learn from information that is not present in the training data. The machine learning setup in this chapter performs proficiently at tasks that humans would certainly not be capable of. In that aspect, it represents a different approach to the problem: machine learning algorithms do not attempt to emulate human behaviour. Instead, the problem is reformulated in a way where the algorithms can exceed human performance.

Given the measurements made by shown in figure 5.1 for HH incidence angle dependence, very young ice like nilas seems to have a similar incidence angle dependence to first-year ice, which is mostly lower than that of multi-year, grey and young ice. Based on this rough estimation, knowing the incidence angle dependence restricts the plausible ice classes considerably. If independent correlations of the same complexity also exist for the HV band (as the algorithm will later suggest), this could be used to further narrow down the possible ice class. The potential of being able to solve a considerable amount of the classification problem near sensor resolution and without any labels (thus applicable to all scenes and in principle, scaleable to all possible ice conditions, as no additional data except the SAR scene is needed) sounds intriguing. These classification approaches, where part of the data is used without labels on an adjacent task, are commonly termed self-supervised learning and have shown success in a variety of tasks where ground truth is sparse (consider Liu, Zhang, et al., 2021 for an overview). Whilst the slopes could be directly useful for classification, the

internally learned features are probably the real treasures of the method,a s they parametrise the spatial distribution of both backscatter and slopes, which should be strongly related to ice types.

5.3 Data

In this section, a brief overview of the data products used in this study is given.

The primary source of data is Sentinel-1 SAR in EW (Extended Wide) mode. This product consists of approximately 400km by 400km scenes with a pixel size of 40 metres. This study uses the Ground Range Detected (GRD) product, which gives pixels projected to sea level (in the case of sea ice), using an earth ellipsoid model. The scenes have an incidence angle range of 20 to 50 degrees. The scenes have been corrected for thermal noise as well as scalloping and calibrated to σ_0 using the snap library *SNAP - ESA Sentinel Application Platform* 2022 and corrections developed by the Nansen Center and detailed in Park, Korosov, Babiker, Sandven, et al., 2018; Park, Won, et al., 2019; Korosov et al., 2022. This significantly mitigates the effect of sensor artefacts on the study. The second data product used is the ATL-10 sea ice freeboard measurement derived from altimetry measurements of ICESat-2. ICESat-2 operates using a 532nm laser at 10,000 pulses a second. The returned photons are used to produce altimetry measurements accurate to approximately 3cm, by aggregating a fixed number of photons to measurement. The ATL-10 product has variable spacing - depending on the photon returns. The footprint of a single pulse is around 17 metres, as published in Neumann et al., 2019. Do obtain a freeboard height, the absolute heights measured are corrected for the ocean surface elevation by identifying leads through their reflectivity characteristics. The open leads can then serve as a reference height to correct the freeboard. To minimise errors, only the three strong beams from ICESat-2 are used in this study. The SAR data consists of 59 pairs of Sentinel-1 imagery (fig. 5.2) with an ICESat-2 overflight within 10 minutes of the SAR measurement and considerable overlap between both measurements. The SAR data is used to train the model and the altimetry measurements only serve to validate the findings. The validation is done by identifying various regions of ice using their measured freeboard. The ice types to be extracted are first-year ice (FYI) and multi-year ice (MYI) (second-year ice is included in the MYI class). For that reason, the study time is constrained to October and November. During this part of the ice development cycle the freeboard differences between these two ice types are substantial and easily observable from altimetry data: FYI is still young and thin with a significantly smaller freeboard than the MYI that survived the summer. Later in the season, the ice thickness of the FYI will catch up and get closer to the MYI thickness because of the higher thermodynamic ice growth rates of thinner compared to thicker ice. As the snow depths on the floes is unknown, and thicknesses vary with region and time, using constant thresholding on the freeboards to separate classes is not feasible. Instead, those scenes are used where the altimetry freeboard distribution is clearly bimodal – allowing for the separation of the

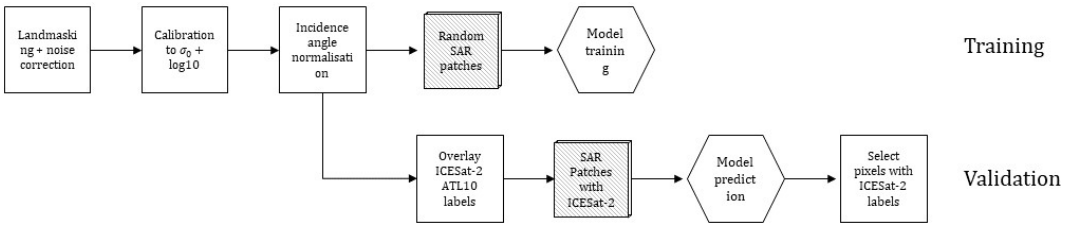


Figure 5.5: Depiction of the flow of data during training and validation of the network.

two different stages of ice development from the freeboard alone. An example of overlapping measurements of ICESat-2 and Sentinel-1 and the bimodal freeboard distribution can be seen in figure 5.6. The thresholds between FYI and MYI are then simply selected to lie at the local minimum between the two distributions. Due to the variable snow cover and subpixel resolution structure, the two distributions are partially overlapping. Thus, some false mixing of classes cannot be avoided. Because the scenes are dominated by MYI this mixing will be more present in the FYI class (relative to total class size). The lowest 5 cm of freeboard measurement for the FYI class are disregarded: ice in the early stages of development has significantly different backscatter properties and would thus skew the distribution.

5.4 Methodology

In this section, the exact method and the various network's used as a part of it are described in detail. The phrase 'incidence angle normalised' used in this section denotes that SAR acquisitions have been corrected so that the average backscatter across all scenes is approximately constant across all incidence angles. After logarithmic scaling a linear transformation is sufficient to achieve this (eq. 5.3). This one transformation is calculated from all scenes and then applied to them. Thus, all the variations in brightness across the incidence angle range come from the differences in incidence angle dependencies between classes, not the total average brightness decay.

5.4.1 Motivating Existence

As was mentioned in the introduction, the incidence angle dependence of different sea ice types can only be measured through multiple observations. The question 'is it possible to predict the incidence angle slope for every pixel from a single patch of a SAR image?', has no obvious answer. To motivate the existence of such a prediction, let us first look at a simplified version of this task by posing the following problem. Given an incidence angle normalised SAR acquisition over sea ice (or a patch thereof) and an incidence angle, is it possible to tell if the measurement was made at the given incidence angle or not? If the answer to that question were yes that would imply some understanding of what different regions (ice types) are expected to look like at certain incidence angles. Because the image

has been normalised for the overall trend first, the mean brightness of the patch is not a useful feature. Only the relative brightnesses of the ice types are useful features. The discriminating task can be described as follows:

Given a dataset of SAR patches of sea ice and the incidence angles they were acquired at (patch, θ), as well as a set of patches with incidence angles they were not acquired at (patch, $\neg\theta$), does there exist a discriminative function D, such that

$$\begin{aligned} D &: (patch, \theta) \mapsto 1, \\ D &: (patch, \neg\theta) \mapsto 0? \end{aligned} \tag{5.1}$$

Solving this task is thus only possible if the algorithm can recognise certain ice and knows how bright it should be at different incidence angles. If an algorithm is able to learn the relationship between brightness and ice class implicitly, it should then also be possible to predict these relationships explicitly.

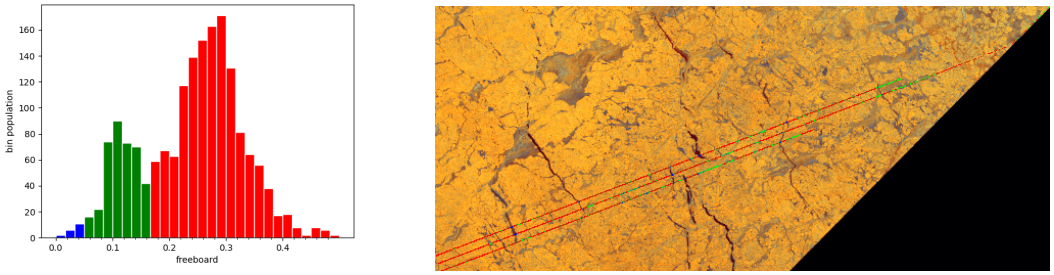
It is worth noting that this task of distinguishing between patches with correct and incorrect incidence angles could never be done by humans at high accuracy. In fact, human ice analysts will find the discrimination between the two pairs very challenging, as the differences between the ice classes slopes are small compared to the overall trend (average trend across classes $\approx$ 0.2 dB/Deg, difference between first-year and multiyear ice $\approx$ 0.04 dB/Deg). Machine learning algorithms, however, excel at this problem: Such discriminators (with incidence angle normalisation in place) achieve accuracies surpassing 98%, whe the false incidence angles $\neg\theta$ are off by at least 5° def. Note that it is imperative here that the normalisation of the incidence angles is carried out first. Otherwise, the discriminator could complete this task easily by simply learning the average brightness trend and having little knowledge of the ice class-specific incidence angle dependencies.

To predict the incidence angle slopes using such a discriminator, a second network (the generator) is added that estimates incidence angles dependencies (slopes) for each pixel in a given patch. Using these slopes, one can then calculate what the patch would look at an arbitrary incidence angle. If these newly generated patches fool the discriminator - making it judge they are real - the probability is high that the estimated slopes are close to reality. This is the core idea of the network setup.

5.4.2 The Network

We have motivated the existence of a prediction algorithm and also introduced the concept of a discriminator that can reliably separate patches with correct and incorrect incidence angles. To turn the implicit understanding of the discriminator to an explicit prediction of incidence angle dependence, a generative adversarial network (GAN) setup (fig. 5.3) is used. Before diving into detail, it is sensible to outline the core idea of the architecture. The core of the concept consists of two networks with competing tasks:

The generator network is given a patch of a SAR acquisition and a single incidence angle it was acquired at (centre of the patch) and then is tasked with predicting a linear function



(a) Freeboard histogram

(b) ICESat-2 Labels overlayed with a Sentinel-1 scene

Figure 5.6: Freeboard distribution from ICESat-2 (a) and derived labels overlayed on top of a false colour composition (channels are HV, HH-HV, HH/HV) from Sentinel-1 SAR(b). Blue labels correspond to open water and young ice, green to FYI and red to MYI. . The data were acquired on the 2nd of November 2021, within 10 minutes of one another.

(i.e. a slope and an intercept) for the backscatter for each pixel depending on the incidence angle. Using these predicted functions, one can then easily simulate what this patch would look like at any arbitrary incidence angle.

The discriminator is given three types of patch and incidence angle pairs: Observed patches with correct incidence angles, observed patches with incorrect incidence angles and simulated patches with the incidence angles they were simulated at. It is tasked with separating the real patches with real incidence angles from the rest. This is a binary classification task conditional on the incidence angle.

During the training of these networks, the generator is tasked with fooling the discriminator into thinking the patches simulated using the predicted slopes are real. This way, the two networks are in constant competition, facilitating continuous improvement. With this setup, it is possible to reformulate the task of predicting incidence angle dependencies of the pixels to a domain transfer task, in which GAN setups excel (e.g. Lin et al., 2018; Lu et al., 2017; Wei et al., 2017). Note also that this is only feasible because the relationship of backscatter and incidence angle is known to be linear for a given ice type; without this known physical constraint, the task would become significantly more difficult.

In the following the three components of the setup (generator, simulator, discriminator) are described in detail.

The Generator

The generator gets an input patch with three channels: HH, HV and θ_{IA}. Its output is a patch with six channels that parametrise the pixels backscatter based on incidence angle. These parameters are the slope, intercept and a noise parameter for each of the bands (see section 5.4.2 for details). This physics-constrained linear behaviour of the patch under change of incidence angle significantly simplifies the task to solve.

We first tried to directly predict the output with a simple Unet model (Ronneberger

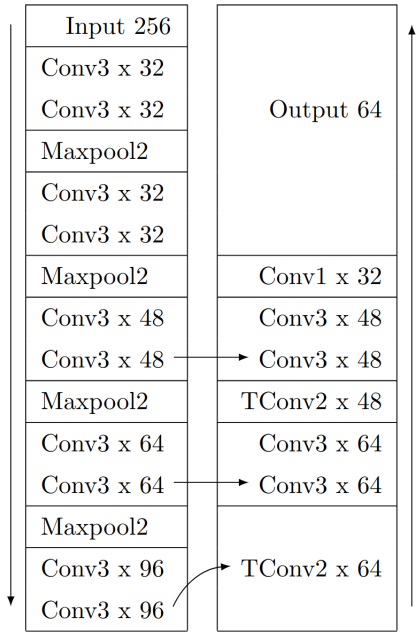


Table 5.1: The Unet architecture used in this paper as the encoder. The input has dimensions 256x256x3 and the output has dimensions 64x64x32. The architecture was originally developed and published in Ronneberger et al., 2015. The ReLU activation is used throughout the network and the padding is set to 'same' for all layers.

Input 64
Conv3 x 32
TConv3 x 32
Conv3 x 32
TConv3 x 32
Conv3 x 32
TConv3 x 32
Conv3 x 24
TConv3 x 24
Conv3 x 16
TConv3 x 16
Conv3 x 12
TConv3 x 12
Conv3 x 8
Conv1 x 6

Table 5.2: The decoder architecture used in the paper. The ReLU activation is used throughout the network. The padding is set to 'same'. The final Conv1 layer has no activation function - it is just a linear pixelwise mapping.

et al., 2015) as a generator. However, the GAN failed to reliably converge during training, as the generator quickly collapsed to a mode where it would simply map the backscatter channels to the constants of the linear function and set the slopes to zero. Thus, the addition of an encoding task also served to circumnavigate this collapse of the generator, forcing the network to make use of spatial information. The Unet (encoder), shown in table 5.1) is now tasked with generating a 32-dimensional feature vector for each pixel. Then the majority of these feature vectors are randomly dropped out, and only 1/32 in the patch are kept. From this sparse representation, a second network (decoder), shown in table 5.2, made up of consecutive convolutions and transpose convolutions, then reconstructs the desired output channels from the sparse representation. This solved the aforementioned training collapse. The same idea of compression using dropout has been shown to be effective in biomedical imaging by Zhou, Gao, et al., 2019. The output of the generator is a 6-dimensional vector for each pixel. The exact parameters of the Unet and the decoder as used in this model, can be found in the appendix. The number of convolutions and transpose convolutions are chosen so that the theoretical perceptive field of an encoded vector is 27 x 27 pixels, which is large enough to reliably decode images without gaps due to too many randomly dropped out encoded vectors. (The probability of not finding another encoded vector in a 27 x 27 patch at a dropout rate of 31/32 is smaller than 10^{-10}). The reason random dropout is used instead of a bottleneck with lower dimensions is that this way, every pixel gets its own local parametrization.

The Simulator

The simulator network consists of a single non-trainable layer that simply takes the six parameters $(a_{\mathrm{HH}}, a_{\mathrm{HV}}, b_{\mathrm{HH}}, b_{\mathrm{HV}}, \Delta_{\mathrm{HH}}, \Delta_{\mathrm{HV}})$ predicted by the generator and an incidence angle θ, to simulate a new patch by evaluating two linear equations of the form

$$\sigma = a \cdot \theta_{\mathrm{ia}}^{\mathrm{sim}} + b + \mathcal{N}(0, \Delta) \tag{5.2}$$

for both channels, where $\mathcal{N}(0, \Delta)$ denotes sampling from a normal distribution. Four parameters are the intercepts a and slopes b of the two linear functions modelling the incidence angle behaviour of the HH and HV bands. The remaining two values Δ parameterise the uncertainty of the pixel backscatter, which is implemented as a random sample of a normal distribution. It is useful to formulate this simulation as a network layer to integrate it into the training loop for the GAN. This layer constrains the network to model the observed physical behaviour of backscatter under incidence angle change.

The Discriminator

The discriminator architecture used is a ConvNext-T architecture that was developed by Liu, Mao, et al., 2022 and tested as a classifier in chapter 4. The architecture is shown in table 4.5. The discriminator is given a patch with three layers (HH, HV and θ_{IA}) and performs a binary classification, separating real patches with correct incidence angles from all other

combinations (real patches with wrong incidence angles and simulated patches with any incidence angles). As the result is conditional on the incidence angle, such a discriminator is called a conditional discriminator as per Mirza and Osindero, 2014.

5.4.3 Training

In this section, the training loop and the corresponding loss function are explained in detail. A graphic representation of the forward propagation through the network is shown in fig. 5.3. The data used during training consists of 59 Sentinel-1 scenes from the Arctic acquired in October or November. The scenes were chosen for the existence of an ICESat-2 overflight with ATL10 ice freeboard observations in the same area within ten minutes. This will become relevant for the evaluation of the algorithm at a later stage. For each batch during training the samples are randomly selected from all scenes in the dataset at runtime. A sample consists of a 256x256 pixel patch with two channels (HH, VV), which is denoted as p^{real}, and the incidence angle at the centre of the patch $\theta_{\text{ia}}^{\text{real}}$. In addition, a second set of incidence angles $\theta_{\text{ia}}^{\text{sim}}$ is generated by copying $\theta_{\text{ia}}^{\text{real}}$ and then randomly offsetting half of them by at least 20% of the incidence angle range of the sensor (approximately 6 degrees in the case of Sentinel-1). The other half of $\theta_{\text{ia}}^{\text{sim}}$ remains equal to $\theta_{\text{ia}}^{\text{real}}$.

The patches' HH and HV channels are then normalised for the incidence angle dependence using a linear approximation of the average patch backscatter $\overline{\sigma_0}(\theta_{\text{IA}})$. To scale the pixel backscatter to lie mostly between 0 and 1, the standard deviation σ_Δ of the means after subtracting the linear approximation is used. This is purely to stabilise the training of the network and has no physical reasoning. The normalised pixel values σ_{norm} are then computed from the σ_0 values as

$$\sigma_{\text{norm}}(\theta_{\text{IA}})\,[dB] = 0.5 + \frac{\sigma_0 - \overline{\sigma_0}(\theta_{\text{IA}})}{12\sigma_\Delta}. \tag{5.3}$$

Thus, the mean backscatter of each patch is 0.5 on average. The denominator of $12\sigma_\Delta$ ensured that less than 1% of individual pixels were outside of the $[0, 1]$ interval.

The correct incidence angles are then concatenated with the patches as an extra channel, which prepares the input for the generator. The generator returns a patch with six channels. The first four contain the parameters $a_{\text{HH}}, a_{\text{HV}}, b_{\text{HH}}, b_{\text{HV}}$ for both linear functions $\sigma_{\text{norm}}^{\text{HH}}, \sigma_{\text{norm}}^{\text{HV}}$ that describes that pixels' HH and HV backscatter depending on incidence angle. The final two outputs are standard deviations Δ of both bands that parametrise the Gaussian random noise added to the pixel after calculating the backscatter to allow for some statistical simulation.

The simulator S then creates new patches using the output of the decoder, calculating the simulated pixel backscatter as

$$\sigma_{\text{norm}}\,[dB] = a \cdot \theta_{\text{ia}}^{\text{sim}} + b + \mathcal{N}(0, \Delta) \tag{5.4}$$

for each pixel and band. Here $\mathcal{N}$ denotes the normal distribution. This constitutes one pass through the generator and simulator. The resulting simulated patches from the generator

G can be described by:

$$p^{\text{sim}} = S(G(p^{\text{real}}, \theta_{\text{ia}}^{\text{real}}) \oplus \theta_{\text{ia}}^{\text{sim}}), \tag{5.5}$$

Where $\oplus$ denotes the concatenation of the patch, and the incidence angle expanded to the image dimension. The next step is the evaluation of the discriminator's performance on real and simulated patches. The incidence angles $\theta_{\text{ia}}^{\text{sim}}$ are added as an extra layer to the simulated patches by concatenation to prepare the input of the discriminator D. The resulting discriminator output d^{sim} is thus

$$d^{\text{sim}} = D(p^{\text{sim}} \oplus \theta_{\text{ia}}^{\text{sim}}) \tag{5.6}$$

The discriminator is also given some real patches with correct and incorrect incidence angles ($\theta_{\text{ia}}^{\text{sim}}$), generating output

$$d^{\text{real}} = D(p^{\text{real}} \oplus \theta_{\text{ia}}^{\text{sim}}). \tag{5.7}$$

The first loss term of the generator is a pixel-to-pixel loss, which minimises the distance between the real patch and the patch simulated at the same incidence angle:

$$L_{\text{G}_{\text{px2px}}} = L_{\text{MSE}}(p^{\text{real}}, p^{\text{sim}})|_{\theta_{\text{ia}}^{\text{real}} = \theta_{\text{ia}}^{\text{sim}}}, \tag{5.8}$$

where L_{MSE} is the mean squared error. Note this is only computed for the patches where the set of modified incidence angles $\theta_{\text{ia}}^{\text{sim}}$ are unchanged from the originals $\theta_{\text{ia}}^{\text{real}}$. This loss term thus encourages that patches simulated at the original incidence angle and should be the same as the real patch. The second part of the generator loss is a binary cross entropy term L_{BC}. For a set of predictions $\hat{Y}$ and corresponding labels Y it is defined as

$$L_{\text{BC}}(Y, \hat{Y}) = \sum_{\hat{y} \in \hat{Y}, y \in Y} y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}). \tag{5.9}$$

In the case of the generator, it describes the objective of fooling the discriminator into thinking the simulated patches are real

$$L_{\text{G}_{\text{disc}}} = L_{\text{BC}}(\mathbb{1}, d^{\text{sim}}). \tag{5.10}$$

The total loss of the generator is simply the sum with some coefficients c_{i}.

$$L_{\text{G}} = c_0 L_{\text{G}_{\text{px2px}}} + c_1 L_{\text{G}_{\text{disc}}} \tag{5.11}$$

The discriminator loss is made up of two components. The first describes the discrimination between real patches with real incidence angles and real patches with wrong incidence angles:

$$L_{\text{D}_{\text{real}}} = L_{\text{BC}}(\mathbb{1}|_{\theta_{\text{ia}}^{\text{real}} == \theta_{\text{ia}}^{\text{sim}}}, d^{\text{real}}), \tag{5.12}$$

where $\mathbb{1}|_{\theta_{\text{ia}}^{\text{real}} = \theta_{\text{ia}}^{\text{sim}}}$ denotes a vector which is equal to 1 where $\theta_{\text{ia}}^{\text{real}} == \theta_{\text{ia}}^{\text{sim}}$ and 0 otherwise. This means the patches with real θ_{ia} should be mapped to 1 and those with offset θ_{ia} to 0 to

minimise the loss. The second discriminator loss term monitors the success of recognising the simulated patches (mapping those to 0), again using a binary cross-entropy loss.

$$L_{\mathrm{D_{sim}}} = L_{\mathrm{BC}}(0, d_{\mathrm{sim}}). \tag{5.13}$$

The total loss for the discriminator is given by the sum

$$L_{\mathrm{D}} = c_2 L_{\mathrm{D_{real}}} + c_3 L_{\mathrm{D_{sim}}}. \tag{5.14}$$

The coefficients used for training are as follows:

$$c_0 = 50, c_1 = 0.1, c_2 = 0.1, c3 = 0.1, \tag{5.15}$$

Finally, some reasoning for the key hyper-parameters is given here. At first glance, it might seem that the pixel-to-pixel loss weighted with c_0 is far more important than the rest. However, this is not true, as the loss term is a mean pixel-wise squared error, which returns small values (on the order of 10^{-3}) for images scaled to pixel brightnesses that lie mostly in the interval $[0, 1]$. Note that they were initially chosen heuristically and then iterated on empirically to obtain a reliably converging network. The model was trained with the commonly used Adam optimiser from Kingma and Ba, 2014. The learning rate was set to 10^{-6}. Low learning rates help stabilise adversarial training. The patch size of 256 x 256 pixels is a compromise of contextual information and computing power, but it is also worth noting that the task of the discriminator - which was already the more powerful network in this setup - is made simpler with larger patches. This imbalance could pose additional problems at larger patch sizes and is one of the reasons why the incidence angle extraction is more difficult at full resolution. The output patch size is 64x64, a quarter of the input resolution.

The described architecture was trained in approximately 10^7 samples, taking 24 hours on a NVIDIA RTX 2060 SUPER graphics card (specifications can be found in Nvidia, n.d.), using the tensorflow, numpy and scipy libraries (see Martıń Abadi et al., 2015; Harris et al., 2020; Virtanen et al., 2020 for implementation). Evaluation of a full scene takes approximately 5 minutes.

5.5 Evaluation

Setting Expectations

Before jumping to the results of the classifier, setting some expectations for the performance seems reasonable, outlining what this algorithm can and cannot achieve. The GAN method improves the generator's prediction by fooling the discriminator. As the discriminator only sees the newly generated scenes, not the dependencies themselves, there are some implications for the results. Firstly, the spatial resolution of the extracted incidence angle dependencies is most probably not going to be at the resolution of the SAR data. As the gradients only have to be close enough to bring the backscatter into a reasonable range for

the newly generated patches, the edges between different ice types and other small-scale features are most likely not going to be predicted accurately, as they are not large enough to be useful for discrimination. For the same reason, one should expect the gradients to be a little smaller (in magnitude) than the measurements: they just need to be in a believable range to be realistic. As sea ice naturally has a broader spread of intensities, even for ice of the same class, it is enough to fool the discriminator with gradients a little smaller than observed in nature.

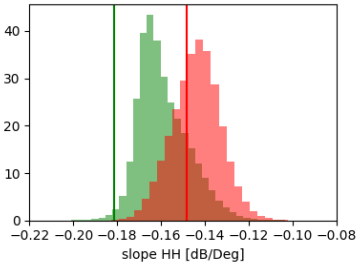
Apart from discussing these shortcomings of the method, one should also define what would be considered a success: The distribution of predicted incidence angle dependencies for FYI and MYI should be clearly different from one another (at least in HH, where this has been observed previously) and most importantly the predictions should be consistent with previous measurements and consistent across ice of the same type (identified visually). For example, it is expected that the first-year ice slopes are of greater magnitude than those of multiyear ice in HH. From a qualitative perspective, it also means one should be able to clearly see the separation between younger and older when looking at images of the in the gradients.

Results

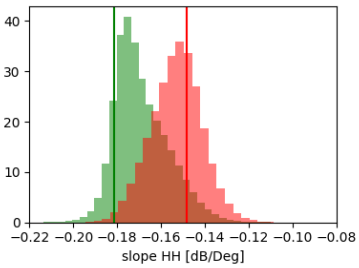
After training the above-described architecture, the results do not immediately show if the predicted incidence angle behaviours of the generator are in accordance with measurements. Although one cannot measure the incidence angle behaviour directly for each pixel, it is possible to calculate it for an ensemble of pixels of the same ice type. To get such ice types, the near coincident ICESat-2 measurements mentioned earlier in the paper are used.

The ICESat-2 ATL10 measurements are spaced by approximately $13m$, as also given by Neumann et al., 2019, resulting in 3-4 measurements for one 40x40m Sentinel-1 pixel. Patches of 4x4 pixels are grouped by downsampling the Sentinel-1 data, giving roughly 15 ICESat-2 measurements per group for increased certainty of the measurements. Then, it is possible to use the mean of these measurements to do the selection of classes. To measure the incidence angle dependence of each class, the Sentinel-1 backscatter σ_{norm} at each pixel belonging to that class is placed into one of twenty bins in the incidence angle range. The mean backscatters in the bins are then fitted with a linear function, giving a measured slope (indicated as vertical lines in fig. 5.7).

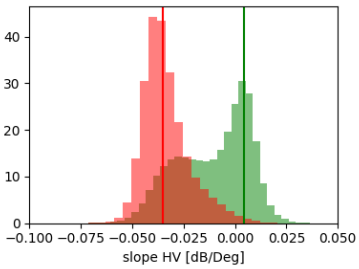
Comparing the slopes measured using the ICESat-2 data with previously derived incidence angle slopes for varying ice types (fig. 5.1), the MYI slope lies in the expected domain, whilst the FYI slope is smaller than in previous studies. Part of the reason might come from looking at the earlier freeze-up season, whilst most of the measurements compared with are from winter. Another reason could be that the ice grouped in this class is selected by freeboard thresholding alone. Natural ice diversity and varying snow cover will result in other ice types also lying in this freeboard range and thus the classes are not pure. FYI is reported to have the smallest incidence angle slope (fig. 5.1). Thus the accidental inclusion



(a) HH slopes uncalibrated



(b) HH slopes calibrated



(c) HV slopes

Figure 5.7: Histograms depicting the model's predicted slopes for FYI (green) and MYI (red). Values measured using the coincident ICESat-2 data are shown as vertical lines.

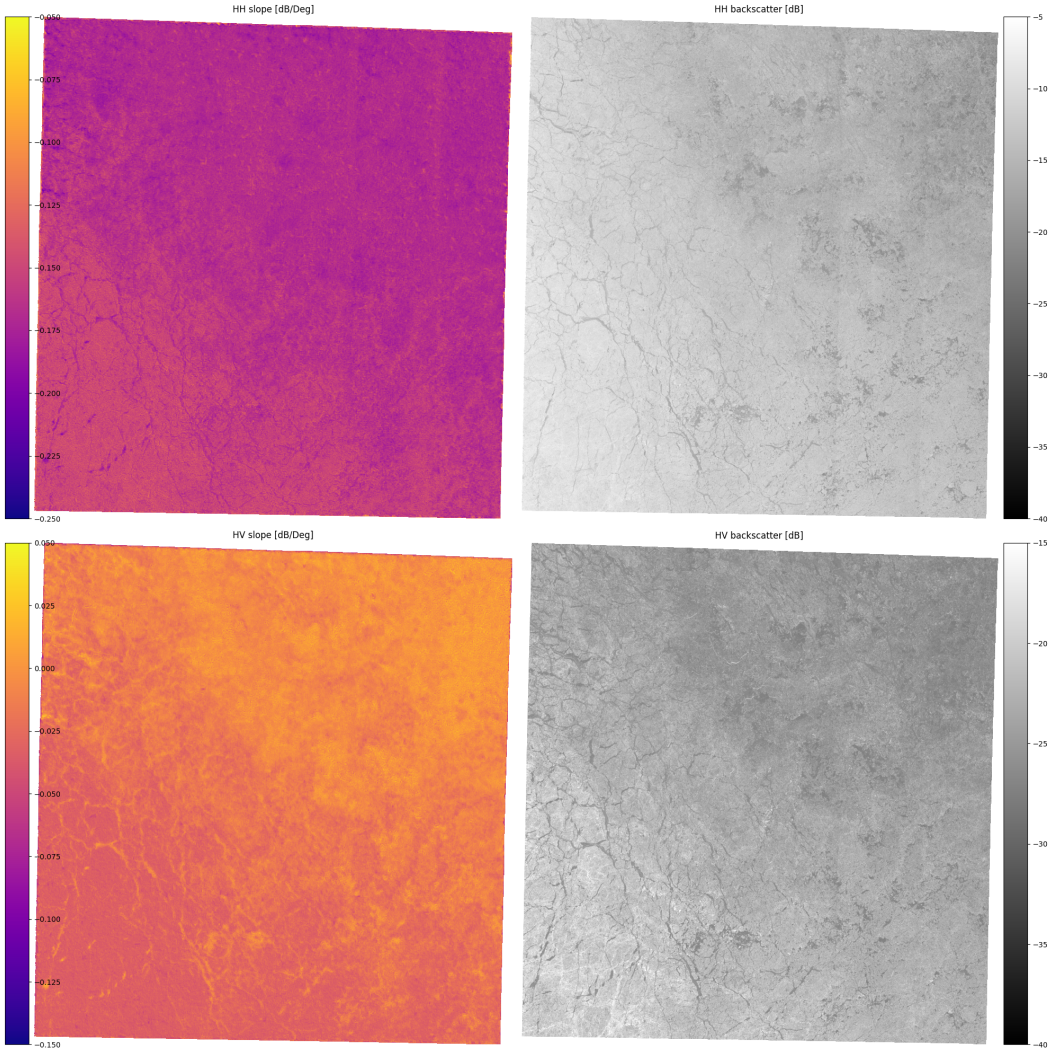


Figure 5.8: Extracted slopes (left column) for a ICESat-1 EW acquisition from the 31. October 2021 over the Arctic sea ice and the original HH and HV channels (right column) in dB. The HH slopes are depicted between -0.25 to -0.05 and the HV slopes between -0.15 and 0.05.

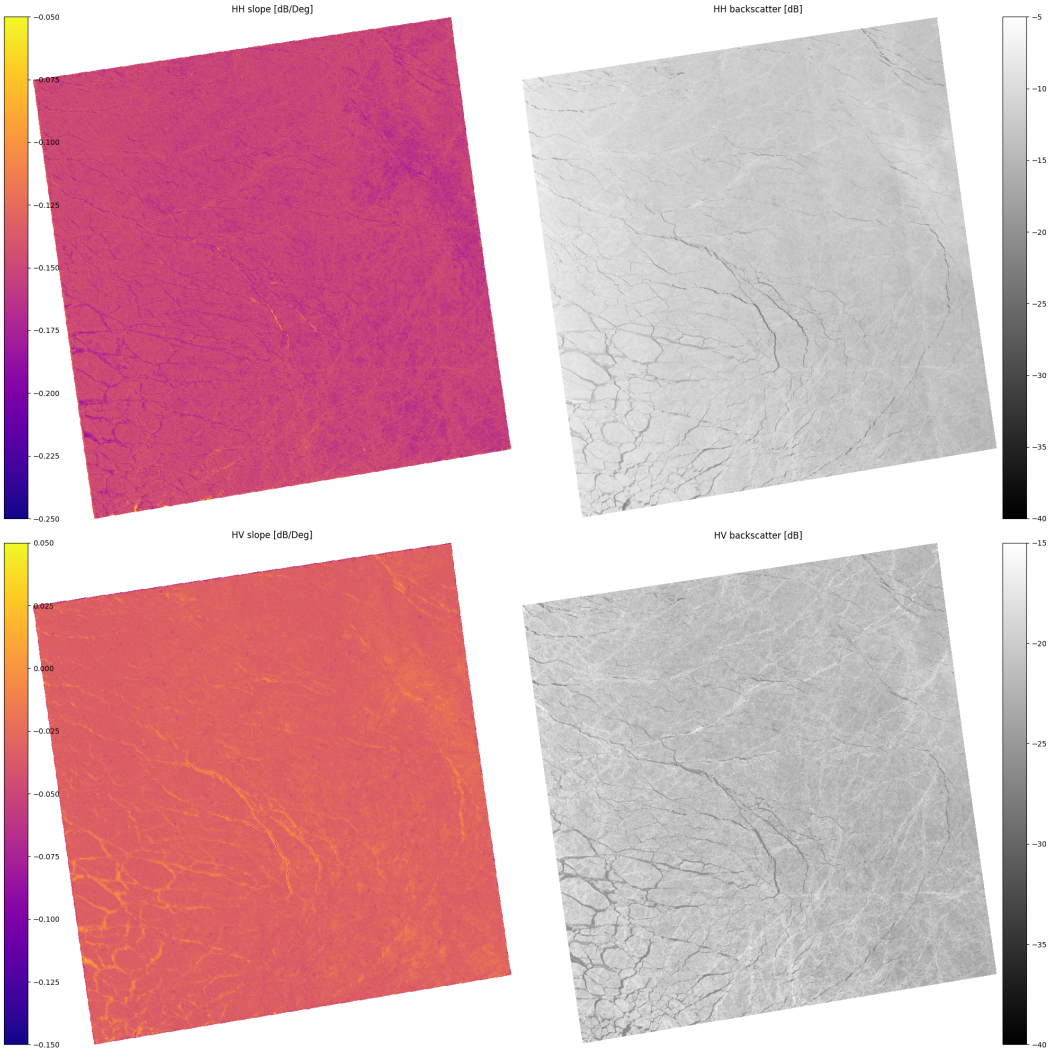


Figure 5.9: Extracted slopes (left column) for a Sentinel-1 EW acquisition from the 10. October 2019 over the Arctic sea ice and the original HH and HV channels (right column) in dB. The HH slopes are depicted between -0.25 to -0.05 and the HV slopes between -0.15 and 0.05.

of other ice in what was here defined as FYI could be a reason for the values measured here. For example, the lower threshold might not reliably keep out young ice with high incidence angle dependencies. It is worth noting that this has no bearing on the trained network.

To calculate the model predictions, the generator is run on each patch, which contains labelled pixels and gathers the parameters from the output. These parameters are then plotted in a histogram and compared those with the slopes measured using the ICESat-2 derived classes (fig. 5.7 a, and c).

In figure 5.7, one can see that there is a separation between slopes predicted from first-year ice and multiyear ice, similar to the ICESat-2 measurements. The measured values are marked as vertical lines. From the separation of the distributions, one can conclude that the algorithm has managed to successfully distinguish FYI and MYI incidence angle dependencies. In the HH band, a general underestimation of the slopes (by approximately 6.5%) is seen, which aligns with the expectations set for the algorithm earlier. To compensate for underestimation, the results are calibrated by multiplication with a constant. This was chosen so that it minimises the squared distance between the modes of the predicted distributions and the measurements from ICESat-2. The calibrated results are also shown in 5.7.

As was previously mentioned, some mixing of the classes will have occurred, which skews the distributions slightly. Remarkably, the FYI distribution in the HV band has split into one part partially overlapping with and one part clearly separated from the MYI distribution. This reinforces the notion of increased class mixing happening here. The results for the HV band also suggest that the slopes here are just as valuable to separate ice types as the HH slopes.

5.6 Demonstration

In this section, an example of extracted incidence angle dependencies from a Sentinel-1 scene and a clustering analysis for improved visualisation of the results is shown. Then the method is compared with the global clustering based method from Cristea et al., 2020, and the use of the encoded intermediate features for ice classification is explored on an example basis.

The following analysis uses two of the 59 scenes in the dataset, which are representative of the performance of the classifier and exemplify both the strengths and difficulties of the approach. Figure 5.8 shows the extracted slopes for a Sentinel-1 scene from the 31. October 2021, which shows an area where FYI (upper right) meets older MYI (lower left). Some FYI in leads can also be identified in the MYI region. Visually it is quite clear that the extracted parameters correlate spatially with different stages of ice development: As expected from previous measurements the younger ice has lower HH slopes than the older ice. The leads are also well picked out by the slopes. In a scene from the 10th of October 2019 (fig. 5.9) on can see that leads in the older ice are easily picked out by the classifier. An added feature are some leads which show higher HH slope. Past measurements have shown increased slopes

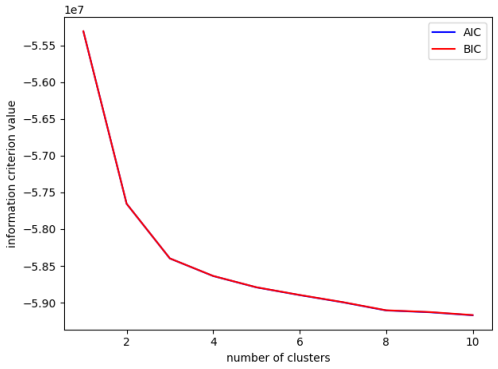


Figure 5.10: Information criteria scores (y-axis) of GMMs fitted to the extracted parameter space for various amounts of clusters (x-axis). Note that the BIC (red) and AIC (blue) are largely overlapping.

for young ice (fig. 5.1), although not quite as large as these predicted here. Nonetheless, it is promising that even differentiation between sparse young ice types might be possible with this approach.

Having extracted the incidence angles without any prior knowledge of different ice types, a reasonable next step could be to cluster the extracted parameter space of slopes, intercepts and noise. To do so a Gaussian mixture model (GMM) as in Reynolds, 2009, implemented the scipy library by Virtanen et al., 2020, is used. As the number of clusters that are present in the data is not known, first the number of clusters is varied continually while observing the model performance using Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (fig. 5.10). This is a common practice approach when fitting clustering models (e.g. Fraley and Raftery, 2002). As both criteria are continuously decreasing, the analysis suggests that the data extracted does not originate from clusters that are separable in this feature space (or not from clusters at all, but from continuous distributions). As a compromise, a GMM using six clusters is used in the following.

After selecting and training a GMM, the coincident ICESat-2 data are used to qualitatively relate the clusters to the ice stage of development for visualisation purposes. To achieve this, all patches with ICESat-2 derived labels are evaluated, giving six parameters for each pixel (slopes, intercepts and noise values), to be used for clustering. Then the clusters are sorted according to average freeboard, where cluster 1 has the lowest and cluster 6 has the highest. Mind that not all ice types mentioned in fig 5.1 are separable by freeboard, as other properties can play a role. The results of clustering the parameter space for the scene from figure 5.9 can be seen in figure 5.11. One can see that the lowest classes correspond to young ice areas, whilst the highest class is likely related to ice deformation features such as ridges. The classes in between cover a spectrum from level first-year to multiyear ice. Overall the results show off strong correlations between different ice classes and the feature space that is clustered. However, a small area in the right of the scene seems to have some results that don't quite align with expectations, as the ice predicted to be of cluster 3

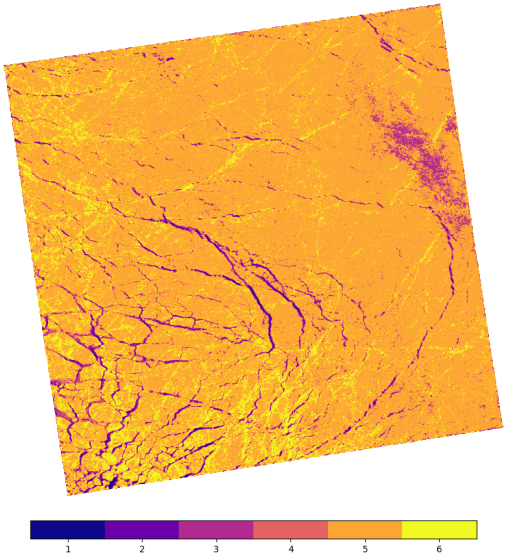


Figure 5.11: Clustered parameter space of the example scene shown in figure 5.9 with clusters sorted by increase in freeboard according to ICESat-2 data.

looks very similar in the satellite imagery to that belonging to cluster 5. This scene is quite representative of all the 59 clustered scenes - largely, the clustering looks like a reasonable classification, but there are some regions that do not quite align with expectations. This could have a multitude of reasons and is a common difficulty of SAR sea ice classifications. Visually the incidence angle dependence clustering seems to have a comparable quality to existing ice classification methods without having used any labels for training.

5.7 Comparison with GMM

So far, the results of the GAN setup have looked promising. To truly gauge the quality of the incidence angle dependent slopes, it is compared with a gaussian mixture model (GMM) clustering based approach by Cristea et al., 2020 for all scenes. It is immediately apparent that the two ways of measuring give different results, with the global clustering method qualitatively looking significantly better than the local GAN based results, with higher contrast and sharper outlines between classes. The GAN based approach does seem to be better at handling the swath edges, however. An example of this is shown in figure 5.12.

Given this result, using the GAN based technique to extract incidence angle dependences should not be the preferred option. In the cases which are difficult for the clustering technique, as outlined in the introduction, it is probably still better to adapt the clustering based method than to use the GAN method. However, the original aim of using the incidence angle dependencies was not purely to predict them directly but also to serve as a proxy label for ice types. Apart from the naive clustering of output variables above, one can also investigate

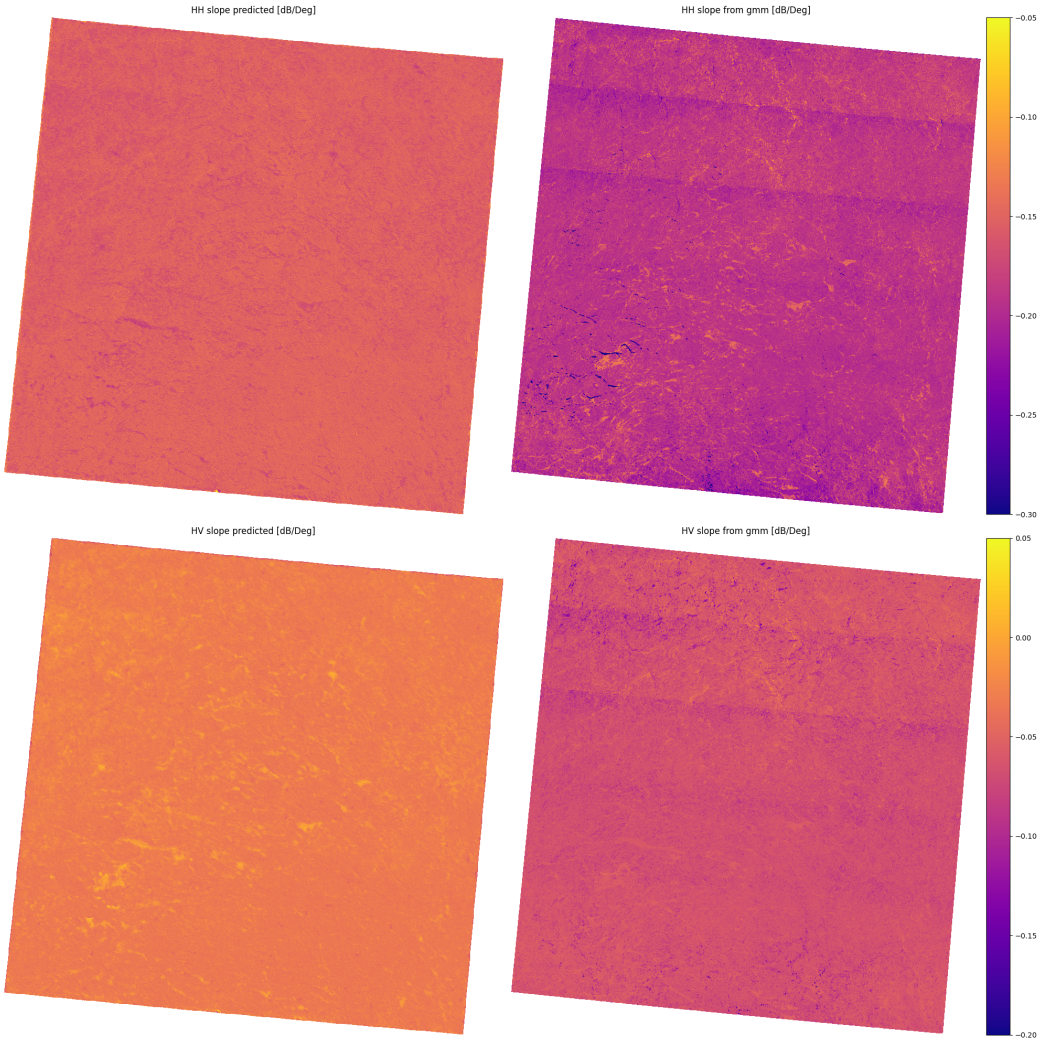
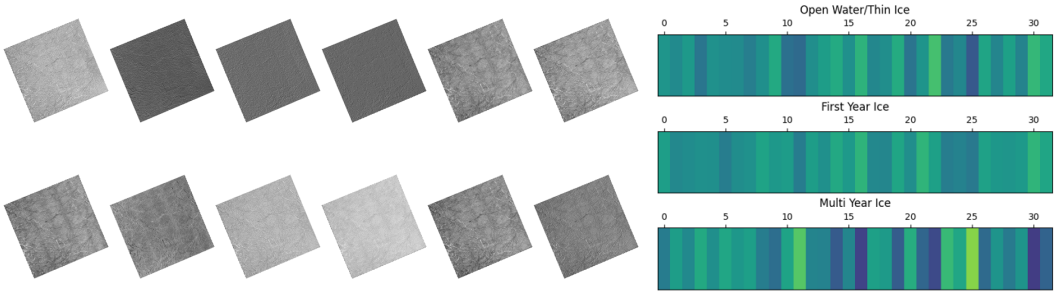


Figure 5.12: Incidence angle dependent slopes from a Sentinel-1 EW scene of the central Arctic. Once predicted by the GAN setup introduced here (l) and once calculated using a Gaussian mixture model (right) according to work published by Cristea et al., 2020. The scene is from the third of November 2018.



(a) 12/32 features visualised spatially

(b) average feature vectors

Figure 5.13: Visualisation of the feature space extracted by the GAN setup. The features space is 32 dimensional. 12 of these features are shown in a. for an Sentinel-1 scene. The average feature vector for three ice classes - as given by the ICESat-2 data - are shown in b.

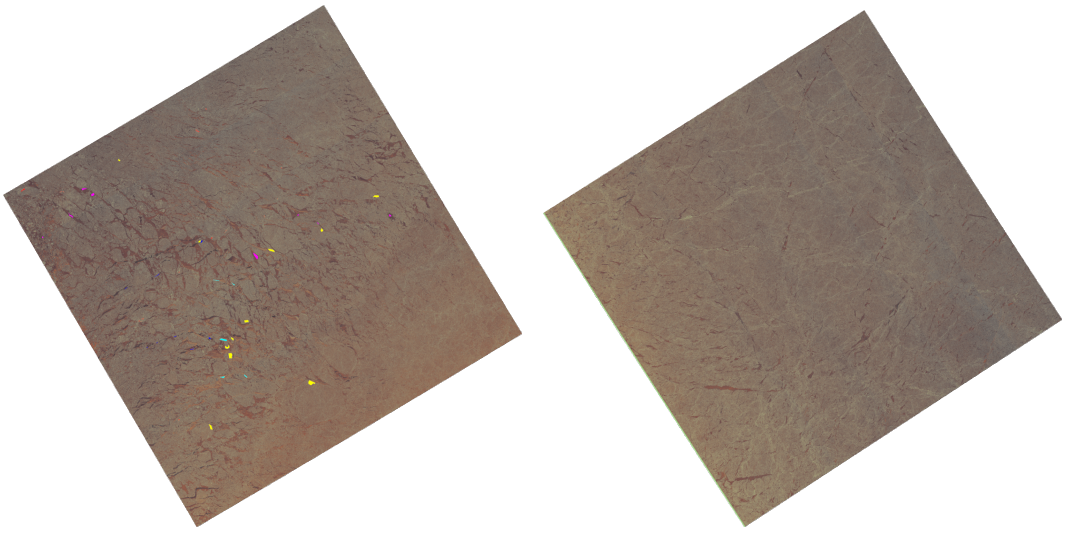
the intermediate encodings, which parametrise both spatial and incidence angle dependence information. The information contained in the encoded space is 32 dimensional per pixel. Quantifying the information content in these 32 features is a not an easy task. Visualisation of a subset of these features in 5.13a shows that they are diverse and have spatial complexity similar to that of the original image. To understand how these features relate to different ice classes, the average feature vector for OW/YI, FYI and MYI classes are computed as suggested by the ICESat-2 ATL-10 data in figure 5.13b. This visualisation clearly shows how the features do vary by ice type and gives the impression that features valuable to ice type classification have been learned by the network without any supervision.

A quick experiment is conducted as per the usefulness of these features for classification. Five of the 59 scenes in this study are coarsely labelled manually, with 5 ice classes spanning open water, young ice, first-year ice, multi-year ice and heavily ridged multi-year ice. An example of a scene with labels is shown in figure 5.14a. Using these manual labels, a simple dense network with one hidden layer was trained to predict the labels from the 32 dimensional features space as inputs. This is compared with a Unet, which is trained using the SAR data as an input instead. Because the encoder that generated the 32 dimensional features space is also a Unet, the architecture of both of these models is essentially the same; only one has been pre-trained in the GAN setup, and the other has not.

The results of the comparison in figure 5.14c versus figure 5.14d speak for themselves. The Unet has no chance to extrapolate the sparse manual labels in a sensible way, strongly overfitting to the training samples. In contrast, the model learning from the feature space manages to extrapolate the labels and gives reasonable results on unseen data.

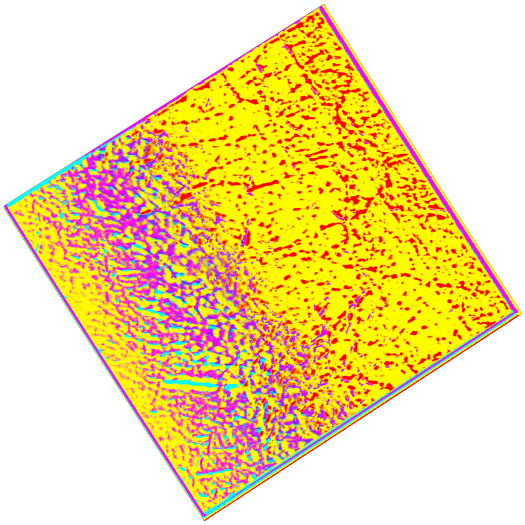
5.8 Discussion

In light of the expectations set earlier, the incidence angle predictor is performing at a reasonable level: There is a clear separation between the classes, and the predicted values

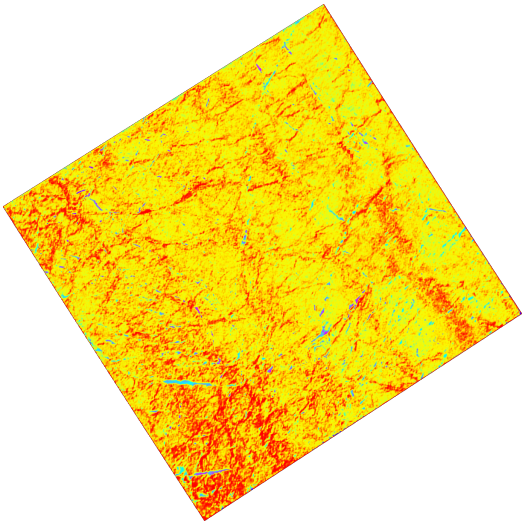


(a) manual labelled scene

(b) test scene



(c) unet classification



(d) GAN classification

Figure 5.14: Examples from a quick comparison demonstrating the use of the derived features, as seen in figure 5.13. Subfigure a. shows manual labels from five classes overlayed with a false colour composition of a SAR scene. The classes are heavily deformed ice (red), MYI (yellow), FYI (magenta), YI (teal) and open water (blue) The RGB channels are made up of HH, HV, and HH/HV. In subfigure b., the test scene is shown with the same colour composition. In c., the Unet classification is shown for the test scene, and in d., the classification that used the GAN-derived features is given.

match up with measurements using ICESat-2 altimetry. Spatial features such as leads or young ice inclusions are clearly visible in extracted slopes, as are transitions between older and younger ice areas. The predictor also suggests diverse incidence angle gradients across ice types in the HV polarisation, and the derived feature space is sensible to ice types. As dependencies from similar HV measurements are not readily available, it will be interesting to see how the derived gradients compare with future observations. High gradients in young ice as observed in Liu, Guo, et al., 2015 are also predicted by the classifier. Although these young ice occurrences are rather rare in the scenes used here, the classifier has still learned about the diversity of the gradients of young ice. in principle, this sensitivity, even to less common phenomena, is a promising sign that even sparsely present ice types are picked up by both discriminator and generator. The results for FYI and MYI are validated by coincident ICESat-2 measurements, but they only can confirm that the average trend is correct. A comparison with the clustering based method by Cristea et al., 2020 shows that the extracted incidence angles seem of lower quality both in the spatial resolution and in the precision of the extraction sense.

Whilst the calibration using ICESat-2 could be improved with some more detailed ice type maps with more different ice types and less class mixing, this would not bring up the quality enough to match the other method, as the calibration can only affect the absolute values, but not the spatial aspect or relative contrast between ice types.

The other aim of the algorithm is not to use the incidence angles explicitly but instead to use the outputs of the encoder in the middle of the network for sea ice classification. The idea behind this is that to be able to predict the incidence angle dependence of ice from only local data, one would also need to implicitly learn about the ice type, or at least a variety of variables, which are also useful for ice type classification. The common approach to make use of the symbiotic nature of these two tasks is to classify from the encoded layer, as was briefly demonstrated above. The approach discussed here opens an avenue to leverage the vast amounts of satellite scenes captured to learn about the diverse states of sea ice without additional data. A promising result of this setup is that the network's performance is not limited by human inputs. This marks a step in the direction of sea ice property retrieval from SAR with capabilities surpassing those of a human observer. Because it can be trained with as many scenes as available, this should make the predictions very robust.

Unsupervised ice-water discrimination from incidence angle dependence has been demonstrated in Cristea et al., 2020, making use of the significantly steeper incidence angle slope of water in the HH channel. Whilst studies from Mäkynen, Kern, et al., 2014 have shown little correlation with melt pond fraction and the radar backscatter, open water and melt ponds have significantly different incidence angle dependence compared to sea ice. Thus, this technology might open up a way to predict meltpond fractions from SAR, as the incidence angle dependence should be correlated with the melt pond fraction.

5.9 Conclusion

In this chapter, the possibilities for extracting information important to sea ice classification were explored, using incidence angle dependencies as a proxy label for ice types. The research shows that the slopes calculated from local data only cannot match the quality of those calculated from global clustering. However, the features calculated in the encoding part of the network prove very useful for ice classification, especially for sparse labels. In summary, the work in this chapter demonstrates how incidence angle dependencies can be used as a proxy for ice class labels in a fruitful manner. Combined with spatial encoding, a deep learning setup can learn features which are very useful for ice classification from entirely unlabelled data, greatly reducing the need for large amounts of ground truth data that has typically held back machine learning approaches in sea ice classification. So far this has only been demonstrated for coarse manual labels and in an example-based manner. Also, the encoding was only at a quarter of the resolution of the input scene. If the same holds true for high-resolution labels and the full resolution of the input data has not yet been determined.

Chapter 6

Extrapolating Sparse Measurements - From ICESat-2 Topography to Sentinel-1 SAR

6.1 Overview

The last chapter showed some first results, that incidence angle dependencies might serve as valuable proxies for ice type labels. However, the incidence angle dependent slopes predicted by the GAN setup were not of as high quality as those extracted using a linearly varying Gaussian mixture model. This suggested that the intermediate features that were extracted, were also suboptimal. It also was not yet clear if the features are of high enough fidelity to be useful for high resolution classification. In this chapter, directly using the slopes extracted using the global clustering technique as proxy labels is explored to extrapolate the ice development as derived from ICESat-2 freeboard measurements. The resulting method is an architecture agnostic approach that deals with the ground truth problem of sea ice remote sensing. When using it, one only needs a small amount of ground truth data, whilst the vast majority ($> 99.99\%$) of network parameters are trained on unlabelled data. A paper containing most of the work presented here has been submitted to a peer-reviewed journal and is currently under review. This paper is first-authored by the author of this thesis, who conducted the research presented in the paper and in this chapter.

6.2 Introduction

One approach to monitor sea ice that has been discussed at length in the previous chapters, are synthetic aperture radar (SAR) instruments, capable of delivering year-round backscatter measurements that are sensitive to changes in the ice cover, yet more difficult to evaluate than optical satellite due to the diverse backscattering properties that sea ice admits in all its different stages of development. An alternative approach, which was briefly introduced to generate validation data in the previous chapter, is the use of altimeters, which measure the distance to the ground in nadir with footprint sizes on the order of tens of meters. Altimeters are very accurate and give precise information on the distance of the satellite to the scatterer on the ground. If leads in the ice open up and can be detected, this distance can then be converted to a freeboard measurement. This total height of the ice and snow above the water is typically very indicative of the development of the ice underneath. The big drawback of the altimetry measurement is that it is spatially very sparse in the transversal direction of the flight path. Essentially, the measurement is only along thin lines over the Arctic. Measurements from multiple flights can in principle be combined to give a more detailed perspective. Because of the dynamic nature of sea ice, however, it is not so easy to spatially relate multiple overflights, and resulting products are typically constrained to a more regional scale, where drift can be ignored.

In some sense, the SAR and altimetry measurements are complementary: SAR has great two dimensional coverage, but it is not easy to convert the collected data to information concerning the development of the ice. Altimetry, on the other hand, has very limited spatial coverage, but the variable it measures is very precise, easy to interpret and gives concrete information about the underlying ice. An existing method by Karvonen, Rinne, et

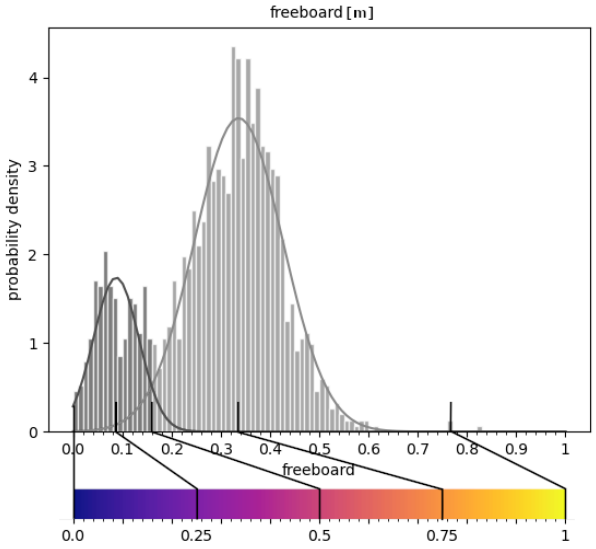


Figure 6.1: Visualisation of how the freeboard distribution (top) from an ATL-10 track is converted to an ice development index (bottom).

al., 2022 combines Sentinel-1 and Cryosat-2 measurements of ice thickness. Their approach focuses on larger spatial scales in the region of kilometres and on situations where both measurements are available, using the SAR data to interpolate between the altimetry data. In this chapter, a deep learning method is introduced to extrapolate ICESat-2 altimetry-derived ice development information to Sentinel-1 SAR scenes at 100m resolution. This builds on the knowledge generated in the previous chapter, using supervised learning of the incidence angle slopes acquired from the global clustering technique by Cristea et al., 2020 to replace the GAN setup. The final model is no longer dependent on the altimetry data being present in the area. At this resolution, one is still able to capture processes such as leads opening up, the formation of large ridged areas and the breakup of the ice sheet in the marginal ice zone.

The aforementioned spatial decorrelation of non-simultaneous acquisitions due to the sea ice drift plays a central role in this endeavour. In this work, the SAR scenes are resampled to 100 by 100 metres, to mitigate the effect of speckle on various parts of the procedure. In contrast to the previous chapter this approach does not downsample the encodings by a factor of four, so there is still an effective upgrade in resolution from 160 to 100 metres. To gauge the maximum sensible time difference between two measurements, one can assume a strong drift will result in displacement of 12 kilometres a day or 500 metres an hour. Then, 50 m (half of the pixel size of the downsampled SAR product) is covered in 10 minutes. Thus, constraining the time difference between the two measurements to at most 10 minutes should be sufficient for the measurements to still be spatially correlated and the same data as in the previous chapter can be used.

Relating freeboard height to ice development is not straightforward because the variable

layer of snow on top obscures the relation between freeboard and thickness. To obtain a somewhat meaningful measure relating the altimetry to ice development that generalises to a variety of ice conditions, this first foray into ICESat-2 and Sentinel-1 fusion will concern itself only with scenes acquired in October and November, when the older multiyear ice and freshly consolidated first-year ice are able to be separated from the bimodal nature of the freeboard distribution (fig. 6.1). In this instance the characteristics of the distribution determine the ice development measure.

The constraint to small time differences immediately leads to only a small amount of overlap between the measurements. Part of the reason is that the availability of the ATL-10 freeboard product of ICESat-2 requires both the absence of clouds as well as the presence of leads in the ice. Further constraints to a certain part of the year mean that the amount of available overlapping data is not large. Between 2018 and 2021, there are approximately 60 overlapping measurement pairs that fulfil these criteria and have a substantial overlap of more than a hundred Sentinel-1 pixels. Of these scenes, 48 admitted a clearly bimodal freeboard distribution needed to obtain a consistent measure of ice development that allowed to extract an ice development index as outlined below in the methodology section. The number of overlapping resampled Sentinel-1 pixels (100x100m) in this dataset is around 20000, which is not a large amount of data when looking to train data-driven algorithms. In this work, the input window is a 256 x 256 (= 65536) pixel image patch, so the total number of labelled pixels corresponds to less than a third of a single image patch.

The results from chapter 4 showed, that a substantial difference is made when the high-resolution classifier is able to learn from and make use of the spatial distribution of labels/ice classes. Because this 2-dimensional spatial distribution cannot be present in the 1-dimensional altimeter track, it is expected that traditional neural network-based approaches not tuned to deal with the sparsity of the data will not give satisfactory results here.

Because data sparsity can be a substantial difficulty for data driven models, the work in the previous chapter is expanded upon to show that incidence angle dependencies of the SAR backscatter – obtained from a linearly variable Gaussian mixture model Cristea et al., 2020 – can be used as a proxy measurement during training. This is done through a range of experiments and a novel transfer learning approach. To recap the core idea, remember that ice development and incidence angle dependence are correlated (see fig. 5.1), and thus the features needed to learn one of these variables are very similar to the features needed to learn the other. A model trained on the proxy data must therefore have learned high-level features that can be leveraged to fit the real (target) data. This is a concept known as transfer learning, which is very well established in machine learning (for an overview, see for example Weiss et al., 2016). In the previous chapter this task was formulated as a self-supervised task, but results of the incidence angle failed to live up to the global technique. The use of incidence angle dependencies for sea ice classification is explored in existing classifiers Lohse et al., 2021; Guo et al., 2022, which do not make use of deep learning or machine vision techniques. A large range of deep learning classifiers have been researched

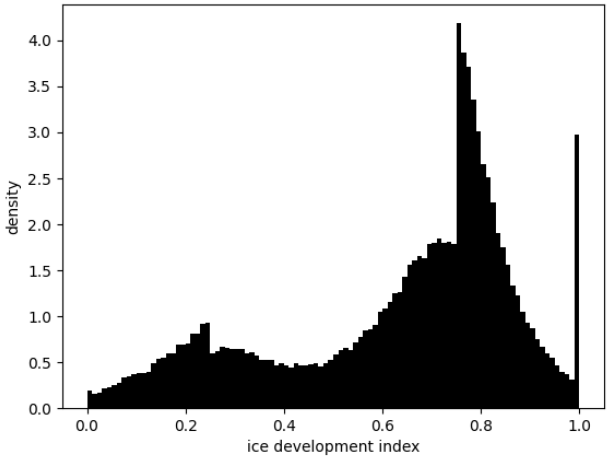


Figure 6.2: Histogram of label occurence for ice development index in the training dataset. The density is used as an inverse weight in the later training of the networks, to obtain a more balanced predictor.

that use manual annotations or ice charts as 'ground truth', but they are not designed to work with only sparse training data (e.g. Ressel, Frost, et al., 2015; Boulze et al., 2020; Ullah et al., 2021; Wang and Li, 2021; Nagi et al., 2021; Murashkin and Frost, 2021; Ren, Li, et al., 2022; Kortum, Singha, and Spreen, 2022). This includes all algorithms mentioned in the literature overview section 2.9. Another concept that is explored in this work is to further enrich the extracted features by combining the transfer learning task with an encoding task, the same as in the last chapter. This forces the network to encode a parametrisation of the local backscatter values and the predicted incidence angle dependencies at the same time. Autoencoders have been a part of the machine learning landscape for decades, dating back to the 1980's Goodfellow, Bengio, et al., 2016. The final learning of the ice development measure is then trained from the encoded layer (latent space). Compressing complex image data into a latent space before further processing is also a well established approach in the machine learning field (e.g. Ha and Schmidhuber, 2018; Wahlström et al., 2014).

6.3 Data

The data used in this study is the same as in the previous chapter, except that the SAR data is resampled to 100 by 100m pixel size, to obtain more freeboard measurements per pixel. In the previous chapter this was not necessary, as the encoder network downsampled the product by a factor of 4 to 160 by 160m pixel size. For details of the data see section 5.3. An example of IceSAT-2 and Sentinel-1 data can be seen in figure 6.3

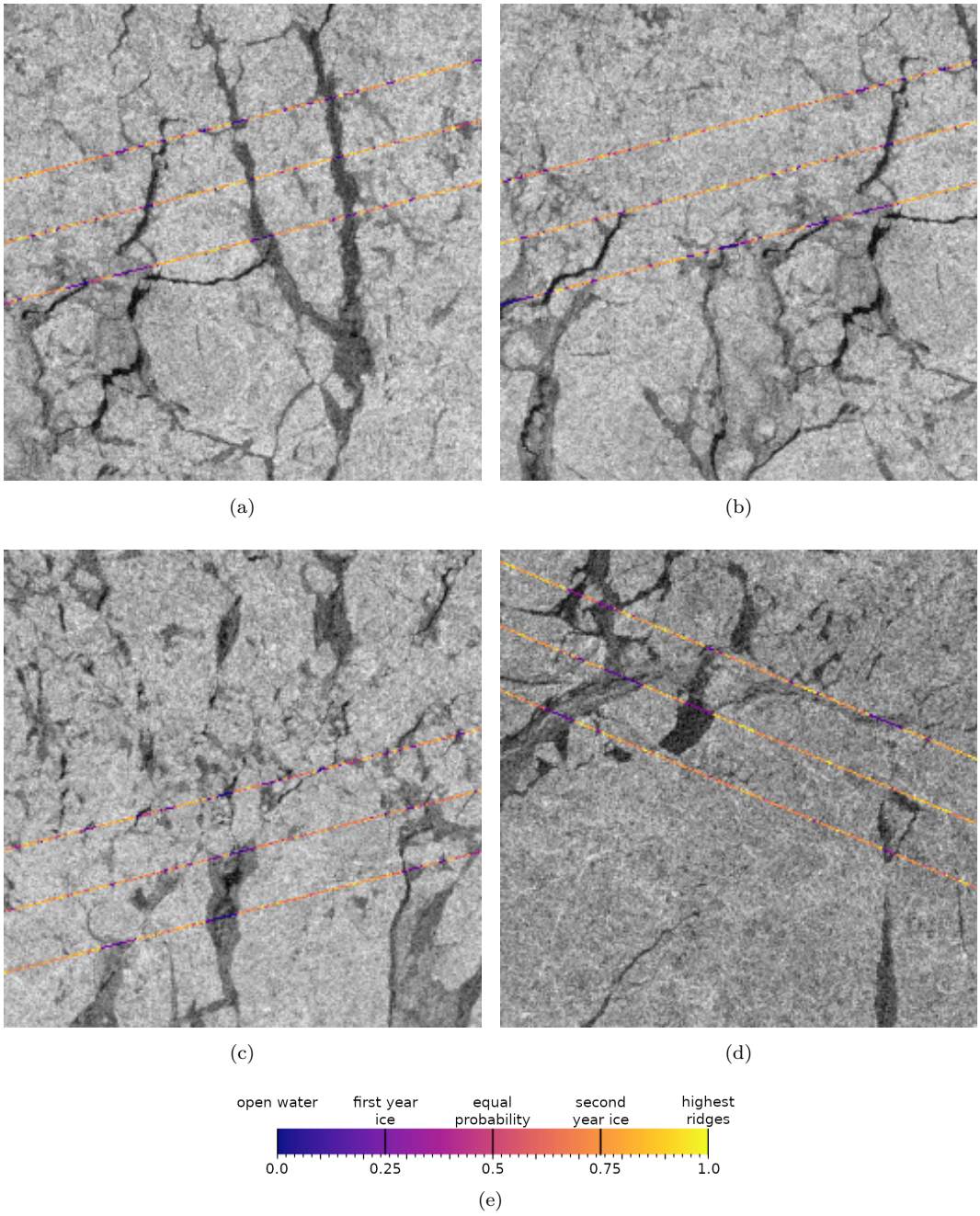
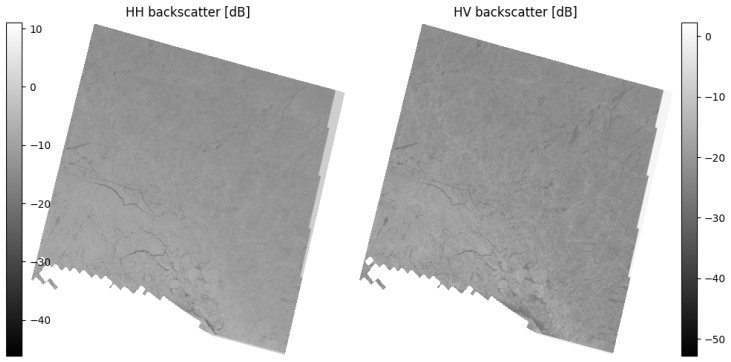
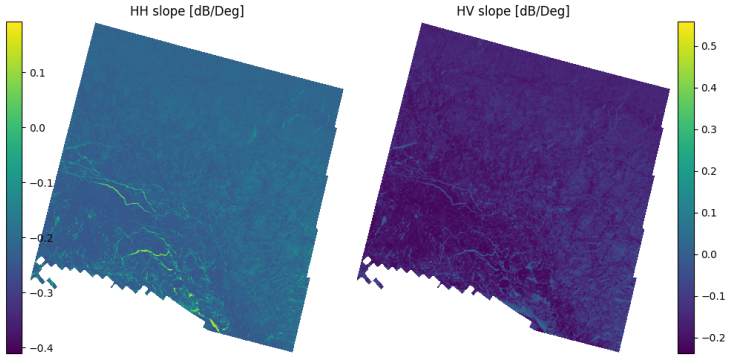


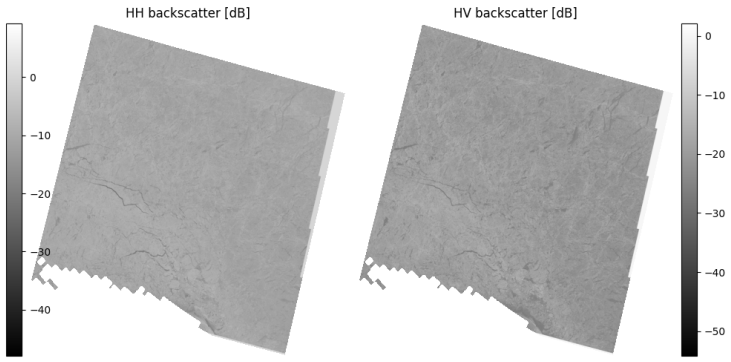
Figure 6.3: A 256 x 256 training patch of Sentinel-1 EW HH backscatter overlayed with the ice development index extracted from ICESat-2 altimetry as in figure 6.1.



(a) Backscatter



(b) Slopes



(c) Backscatter calibrated to 30°

Figure 6.4: Plots of the backscatter and slopes from a chosen scene captured north of greenland next to the entrance of the Fram Straight on November 30th 2019. The last image shows the backscatter once it has been calibrated to 30° across the scene, using the slopes above.

6.4 Methodology

6.4.1 Ice Development Index

The freeboard derived from the altimetry measurement is only 1-dimensional. Whilst it might be possible to segment this to obtain a number of classes, that would lead to a loss of information because of the discretisation of a continuous variable. Predicting freeboard or ice thickness from SAR directly, however, is very difficult because it is heavily dependent on snowfall and electromagnetic waves at C-band only have small penetration depth in arctic sea ice. Instead, we aim to also extract a one-dimensional index of ice development, where the snow on top of the ice can then be interpreted as some label noise. This makes it significantly easier to deal with than the bias one would introduce by segmenting this into classes. In chapter 4, the local surface roughness was additionally available to segment classes, so this effect was not as prevalent there. The index is constructed by using a Gaussian mixture to fit the bimodal freeboard distribution of a given ATL-10 track. Index 0 corresponds to 0m freeboard. The mean of the Gaussian fitting the first-year ice distribution is set to 0.25. The point where the probability of belonging to multiyear or first-year is equal corresponds to development index equal to 0.5 . The mean of the Gaussian fitting the multiyear ice is set to index value 0.75; all values greater than the MYI mode plus four times the standard deviation of that Gaussian is set to 1. All values in between are linearly stretched between the above-mentioned tie points. The index definition is illustrated in figure 6.1. From the index more concrete classifications into first- and multiyear ice, heavily ridged areas or young ice areas can be easily achieved via thresholding.

6.4.2 The Incidence Angle Dependence Estimation Method

In this work, the incidence angle dependence of sea ice backscatter will play a central role. To briefly summarise what has been established in the previous chapters: This dependence has been observed to be approximately linear in the past Onstott, 1992 and has been shown to correlate with the sea ice development (figure 5.1). In principle, it cannot be measured with a single satellite acquisition. Under the reasonable assumption that the same ice type occurs in a scene at various incidence angles, however, it can be approximated using a clustering approach: To extract the incidence angle dependence (slope) for every pixel in the image, the previously mentioned technique developed by Cristea et al., 2020 is used, which makes use of the linearity of the relationship of brightness and incidence angle. Remember, that the core concept is to fit a multivariate (for HH and HV channels) Gaussian mixture model to the measured backscatter values, where the mean of the Gaussian clusters is allowed to vary linearly by incidence angle whilst variance is kept constant. The linear variation (slope) of the clusters then corresponds to the incidence angle dependence of sea ice in that cluster. To obtain continuous slopes for this work, slopes are interpolated from the slopes of the different clusters for each pixel, weighted with the probability of belonging to the corresponding cluster. It is crucial to know for further tasks in this network, that the

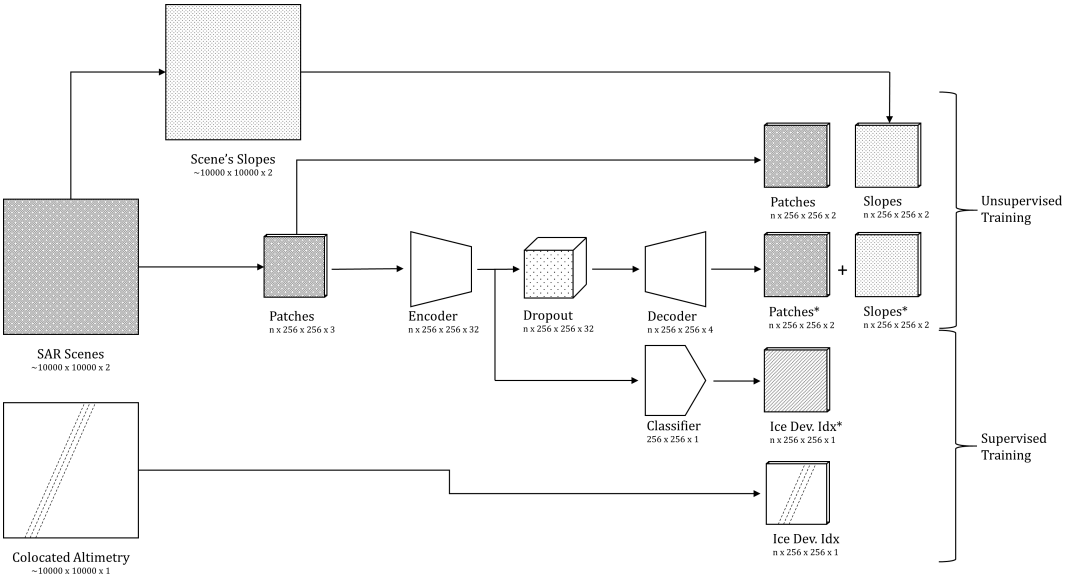


Figure 6.5: The network configuration presented in this chapter. Network predictions are additionally labelled with a * symbol. In the context of transfer learning, the unsupervised task is the transfer task and the supervised task is the target task.

incidence angle variance in a single patch of the image used for training is not large enough to infer the incidence angle dependent slopes in this manner. This clustering technique is only feasible on a scene with a larger incidence angle span or a multitude of scenes spanning the space.

6.4.3 Networks

Here the network architectures used in this work are described. None of them are new to the reader of this thesis, but the information below may serve as a reminder. In general, however, the developed methods are architecture-agnostic, meaning they do not depend on the exact architecture used. This means the core idea is built on focusing on how to use available data for models to converge to a desired solution, not to find the best possible network architectures to do this. As a result, the techniques described here should translate freely to any other network architectures with image-like output dimensions. A visualisation of the networks is shown in figure 6.5. One of the building blocks used in this work is a simple out-of-the-box U-net Ronneberger et al., 2015. The U-net has established itself as a reliable baseline image-to-image model and has also been shown to perform competitively in ice classification tasks from SAR (e.g. Nagi et al., 2021; Ren, Li, et al., 2022).

To extract features valuable for ice classification without any labels available, this work follows two strategies as devised in the previous chapter: Firstly, the incidence angle slopes are used as a proxy for ice type labels. Because these two quantities are correlated, it is reasonable to assume that the features needed to predict one of the quantities are transferable to predicting the other. As mentioned before the spatial distributions of the two should be

very similar, therefore fulfilling the criterion for good classification set in chapter 4. Secondly, the surrounding local spatial information of the scene is encoded into the feature vectors. As in the last chapter, these two strategies can be aligned by building a network configuration, where for every pixel an encoding is generated that parameterises the local neighbourhood of the SAR image as well as the neighbourhood of the incidence angle dependencies. In contrast to the last chapter, the mixture model extracted slopes allows to generate these encodings at full resolution of the original (rather than a fourth of the resolution).

The overarching idea of this transfer learning is to train the majority of parameters of a network on a proxy task, where data is abundant and where the internal features derived by the network are also valuable for the target task. Once the network has converged on the proxy data, the final layer(s) are reinitialised and retrained on the sparse data describing the target task. The rest of the network parameters are kept constant during this final training step. In our case the target task is deriving the altimetry derived ice development index, and the proxy task is parametrising the SAR image and the local incidence angle dependencies. The idea of fitting only a small number of parameters to the sparse data is to avoid overfitting and to generalise well, a strategy that was shown to be successful in chapter 3.

The Encoding Task

As mentioned above, the image-to-image network is set up in such a way that the output encodes a parametrisation of the local SAR backscatter as well as a prediction for the incidence angle dependencies for every pixel. This means the input to the encoding task is the backscatter in both polarisation channels (HH and HV) as well as the incidence angle, and the output is also the two backscatter channels plus two channels of the respective incidence angle dependencies. In a convolutional (auto-)encoder setup, this would mean having a bottleneck in the centre of the network, which has reduced spatial dimensions and thus would not result in encoding for every pixel. Instead, the approach here is to take a given image-to-image network as an encoder E (we will use a U-net in this investigation, but any other image-to-image architecture could be used) and to expand its output channels to the number of encoded features and then drop out the majority of encoded pixels, so that all those that remain need to contain the information about the neighbourhood to be able to reconstruct the entire image patch later on. For this research, we decided on a number of 32 features. Thus, the encoder E can be described as

$$\begin{aligned} E : \text{Input} \in M_{[256,256,3]} &\mapsto \text{Output} \in M_{[256,256,32]} \\ (\sigma_{HH}, \sigma_{HV}, \theta) &\rightarrow (\text{ Features }), \end{aligned} \tag{6.1}$$

where M_{dims} denotes the space of real matrices with dimensions $dims$ and θ is the local incidence angle. We use the hat $\hat{\sigma}$ to describe the predicted values and no hat to describe real values.

Additionally, a decoder D is introduced. This is another image-to-image network, that

has as input the encoded image with 32 channels and as output the four original output channels (backscatter and incidence angle slopes), expressible as

$$\begin{aligned} D : \text{Input} \in M_{[256,256,32]} &\mapsto \text{Output} \in M_{[256,256,4]} \\ (\text{ Features }) &\rightarrow (\hat{\sigma}_{HH}, \hat{\sigma}_{HV}, \hat{\partial_\theta \sigma}_{HH}, \hat{\partial_\theta \sigma}_{HV}). \end{aligned} \tag{6.2}$$

Here the incidence angle slopes of polarisation channel XX are written as $\partial_\theta \sigma_{XX}$. Between passing from the encoder to the decoder, there is a dropout layer $DO_{.98}$, that randomly drops out 98% of encoded pixels. Thereby, the decoder is forced to reconstruct the outputs from the resulting sparse matrix. Before dropout, this results exactly in an encoded space as desired, with encodings for every pixel. The exact architectures of the encoder (U-net) and the decoder used in this work are found in the appendix 6.7, tables 6.3 and 6.5. The entire encoding task from start to end can be expressed as

$$\begin{aligned} &Enc = (D \circ DO_{.98} \circ E) \\ &Enc : (\sigma_{HH}, \sigma_{HV}, \theta) \rightarrow (\hat{\sigma}_{HH}, \hat{\sigma}_{HV}, \hat{\partial_\theta \sigma}_{HH}, \hat{\partial_\theta \sigma}_{HV}) \end{aligned} \tag{6.3}$$

The total loss function L_{Enc} of this step is just a sum of the mean squared errors:

$$\begin{aligned} L_{Enc} = 0.5 \; &MSE((\sigma_{HH}, \sigma_{HV}), (\hat{\sigma}_{HH}, \hat{\sigma}_{HV})) \quad + \\ 500 \; &MSE((\partial\sigma_{HH}, \partial\sigma_{HV}), (\hat{\partial\sigma}_{HH}, \hat{\partial\sigma}_{HV})) \end{aligned} \tag{6.4}$$

The co-factors of 0.5 and 500 were adjusted manually to compensate for the different scales of the two properties. All models are trained with the Adam optimiser, developed by Kingma and Ba, 2014, and a learning rate of 10^{-4}.

The encoding task is trained on all available scenes, and patches are selected randomly from the data during training at run-time. In principle, any number of scenes could be used here, as no additional information is required. Because the incidence angle dependencies have to be predicted and are not given as input, the encoder is forced to learn the underlying logic of how to predict the slopes from the intensity images. The same logic should also be useful for predicting the ice development index.

An alternative approach could have been to follow a traditional auto-encoder-like setup and to retrain a new decoder mapping from the bottleneck to the ice development index for the target task. However, such a decoder is naturally rich in parameters and thus would undermine the transfer learning paradigm of keeping the number of parameters learned on the target task small. In turn, this would make the target task more susceptible to overfitting.

The Target Task

For the target task, the entire encoder is kept constant. On top, a tiny model F with only a single five-neuron hidden layer is trained to map the 32 features output by the encoder to the ice development index δ (see appendix 6.7, table 6.4). That way over-fitting to the sparse data can be greatly reduced whilst still outputting a prediction at the resolution of

the input image. One can express the target task as

$$\begin{aligned} Tar &= (F \circ E) \\ Tar &: (\sigma_{HH}, \sigma_{HV}, \theta) \rightarrow (\hat{\delta}), \end{aligned} \tag{6.5}$$

with a standard loss function

$$L_{Tar} = MSE(\delta, \hat{\delta}). \tag{6.6}$$

Here, $\hat{\delta}$ is the ice development index as retrieved from ICESat-2 and δ is the ice development index predicted by the model.

6.4.4 Experiments

The following compares the performance of the transfer learned model with a direct prediction of the ice development index without transfer learning using the same U-net architecture as is used in the encoding task. Due to choosing the same U-net architecture as a backbone to the different approaches, the only difference in performance comes from the configuration of the learning tasks and not from the architecture itself. As mentioned previously the transfer learning and encoding technique presented here should thus be freely transferable to any other architecture.

We also test some other improvements that could be made by exploiting the incidence angle dependences in different ways, by either using them as additional input to the classifier and/or using them to normalise the input image to backscatter as would be seen at 30 degree incidence angle across the entire image.

In total, there are 48 near-coincident (we allow for a maximum time difference of 10 minutes) Sentinel-1 EW mode acquisitions and ICESat-2 ATL-10 tracks. These scenes are randomly split into 30 scenes (12000 pixels) for the training, 9 scenes (3000 pixels) for the development set and 9 scenes (4000 pixels) for the test set. The development set is used to stop training at an optimal point and the test set is what the performances are evaluated on.

Some examples of SAR backscatter overlaid with the extracted ice development indices from altimetry can be found in figure 6.3.

6.5 Results

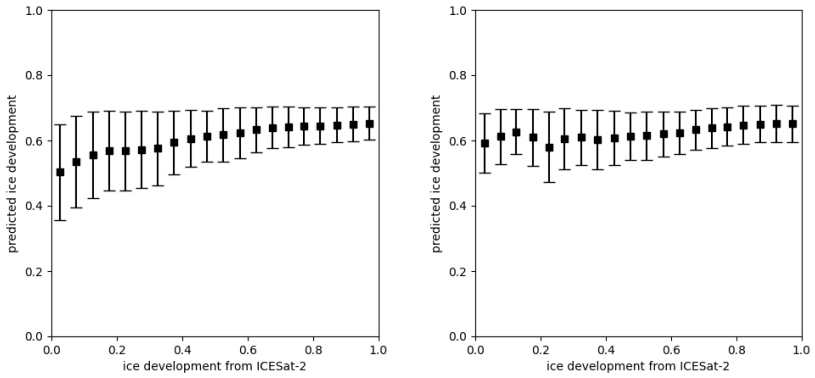
For reference, the losses of the encoding task are given here: The encoding network (without normalising the image first using the slopes from the GMM approach) performed with a root mean squared error between the real and the reconstructed backscatter of approximately 1.5 dB and a root mean squared error of predicting the slopes around 0.05 dB/Deg. When normalising the brightness as it would appear at an incidence angle of 30° for every pixel in the scene, the estimation of the slopes improved to around 0.033 dB/Deg. The backscatter reconstruction remained similarly accurate.

Model	avg. rmse	std. rmse
Base	0.2039	0.00184
Base + Slopes	0.2057	0.00163
Base + IA cor.	0.2001	0.00522
Base + Slopes + IA cor.	0.1983	0.00768

Table 6.1: Test set performances of networks that employed the transfer learning strategy. The root mean squared error and its standard deviation is given across 10 instances of the model trained under the same conditions. 'Slopes' indicates the slopes were also given directly as a feature to the final classifier. 'IA cor.' indicates the scenes were normalised to $30°$ incidence angle using the slopes prior to ingestion into the network. The encoder architecture used here is a U-net with 32 output channels. The decoder architecture uses consecutive normal and transpose convolutions. Both can be found in detail in the appendix 6.7.

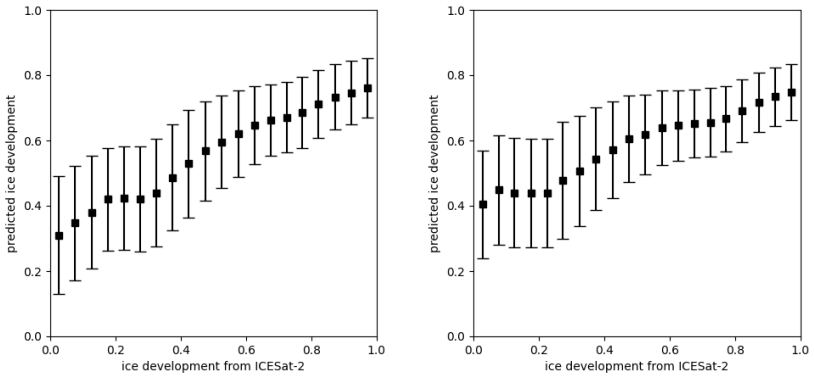
Model	avg. rmse	std. rmse
Base	0.2317	0.01560
Base + Slopes	0.2402	0.01078
Base + IA cor.	0.2399	0.00719
Base + Slopes + IA cor.	0.2404	0.00942

Table 6.2: Test set performances of networks that did not employ the transfer learning strategy. The testing parameters are the same as in table 6.1. The architecture used here is a U-net with one output channel.



(a) Training Set (b) Test Set

Figure 6.6: Model predictions without transfer learning on the training and test sets.



(a) Training Set (b) Test Set

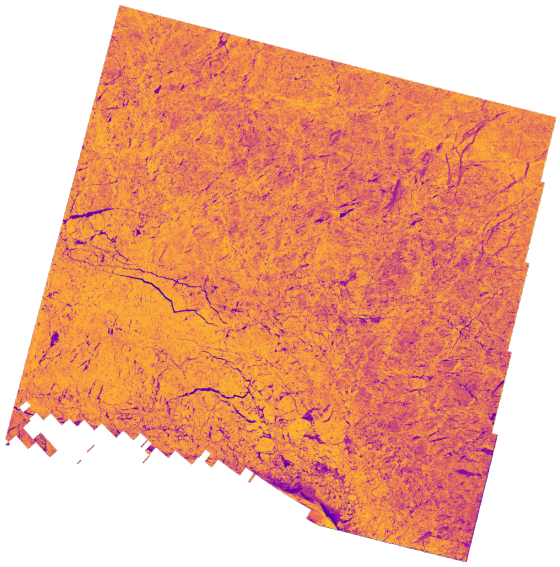
Figure 6.7: Model predictions with transfer learning on the training and test sets.

The tables 6.1 and 6.2 show how the inclusion of the encoding task improves the performance of the network by a rmse of 15-20% from the base model. Although the improvement is already significant, to truly evaluate the performances some more metrics are interesting. For further evaluation, the best architecture trained with and without the encoding task are chosen and their average predictions on the test set divided into 20 equally sized bins of ice development index is plotted. The results for the bare U-net are shown in 6.6. It is clear that even on the training set there is no meaningful separation between different ice development stages and on the test set, the situation looks dire. it would be fair to say that the retrieval task has failed in this case. When looking at the setup with the transfer learning in place in figure 6.7, it is evident that some separation of the ice development stages has occurred, although the extremes of the spectrum are still difficult to predict correctly. One needs to keep in mind, though, that with the uncertainty of snow thickness and the uncertain correlation of freeboard and ice development visible on SAR, the labels have to be considered as being very noisy. The error bars show that this is particularly true for the younger ice regime, where the data is sparsest (6.2). The uncertainty of the model is evident in the fact that the predictions tend to lie close to 0.5. Visual inspection of the results suggests that the network has captured core features of ice development and even advanced characteristics such as heavily ridged areas, which are evidence of past deformation events.

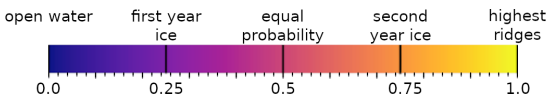
In figure 6.8, the successes of the approach presented here in contrast to a naive prediction without transfer learning is self-evident. Younger and older ice seem well separated with the proposed technique and a difference in development between ice closer to the coast and that drifting out towards the Fram Straight can be seen. Linear kinematic features can be made out, with both leads and ridged areas being visible in the image. Some swath effects do persist despite the denoising efforts that have been made.

To illustrate both the successes and shortcomings of the trained algorithm, we included a predicted scene from the marginal ice zone, which can be seen in figure 6.9. Despite having

(a) Without Encoding/Transfer Learning Task

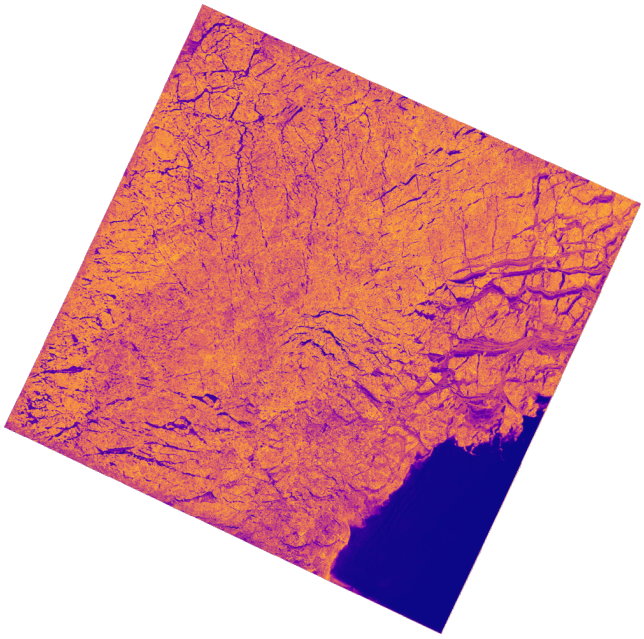


(b) With Encoding/Transfer Learning Task

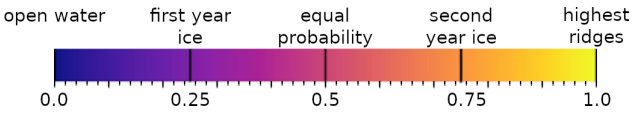


(c)

Figure 6.8: Model predictions without (a) and with (b) transfer learning and encoding of the scene. The scenes backscatter properties are shown in figure 6.4. The ice development index is shown in figure 6.1 and described in section 6.4.1.



(a)



(b)

Figure 6.9: Scene from the marginal ice zone, predicted by a model using the transfer learning and encoding technique. Note that during training no labels were included over open water or in the marginal ice zone. The ice development index is shown in figure 6.1 and described in section 6.4.1.

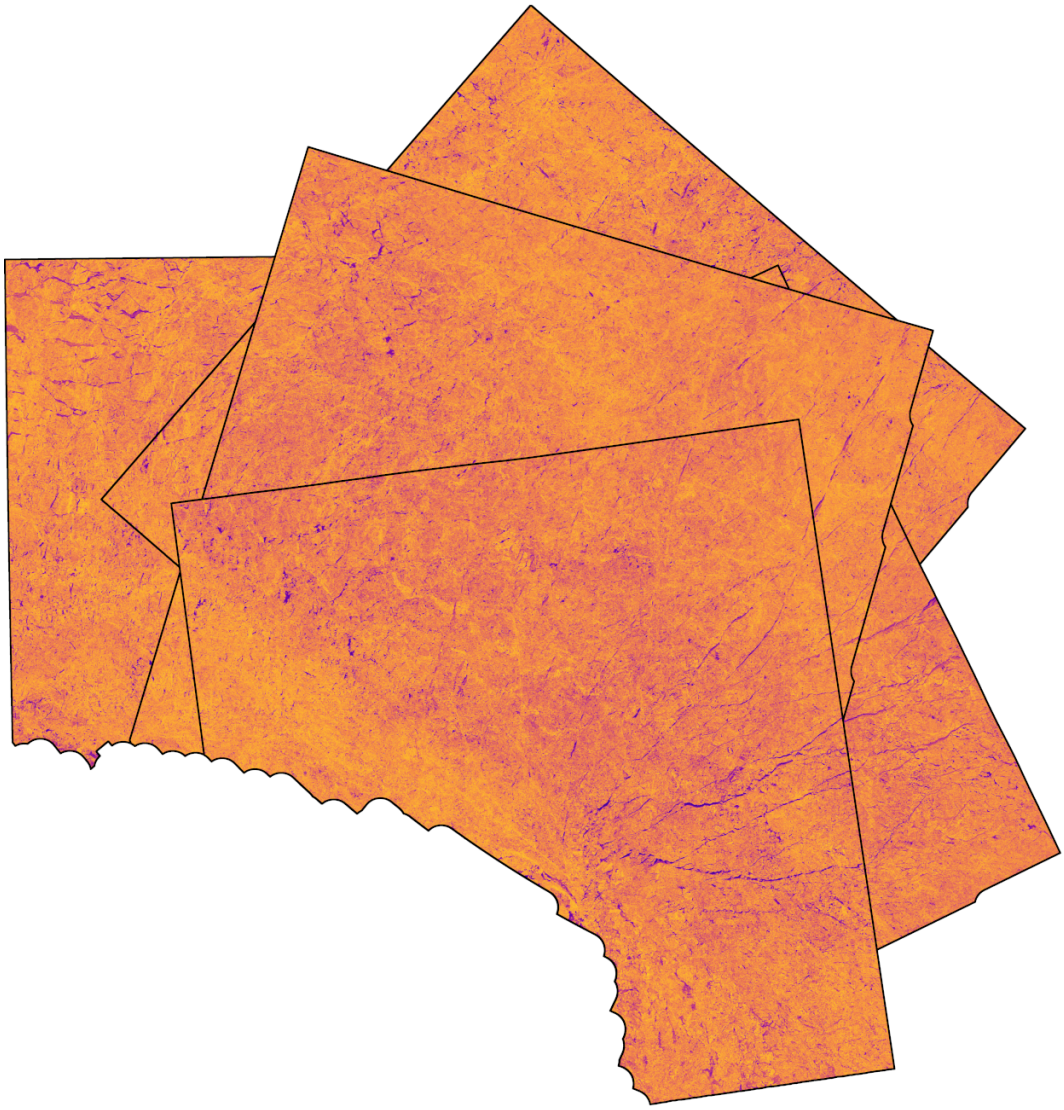
not trained in this area and no labels being available here, the algorithm has no problem identifying open water and extrapolates even to these unseen regions well. However, some of the younger ice close to the ice edge is predicted to have unreasonably high ice development index.

To demonstrate the performance on an area that was more frequent in the training set, a mosaic of five scenes is created in figure 6.10. From a visual persepectve, there seem to be little discontinuities in the images, except for a patch of ice near the cost which is likely young or fast ice (seen in fig 6.11 in the bottom right as a bright region without any multiplicity due to spatial decorrelation between scenes). The other high standard deviations in the classification seem to be due to the drift. Individually, the classification of the leads is consistent across the scenes and the model seems robust to the difference in geometries of the acquisitions.

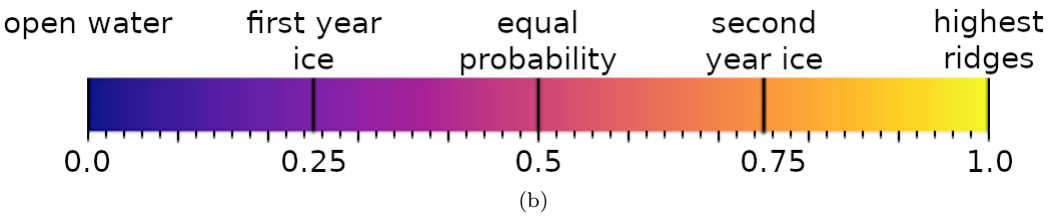
6.6 Discussion

It is clear from both the quantitative as well as the qualitative results of the networks, how the encoding task facilitates meaningful extrapolation of the altimetry-derived ice development, capturing the key trend. The accuracy of the retrieval, however, is still sub-optimal (figure 6.7) and hard to judge due to the underlying label uncertainty. When considering that the amount of training data covers only a fifth of a single input patch to the network, some remaining uncertainty should be expected. What is certain, however, is that without the transfer and encoding methodology, extrapolation cannot be achieved with neural networks. The further uses of the incidence angle slopes as an input feature do not show significant improvement in the retrieval task on average. It is possible that the network is prone to overfitting to these features during training. Using the slopes to calibrate the data as it would be seen at 30° led to a significant increase in the variance in model training when the encoding task is used. With slightly better mean performance and higher variance, it comes as no surprise that the overall top 25% of all 40 models tested all come from instances where this calibration has been undertaken.

So far, the extrapolation has been limited to only a certain season in the year, where the extraction of an ice development index is relatively straightforward. Expanding this approach to other seasons and the marginal ice zone will be more challenging. Part of the reason is, that the amount of overlapping data at 10 minutes of time difference is sparser in other months and non existing inside the marginal ice zone. Also, the freeboard distribution is harder to interpret as snow cover increases and ice thickens. But our results show that even from sparse data meaningful extrapolation is possible, which gives hope that in these conditions this approach can still be successful. Preliminary results from the marginal ice zone in figure 6.9 are promising. Obtaining labels in the marginal ice zone is not possible from altimetry due to the influence of waves, so here some other data or even manual labels would have to be utilised.



(a)



(b)

Figure 6.10: Mosaic of five Sentinel-1 scenes of the central Arctic north of Greenland. All where acquired on November 12th 2021, within 6 hours of one another.

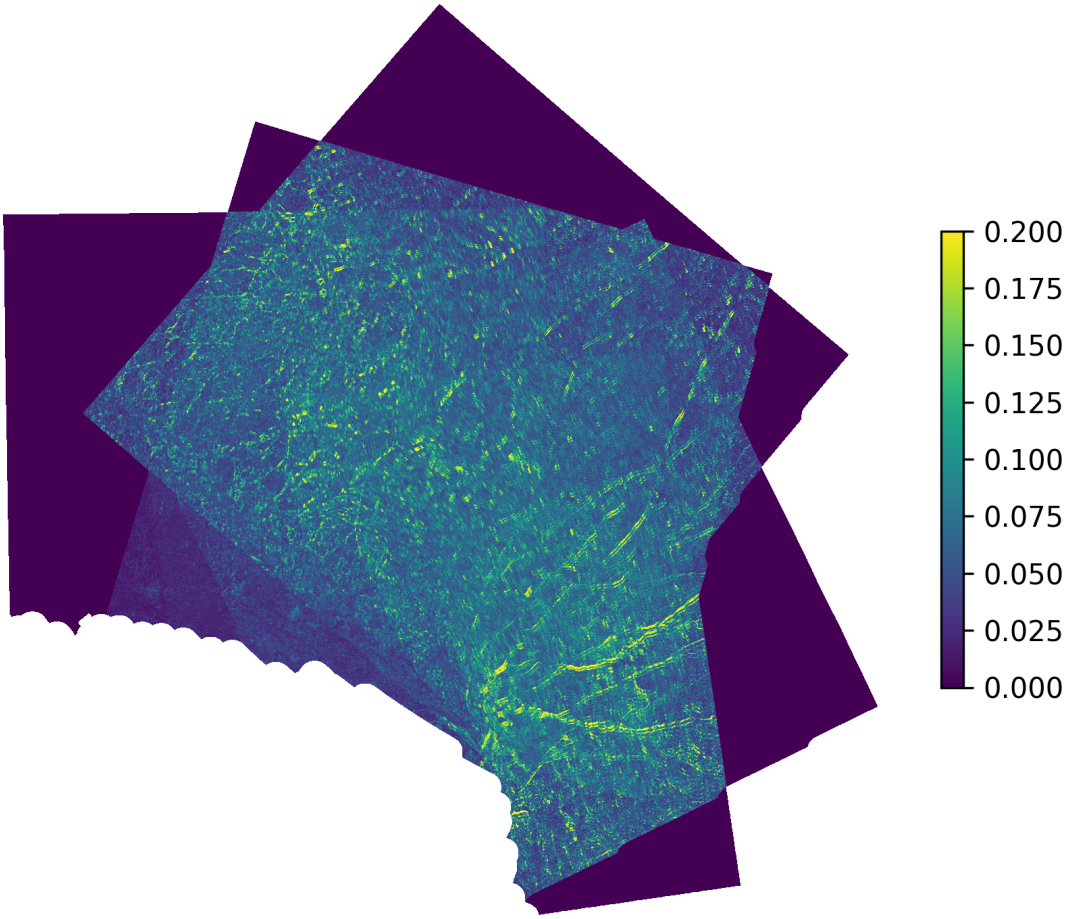


Figure 6.11: Standard deviation (between the different classified scenes) of the mosaic in fig. 6.10. The multiplicity of some features (mainly leads) is evidence of spatial decorrelation due to sea ice drift.

It is worth mentioning that there is one existing paper on overcoming sparse labels in Khaleghian, Ullah, Kræmer, Eltoft, et al., 2021, that uses an established self supervised learning method to overcome these problems. This has only been shown to work on manual labels with centre-pixel classification models so far and has no access to the spatial ice type distribution. Therefore it is not clear how well it would perform on high resolution labels.

There is currently an effort in the scientific community to obtain even more information from the waveforms of the altimeter about the properties of the ice surface. For example Duncan and Farrell, 2022 developed a product increasing the spatial resolution and Happ et al., 2023 are working on advanced ice development detection. If these efforts are successful, the extrapolation to SAR scales would certainly also be of interest, and the techniques developed in this work would be easily transferable.

Whilst altimetry data is a very good example for sparse yet detailed measurements of sea ice, the approach presented in this work of using incidence angle dependence as a proxy for ice development can in principle be used for other parameters. The stronger the two are correlated, the more fruitful this approach will be.

What is particularly exciting about this line of research is that the retrieval models can be improved by learning from unlabelled data, which opens up the opportunity to leverage the treasure trove of data that is the Sentinel-1 archive. The work shows a clear approach to mitigate the problem of ground truth sparsity, which is one of the most fundamental problems of sea ice classification or property retrieval.

6.7 Appendix: Architectures

We briefly present the network architectures used in this investigation. The following conventions are used to keep the figures concise. FCX is short for a fully connected layer with X neurons. ConvX x Y denotes a 2D convolutional layer with filter size X and Y filters. Unless otherwise specified the convolutional layers have stride 1. If a layer has multiple inputs, they are concatenated along the channel dimension before being parsed to the layer.

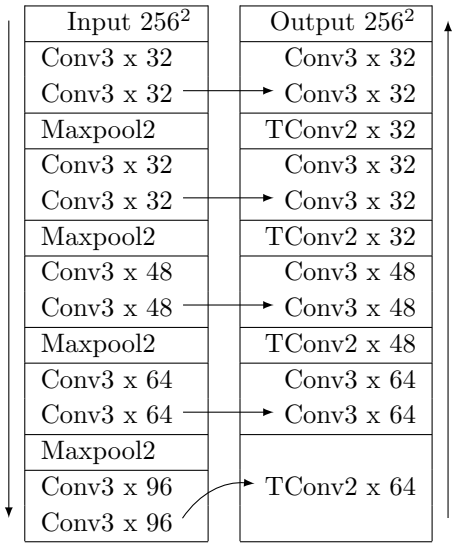


Table 6.3: The Unet architecture as used in this paper as a standalone model with a final Conv1 x 1 layer to map to the ice development index or as shown here ending in the encoded space with 32 features/channels. Originally developed and published in Ronneberger et al., 2015. The ReLU activation is used throughout the network and the padding is set to 'same' where applicable.

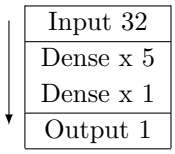


Table 6.4: The final layers mapping from the encoded space to the ice development index. We use the gelu activation function except in the output which uses sigmoid.

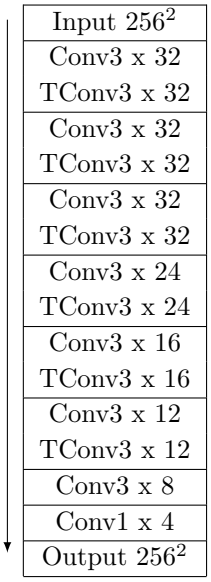


Table 6.5: The decoder architecture used in the paper. The ReLU activation is used throughout the network. The padding is set to 'same'. The final Conv1 layer has no activation function - it is just a linear pixelwise mapping.

Chapter 7

Conclusion

At the beginning of this thesis, two main problems of sea ice remote sensing with SAR were identified: the ground truth problem and the inversion problem. Machine learning and especially deep learning techniques have been at the forefront of research in the past years. To be able to generate valuable information that has the potential to inform climate models, classifiers need to perform near the resolution of the SAR sensor. Whilst convolutional neural networks are adept at handling the inversion problem by using contextual image data for classification, they are highly dependent on large amounts of ground truth. Therefore, it had so far not been possible to build robust classifiers at the desired fidelity that capture even the pixel-scale intricacies of the SAR product, as is made evident by the overview of existing methods, which emphasised that all existing work has been done on manual labels, which do not capture the full detail of the SAR scene and translate that quality to classifiers trained on them. Through a series of studies presented in chapters 3 - 6, machine learning methodologies were advanced to classify near the resolution and at similar fidelity as the underlying SAR product. Through the construction of high frequency timeseries and full resolution ground truth from additional measurements, these developments were tested in the diverse ice conditions present during the uniquely comprehensive observations made during the MOSAiC expedition.

In chapter 3, heuristics for robust classification were developed using the unique opportunity provided by long time series of the same ice from the MOSAiC campaign. It was demoonstrated that a probabilistic treatment of the labels and large contextual windows were critical to a robust algorithm. To facilitate the use of large contextual windows in a way that was not detrimental to generalisation, constraining the network to a low number of parameters was paramount. In the evaluation of the MOSAiC timeseries it was additionally found that very high backscatter floes at X-band classified as heavily deformed ice, were either forming on larger timescales than the observed period and/or in a different part of the Arctic (fig. 3.15). This suggests that they were most likely fractions of perennial ice. With the region around MOSAiC being heavily dominated by first and second-year ice, it could be a good representation for what the majority of the Arctic will look like in the future. In that case, the classification timeseries suggests that heavy deformation to the degree of perennial ice is less likely to form in the future sea ice pack. Airborne laser scanner data confirmed that high roughness regions were forming disproportionately rarely in the observed region. The work in chapter 3 also revealed how the coarse manual labels translate to coarse predictive qualities of the classifier. Thus, the predictions give valuable insight into the diverse regional ice conditions over the course of 7 months. They are, however, not of high enough fidelity to retrieve information about the spatial distribution of ridges or similar high fidelity ice properties.

To train models at such fidelity, one would need labels of that quality as well. Opportunity to gather ground truth data of that nature and have satellite SAR data available at the same time is rare. The joined efforts of the MOSAiC mission facilitated an environment to create such a dataset of SAR data overlapping with airborne laser scanning topography

surveys. The synthesis of this dataset and the subsequent tests of convolutional neural networks is detailed in chapter 4. In the introduction, the open question of how established machine learning model performan on high resolution ground truth was introduced. Using the synthesised dataset, it was shown how strongly the bias of manual labels in other studies was artificially boosting the performance of machine learning models. Previous research had found performances upwards of 95% accuracy on ice charts and manually labelled ice types. Using high resolution ground truth from measurements showed that the best models cannot reach 70% of accuracy consistently. It also unveiled one of the key ingredients that deep learning models need to make reliable predictions: the abililty of learning the spatial ice type distribution. This information is not presented in ice charts and only in very meticulous manual annotations. This poses an additional challenge to the use of machine learning algorithms in automated sea ice retrieval.

The classifications of the MOSAiC timeseries also suggested that first-year ice areas had experienced more deformation than second-year ice areas in terms of spatial coverage, even when taking into account that deformation cannot be detected on top of previously existing deformation. The uncertainty of the retrieval algorithm means this trend cannot be shown definitively; if true, however, this suggests that the younger ice sheet of the future Arctic would see more deformation events (more ridges forming and leads opening), which will, in turn, affect the coupling of ocean, ice and atmosphere, and thus the climate feedback mechanisms.

In the introduction, the question was posed if deep learning models can classify robustly throughout different ice conditions. It was found, that models can be designed towards robust classification (4) using a range of techniques, such as label smoothing, discriminators and parameter constraint. However, in chapter 4, it was found that these techniques depend on the type of training data and that the robustness of larger deep learning models seemed largely independent of the architecture on labels derived from measurement.

The importance of the label distribution to sea ice retrieval is a considerable challenge to retrieval models, because this kind of data simply does not exist on an Arctic-wide scale. Therefore in chapter 5, the idea of using incidence angle dependencies as proxies for ice types was first explored by developing an adversarial setup that allowed to prediction of incidence angle dependencies (slopes) only from a local image patch. Due to the correlations of ice types and incidence angle dependencies the spatial distributions of them are also related and therefore the features to learn one were hypothesised to coincide with the features to learn the other. In the chapter it was demonstrated that a network can learn to predict incidence angle dependent slopes aligning with existing measurements of the same quantity, using only a local image patch as input. However, the measurement proved to be less detailed than an existing global clustering method for getting local incidence angle dependencies. Nonetheless, it was shown that the encoding part of the model had learned features which were very useful for ice classification, without having been given any labels at all. Thus, it suggests there is a way of training a vast majority of a network's parameters for high

resolution ice classification without using any labels. It also proves that the incidence angle dependencies are a useful tool for learning important information from SAR imagery of sea ice without any labels, addressing one of the open questions that was posed in the introduction.

Building on the previous findings, the final chapter introduced an architecture-agnostic transfer learning strategy to make use of the detail of the incidence angle dependencies as extracted by the global mixture model approach and to leverage them for a retrieval task. By also using an encoding task which had proven to be crucial in the generative adversarial setup in the previous chapter, this approach solved a large part of the ground truth problem, as very sparse ICESat-2 altimeter measurements were meaningfully extrapolated to a far larger space of sea ice acquisitions, without overfitting significantly to the training data. With this methodology a vast part of the network can be trained without having any labels at all (99.99% of parameters in the case of the Unet). Using the incidence angle dependencies as proxies means that the ice type distribution shown to be important to successful high-resolution classification in chapter 4, is also present for the network to learn from. Therefore, machine learning networks can finally overcome both the inversion and the ground truth problems at the same time. The strategy also unlocks the potential for big data technologies to learn high resolution information from any amount of SAR scenes without any additional labels, which allows for retrieval near the resolution and fidelity of the SAR product.

Chapter 3 visualised clearly, how the seasons have a big effect on the performance of classifiers, but as the seasons with worst performance were also those with least amount of data, it is unclear if the challenge of classification was just a result of the unavailability of training labels. Having developed a strategy to overcome the two big challenges of sea ice classification with deep learning, it is left to show that this can work on a multi-seasonal scale as well and preserve robustness to even more diverse ice conditions and in the marginal ice zone. Surely this will come with its own set of challenges. Due to the strategies developed in this thesis, however, the entirety of the existing archive of Arctic SAR scenes can, in principle, be used to combat this problem: A robust strategy has been developed, that can increase model performance with SAR data alone. Because the developed methods are architecture agnostic, any current and future developments of advanced machine vision architectures, such as transformers or diffusion based networks, can also make use of this training strategy without adjustment. It therefore presents a clear approach to elevate sea ice observations from SAR to higher resolution, without the need for extensive high resolution ground truth data.

To give some context for the improvements made in this thesis, some applications are discussed, that can make use of observations of the quality that the work in this thesis enables. Novel work by Hutter, Bitz, et al., 2023, for example uses SAR observations to parametrise sea ice dynamic processes, using only simple thresholding of the SAR data to define leads. With better and higher resolution lead classification such studies could be further improved upon. In the future, floe edges in the marginal ice zone can be resolved at high resolution,

revealing information about the mechanical stresses of penetrating ocean waves that lead to floe braking events. Retrieving deformation zones at high resolution enables a more detailed parametrisation of form drag and coupling of ice and atmosphere. Continuous parametrisation of the deformation at such high resolution could also be connected to meltpond coverage, which has a strong climate feedback. A core aspect of understanding future change, is the contrast in behaviour of multi- and first-year ice, which can only be resolved with advanced observations that the models developed in this thesis are working towards. Retrieving such characteristics from the Sentinel-1 archive, for example, and looking at their evolution over the course of Arctic warming, will enable more accurate predictions about how the ice properties will change in light of continuous warming. Ultimately we are currently still lacking effective parametrisations for a variety of dynamic sea ice processes, because the community has no high resolution historic data record of the sea ice surface properties. This situation can be adressed in the near future with the developments presented in this thesis playing an important part in realising such high-performance retrieval models. This will finally allow one to connect the Arctic climate record of the recent past with high resolution observations of the ice conditions.

Appendix A

Acknowledgement

[redacted]

Bibliography

Aldenhoff, Wiebke, Céline Heuzé, and Leif E.B. Eriksson (2018). "Comparison of ice/water classification in Fram Strait from C- and L-band SAR imagery". *Annals of Glaciology* 59.76pt2. DOI: 10.1017/aog.2018.7.

Asadi, Nazanin, K. Andrea Scott, Alexander S. Komarov, Mark Buehner, and David A. Clausi (2021). "Evaluation of a Neural Network With Uncertainty for Detection of Ice and Water in SAR Imagery". *IEEE Transactions on Geoscience and Remote Sensing* 59.1. DOI: 10.1109/TGRS.2020.2992454.

Assur, A. (1960). "Composition of sea ice and its tensile strength". *SPIRE Research Report 44*.

Barber, David G. and Ellsworth LeDrew (1991). "SAR sea ice discrimination using texture statistics : a multivariate approach". *Photogrammetric Engineering and Remote Sensing* 57.

Beyer, L. (2015). *Pydensecrf*.

Bogdanov, A.V., S. Sandven, O.M. Johannessen, V.Yu. Alexandrov, and L.P. Bobylev (2005). "Multisensor approach to automated classification of sea ice image data". *IEEE Transactions on Geoscience and Remote Sensing* 43.7. DOI: 10.1109/TGRS.2005.846882.

Boulze, Hugo, Anton Korosov, and Julien Brajard (2020). "Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks". *Remote Sensing* 12. DOI: 10.3390/rs12132165.

Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan Yuille (2014). "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs". *CoRR. arXiv*.

Clausi, D.A. and Bing Yue (2004). "Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery". *IEEE Transactions on Geoscience and Remote Sensing* 42.1. DOI: 10.1109/TGRS.2003.817218.

Clausi, David A (2002). "An analysis of co-occurrence texture statistics as a function of grey level quantization". *Canadian Journal of Remote Sensing* 28.1. DOI: 10.5589/m02-004. eprint: https://doi.org/10.5589/m02-004.

Clausi, David A. (2001). "Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery". *Atmosphere-Ocean* 39.3. DOI: 10.1080/07055900.2001.9649675. eprint: https://doi.org/10.1080/07055900.2001.9649675.

Clausi, David A. and M. Ed Jernigan (2000). "Designing Gabor filters for optimal texture separability". *Pattern Recognition* 33.11. DOI: https://doi.org/10.1016/S0031-3203(99)00181-8.

Cooke, Colin L. V. and K. Andrea Scott (2019). "Estimating Sea Ice Concentration From SAR: Training Convolutional Neural Networks With Passive Microwave Data". *IEEE Transactions on Geoscience and Remote Sensing* 57.7. DOI: 10.1109/TGRS.2019.2892723.

Cristea, Anca, Jeroen van Houtte, and Anthony P. Doulgeris (2020). "Integrating Incidence Angle Dependencies Into the Clustering-Based Segmentation of SAR Images". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13. DOI: 10.1109/JSTARS.2020.2993067.

Cybenko, G (1989). "Approximation by superpositions of a sigmoidal function". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2. DOI: 10.1007/BF02551274.

Deng, Huawu and D.A. Clausi (2005). "Unsupervised segmentation of synthetic aperture Radar sea ice imagery using a novel Markov random field model". *IEEE Transactions on Geoscience and Remote Sensing* 43.3. DOI: 10.1109/TGRS.2004.839589.

Doulgeris, Anthony P. (2015). "An Automatic $\mathcal{U}$-Distribution and Markov Random Field Segmentation Algorithm for PalSAR Images". *IEEE Transactions on Geoscience and Remote Sensing* 53.4. DOI: 10.1109/TGRS.2014.2349575.

Duncan, K. and S. L. Farrell (2022). "Determining Variability in Arctic Sea Ice Pressure Ridge Topography With ICESat-2". *Geophysical Research Letters* 49.18. DOI: https://doi.org/10.1029/2022GL100272. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL100272.

Eicken, Hajo (2003). "From the Microscopic, to the Macroscopic, to the Regional Scale: Growth, Microstructure and Properties of Sea ice". *Sea Ice: An Introduction to its Physics, Chemistry, Biology and Geology*. Ed. by David Neville Thomas and Gerhard Dieckman. Wiley-Blackwell. Chap. 2.

Fily, Michel and D. Andrew Rothrock (1986). "Extracting Sea Ice Data from Satellite SAR Imagery". *IEEE Transactions on Geoscience and Remote Sensing* GE-24.6. DOI: 10.1109/TGRS.1986.289699.

Fraley, Chris and Adrian E Raftery (2002). "Model-Based Clustering, Discriminant Analysis, and Density Estimation". *Journal of the American Statistical Association* 97.458. DOI: 10.1198/016214502760047131. eprint: https://doi.org/10.1198/016214502760047131.

Fritz, Thomas, J. Mittermayer, B. Schaettler, S Buckreuss, R. Werninghaus, and W Balzer (2007). *Level 1b Product Format Specification*. Version Version 1.3. DLR: TerraSAR-X Ground Segment. December, 2007.

Geldsetzer, T. and J. J. Yackel (2009). "Sea ice type and open water discrimination using dual co-polarized C-band SAR". *Canadian Journal of Remote Sensing* 35.1. DOI: 10.5589/m08-075. eprint: https://doi.org/10.5589/m08-075.

Gélis, Iris de, Aurélien Colin, and Nicolas Longépé (2021). "Prediction of Categorized Sea Ice Concentration From Sentinel-1 SAR Images Based on a Fully Convolutional Network". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14. DOI: 10.1109/JSTARS.2021.3074068.

Geman, Stuart and Donald Geman (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6. DOI: 10.1109/TPAMI.1984.4767596.

Gill, Jagvijay, J.J. Yackel, Torsten Geldsetzer, and M. Christopher Fuller (2015). "Sensitivity of C-band synthetic aperture radar polarimetric parameters to snow thickness over landfast smooth first-year sea ice". *Remote Sensing of Environment* 166. DOI: 10.1016/j.rse.2015.06.005.

Golden, K.M. et al. (1998). "Forward electromagnetic scattering models for sea ice". *Geoscience and Remote Sensing, IEEE Transactions on* 36. DOI: 10.1109/36.718637.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press.

Goodfellow, Ian J., Jean Pouget-Abadie, et al. (2014). *Generative Adversarial Networks*. DOI: 10.48550/ARXIV.1406.2661.

Guo, W., P. Itkin, S. Singha, A. Paul Doulgeris, M. Johansson, and G. Spreen (2022). "Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture". *The Cryosphere Discussions* 2022. DOI: 10.5194/tc-2022-86.

Ha, David and Jürgen Schmidhuber (2018). "World Models". DOI: 10.5281/ZENODO.1207631.

HAN, Yanling et al. (2019). "A Cooperative Framework Based on Active and Semi-supervised Learning for Sea Ice Classification using EO-1 Hyperion Data". *TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES* 62.6. DOI: 10.2322/tjsass.62.318.

Happ, Lena, Stefan Hendricks, Andreas Gerndt, Lars Kaleschke, Riccardo Fellegara, and Stephan Paul (2023). "Employing data-science methods for mapping sea ice surface change on decadal time scales". *International Symposium on Sea Ice Across Temporal and Spatial Scales*.

Hara, Y., R.G. Atkins, R.T. Shin, Jin Au Kong, S.H. Yueh, and R. Kwok (1995). "Application of neural networks for sea ice classification in polarimetric SAR images". *IEEE Transactions on Geoscience and Remote Sensing* 33.3. DOI: 10.1109/36.387589.

Haralick, Robert M., K. Shanmugam, and Its'Hak Dinstein (1973). "Textural Features for Image Classification". *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6. DOI: 10.1109/TSMC.1973.4309314.

Harris, Charles R. et al. (2020). "Array programming with NumPy". *Nature* 585.7825. DOI: 10.1038/s41586-020-2649-2.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015). *Deep Residual Learning for Image Recognition*. DOI: 10.48550/ARXIV.1512.03385.

Hendricks, S. (2019). *Ice Drift - Transformation of GPS positions into a translating and rotating coordinate reference system*.

Howard, Andrew G. et al. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". *CoRR* abs/1704.04861. arXiv: 1704.04861.

Hua, Yuansheng, Diego Marcos, Lichao Mou, Xiao Xiang Zhu, and Devis Tuia (2021). "Semantic Segmentation of Remote Sensing Images With Sparse Annotations". *IEEE Geoscience and Remote Sensing Letters*. DOI: 10.1109/LGRS.2021.3051053.

Hutter, Nils, Cecila Bitz, and Luisa von Albedyll (2023). "Linking the evolution of floe-scale ice characteristics to its deformation history using satellite observations". *International Symposium on Sea Ice Across Temporal and Spatial Scales*.

Hutter, Nils, Stefan Hendricks, et al. (2022a). "Digital elevation models of the sea-ice surface from airborne laser scanning during MOSAiC." *Scientific Data (in review)*.

— (2022b). "Merged Grids of Sea-Ice or snow freeboard from helicopter-borne laser scanner during the MOSAiC Expedition, version 1." *Pangea*. DOI: https://doi.pangaea.de/10.1594/PANGAEA.950896.

Imber, James (2022). "Generative Network For Semi-supervised Sea Ice Classification". DOI: 10.36227/techrxiv.21081136.v1.

IPCC Working Groups I, II and III (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report*. Intergovernmental Panel on Climate Change.

Itkin, Polona et al. (2021). *Magnaprobe snow and melt pond depth measurements from the 2019-2020 MOSAiC expedition*. data set. DOI: 10.1594/PANGAEA.937781.

Jain, Anil K. and Farshid Farrokhnia (1991). "Unsupervised texture segmentation using Gabor filters". *Pattern Recognition* 24.12. DOI: https://doi.org/10.1016/0031-3203(91)90143-S.

Jiang, Mingzhe, Linlin Xu, and David A. Clausi (2022). "Sea Ice and Water Classification of RADARSAT-2 Imagery Based on Residual Neural Networks (ResNet) with Regional Pooling". *Remote Sensing* 14.13. DOI: 10.3390/rs14133025.

Johansson, A. Malin, Camilla Brekke, Gunnar Spreen, and Jennifer A. King (2018). "X-, C-, and L-band SAR signatures of newly formed sea ice in Arctic leads during winter and spring". *Remote Sensing of Environment* 204. DOI: https://doi.org/10.1016/j.rse.2017.10.032.

Johansson, A. Malin, Eirik Malnes, et al. (2020). "Consistent ice and open water classification combining historical synthetic aperture radar satellite images from ERS-1/2, Envisat ASAR, RADARSAT-2 and Sentinel-1A/B". *Annals of Glaciology* 61. DOI: 10.1017/aog.2019.52.

Kacimi, Sahra and Ron Kwok (2022). "Arctic Snow Depth, Ice Thickness, and Volume From ICESat-2 and CryoSat-2: 2018â€"2021". *Geophysical Research Letters* 49.5. DOI: https://doi.org/10.1029/2021GL097448. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL097448.

Karvonen, J. (2014). "A sea ice concentration estimation algorithm utilizing radiometer and SAR data". *The Cryosphere* 8.5. DOI: 10.5194/tc-8-1639-2014.

Karvonen, J., E. Rinne, H. Sallila, P. Uotila, and M. Mäkynen (2022). "Kara and Barents sea ice thickness estimation based on CryoSat-2 radar altimeter and Sentinel-1 dual-polarized synthetic aperture radar". *The Cryosphere* 16.5. DOI: 10.5194/tc-16-1821-2022.

Karvonen, J., M. Simila, and M. Makynen (2005). "Open water detection from Baltic Sea ice Radarsat-1 SAR imagery". *IEEE Geoscience and Remote Sensing Letters* 2.3. DOI: 10.1109/LGRS.2005.847930.

Karvonen, J.A. (2004). "Baltic Sea ice SAR segmentation and classification using modified pulse-coupled neural networks". *IEEE Transactions on Geoscience and Remote Sensing* 42.7. DOI: 10.1109/TGRS.2004.828179.

Karvonen, Juha (2017). "Baltic Sea Ice Concentration Estimation Using SENTINEL-1 SAR and AMSR2 Microwave Radiometer Data". *IEEE Transactions on Geoscience and Remote Sensing* 55.5. DOI: 10.1109/TGRS.2017.2655567.

Khaleghian, Salman, Habib Ullah, Thomas Kræmer, Torbjørn Eltoft, and Andrea Marinoni (2021). "Deep Semisupervised Teacher–Student Model Based on Label Propagation for Sea Ice Classification". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14. DOI: 10.1109/JSTARS.2021.3119485.

Khaleghian, Salman, Habib Ullah, Thomas Kræmer, Nick Hughes, Torbjørn Eltoft, and Andrea Marinoni (2021). "Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks". *Remote Sensing* 13.9.

Kim, Young-Soo, R. Onstott, and R. Moore (1984). "Effect of a snow cover on microwave backscatter from sea ice". *IEEE Journal of Oceanic Engineering* 9.5. DOI: 10.1109/JOE.1984.1145649.

Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization". *CoRR* abs/1412.6980.

Korosov, Anton, Denis Demchev, Nuno Miranda, Niccolò Franceschi, and Jeong-Won Park (2022). "Thermal Denoising of Cross-Polarized Sentinel-1 Data in Interferometric and Extra Wide Swath Modes". *IEEE Transactions on Geoscience and Remote Sensing* 60. DOI: 10.1109/TGRS.2021.3131036.

Kortum, K., S. Singha, G. Spreen, N. Hutter, A. Jutila, and C. Haas (2023). "SAR Deep Learning Sea Ice Retrieval Trained with Airborne Laser Scanner Measurements from the MOSAiC Expedition". *The Cryosphere Discussions* 2023. DOI: 10.5194/tc-2023-72.

Kortum, Karl, Suman Singha, and Gunnar Spreen (2022). "Robust Multiseasonal Ice Classification From High-Resolution X-Band SAR". *IEEE Transactions on Geoscience and Remote Sensing* 60. DOI: 10.1109/TGRS.2022.3144731.

Kortum, Karl, Suman Singha, Gunnar Spreen, and Stefan Hendricks (2021). "Automating Sea Ice Characterisation from X-Band SAR with Co-Located Airborne Laser Scanner Data Obtained During the Mosaic Expedition". *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. DOI: 10.1109/IGARSS47720.2021.9553340.

Kovacs, Austin, U.S. Army Cold Regions Research, and Engineering Laboratory (1996). "Sea Ice: Part I. Bulk Salinity Versus Ice Floe Thickness". *CRREL Report 96-7*.

Krähenbühl, Philipp and Vladlen Koltun (2012). *Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials*. arXiv: 1210.5644 [cs.CV].

Kwok, R., T. Markus, et al. (2019). "Surface Height and Sea Ice Freeboard of the Arctic Ocean From ICESat-2: Characteristics and Early Results". *Journal of Geophysical Research: Oceans* 124.10. DOI: https://doi.org/10.1029/2019JC015486. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JC015486.

Kwok, R., A. A. Petty, et al. (2022). *ATLAS/ICESat-2 L3A Sea Ice Freeboard, Version 5*. Version Version 5. NSIDC: National Snow and Ice Data Center. Boulder, Colorado. DOI: https://doi.org/10.5067/ATLAS/ATL10.005.. November, 2021.

Kwok, Ronald, Eric Rignot, Benjamin Holt, and R. Onstott (1992). "Identification of sea ice types in spaceborne synthetic aperture radar data". *Journal of Geophysical Research:*

Oceans 97.C2. DOI: https://doi.org/10.1029/91JC02652. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/91JC02652.

Li, Fan, David A. Clausi, Lei Wang, and Linlin Xu (2015). "A semi-supervised approach for ice-water classification using dual-polarization SAR satellite imagery". *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. DOI: 10.1109/CVPRW.2015.7301380.

Lin, Jianxin, Yingce Xia, Tao Qin, Zhibo Chen, and Tie-Yan Liu (2018). "Conditional Image-to-Image Translation". *CoRR* abs/1805.00251. arXiv: 1805.00251.

Liu, Huiying, Huadong Guo, and Lu Zhang (2015). "SVM-Based Sea Ice Classification Using Textural Features and Concentration From RADARSAT-2 Dual-Pol ScanSAR Data". *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 8. DOI: 10.1109/JSTARS.2014.2365215.

Liu, Xiao, Fanjin Zhang, et al. (2021). "Self-supervised Learning: Generative or Contrastive". *IEEE Transactions on Knowledge and Data Engineering*. DOI: 10.1109/tkde.2021.3090866.

Liu, Zhuang, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie (2022). *A ConvNet for the 2020s*. DOI: 10.48550/ARXIV.2201.03545.

Lohse, Johannes, Anthony Doulgeris, and Wolfgang Dierking (2021). "Incident Angle Dependence of Sentinel-1 Texture Features for Sea Ice Classification". *Remote Sensing* 13. DOI: 10.3390/rs13040552.

Lu, Yongyi, Yu-Wing Tai, and Chi-Keung Tang (2017). "Conditional CycleGAN for Attribute Guided Face Image Generation". *CoRR* abs/1705.09966. arXiv: 1705.09966.

Lyu, Hangyu, Weimin Huang, and Masoud Mahdianpari (2022). "Eastern Arctic Sea Ice Sensing: First Results from the RADARSAT Constellation Mission Data". *Remote Sensing* 14.5. DOI: 10.3390/rs14051165.

Mahmud, Mallik S., Torsten Geldsetzer, Stephen E. L. Howell, John J. Yackel, Vishnu Nandan, and Randall Kenneth Scharien (2018). "Incidence Angle Dependence of HH-Polarized C- and L-Band Wintertime Backscatter Over Arctic Sea Ice". *IEEE Transactions on Geoscience and Remote Sensing* 56.

Mahmud, Mallik S., Vishnu Nandan, Stephen E.L. Howell, Torsten Geldsetzer, and John Yackel (2020). "Seasonal evolution of L-band SAR backscatter over landfast Arctic sea ice". *Remote Sensing of Environment* 251. DOI: https://doi.org/10.1016/j.rse.2020.112049.

Maillard, P., D.A. Clausi, and Huawu Deng (2005). "Operational map-guided classification of SAR sea ice imagery". *IEEE Transactions on Geoscience and Remote Sensing* 43.12. DOI: 10.1109/TGRS.2005.857897.

Makynen, M.P., A.T. Manninen, M.H. Simila, J.A. Karvonen, and M.T. Hallikainen (2002). "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. DOI: 10.1109/TGRS.2002.806991.

Mäkynen, Marko and J. Karvonen (2017). "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* PP. DOI: 10.1109/TGRS.2017.2721981.

Mäkynen, Marko, Stefan Kern, Anja Rösel, and Leif Toudal Pedersen (2014). "On the Estimation of Melt Pond Fraction on the Arctic Sea Ice With ENVISAT WSM Images". *IEEE Transactions on Geoscience and Remote Sensing* 52.11. DOI: 10.1109/TGRS.2014.2311476.

Malmgren-Hansen, David et al. (2021). "A Convolutional Neural Network Architecture for Sentinel-1 and AMSR2 Data Fusion". *IEEE Transactions on Geoscience and Remote Sensing* 59.3. DOI: 10.1109/TGRS.2020.3004539.

Martıń Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*.

Mirza, Mehdi and Simon Osindero (2014). *Conditional Generative Adversarial Nets*. DOI: 10.48550/ARXIV.1411.1784.

Murashkin, Dmitrii and Anja Frost (2021). "Arctic Sea ICE Mapping Using Sentinel-1 SAR Scenes with a Convolutional Neural Network". *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. DOI: 10.1109/IGARSS47720.2021.9553206.

Murashkin, Dmitrii, Gunnar Spreen, Marcus Huntemann, and Wolfgang Dierking (2018). "Method for detection of leads from Sentinel-1 SAR images". *Annals of Glaciology* 59. DOI: 10.1017/aog.2018.6.

Nagi, Anmol Sharan, Devinder Kumar, Daniel Sola, and K. Andrea Scott (2021). "RUF: Effective Sea Ice Floe Segmentation Using End-to-End RES-UNET-CRF with Dual Loss". *Remote Sensing* 13.13. DOI: 10.3390/rs13132460.

Nandan, Vishnu et al. (2017). "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. DOI: https://doi.org/10.1016/j.rse.2017.06.029.

Neumann, Thomas A. et al. (2019). "The Ice, Cloud, and Land Elevation Satellite – 2 mission: A global geolocated photon product derived from the Advanced Topographic Laser Altimeter System". *Remote Sensing of Environment* 233. DOI: https://doi.org/10.1016/j.rse.2019.111325.

Nicolaus, Marcel et al. (2022). "Overview of the MOSAiC expedition: Snow and sea ice". *Elementa: Science of the Anthropocene* 10.1. DOI: 10.1525/elementa.2021.000046. eprint: https://online.ucpress.edu/elementa/article-pdf/10/1/000046/496065/elementa.2021.000046.pdf.

Notz, Dirk and M. Grae Worster (2009). "Desalination processes of sea ice revisited". *Journal of Geophysical Research: Oceans* 114.C5. DOI: https://doi.org/10.1029/2008JC004885. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JC004885.

Nvidia (n.d.). *20 Series Specs*.

Ochilov, Shuhratchon and David A. Clausi (2012). "Operational SAR Sea-Ice Image Classification". *IEEE Transactions on Geoscience and Remote Sensing* 50.11. DOI: 10.1109/TGRS.2012.2192278.

Onstott, Robert G. (1992). "SAR and Scatterometer Signatures of Sea Ice". *Microwave Remote Sensing of Sea Ice*. American Geophysical Union (AGU). Chap. 5. DOI: https://doi.org/10.1029/GM068p0073. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/GM068p0073.

Park, J., A. A. Korosov, M. Babiker, S. Sandven, and J. Won (2018). "Efficient Thermal Noise Removal for Sentinel-1 TOPSAR Cross-Polarization Channel". *IEEE Transactions on Geoscience and Remote Sensing* 56.3. DOI: 10.1109/TGRS.2017.2765248.

Park, J., J. Won, A. A. Korosov, M. Babiker, and N. Miranda (2019). "Textural Noise Correction for Sentinel-1 TOPSAR Cross-Polarization Channel Images". *IEEE Transactions on Geoscience and Remote Sensing* 57.6. DOI: 10.1109/TGRS.2018.2889381.

Park, J.-W., A. A. Korosov, M. Babiker, J.-S. Won, M. W. Hansen, and H.-C. Kim (2020). "Classification of sea ice types in Sentinel-1 synthetic aperture radar images". *The Cryosphere* 14.8. DOI: 10.5194/tc-14-2629-2020.

Radhakrishnan, Keerthijan, K. Andrea Scott, and David A. Clausi (2021). "Sea Ice Concentration Estimation: Using Passive Microwave and SAR Data With a U-Net and Curriculum Learning". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14. DOI: 10.1109/JSTARS.2021.3076109.

Rantanen, Mika et al. (2022). "The Arctic has warmed four times faster than the globe since 1980". *Communications Earth and Environment* 168. DOI: 10.1038/s43247-022-00498-3.

Ren, Yibin, Xiaofeng Li, Xiaofeng Yang, and Huan Xu (2022). "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. DOI: 10.1109/LGRS.2021.3058049.

Ren, Yibin, Huan Xu, Bin Liu, and Xiaofeng Li (2020). "Sea Ice and Open Water Classification of SAR Images Using a Deep Learning Model". *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. DOI: 10.1109/IGARSS39084.2020.9323990.

Ressel, Rudolf, Anja Frost, and Susanne Lehner (2015). "A Neural Network-Based Classification for Sea Ice Types on X-Band SAR Images". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8. DOI: 10.1109/JSTARS.2015.2436993.

Ressel, Rudolf, Suman Singha, Susanne Lehner, Anja Rösel, and Gunnar Spreen (2016). "Investigation into Different Polarimetric Features for Sea Ice Classification Using X-Band Synthetic Aperture Radar". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.7. DOI: 10.1109/JSTARS.2016.2539501.

Reynolds, Douglas A. (2009). "Gaussian Mixture Models". *Encyclopedia of Biometrics*.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. DOI: 10.48550/ARXIV.1505.04597.

Shorten, Connor and Taghi M. Khoshgoftaar (2019). "A survey on Image Data Augmentation for Deep Learning". *Journal of Big Data*.

Simonyan, Karen and Andrew Zisserman (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv: 1409.1556 [cs.CV].

Singha, Suman, Malin Johansson, Nicholas Hughes, Sine Munk Hvidegaard, and Henriette Skourup (2018). "Arctic Sea Ice Characterization Using Spaceborne Fully Polarimetric L-, C-, and X-Band SAR With Validation by Airborne Measurements". *IEEE Transactions on Geoscience and Remote Sensing* 56.7. DOI: 10.1109/TGRS.2018.2809504.

Skolnik, M.I. (1985). "Fifty years of radar". *Proceedings of the IEEE* 73.2. DOI: 10.1109/PROC.1985.13132.

SNAP - ESA Sentinel Application Platform (2022). Version 8.0.0.

Soh, Leen-Kiat and C. Tsatsoulis (1999). "Texture Analysis of SAR Sea Ice Imagery using Gray Level Co-occurrence Matrices". *IEEE Transactions on Geoscience and Remote Sensing* 37. DOI: 10.1109/36.752194.

Song, Haijun, David B. Kemp, Li Tian, Daoliang Chu, Huyue Song, and Xu Dai (2021). "Thresholds of temperature change for mass extinctions". *Nature Communications* 12.4694. DOI: 10.1038/s41467-021-25019-2.

Song, Wei, Minghui Li, et al. (2021). "Automatic Sea-Ice Classification of SAR Images Based on Spatial and Temporal Features Learning". *IEEE Transactions on Geoscience and Remote Sensing* 59.12. DOI: 10.1109/TGRS.2020.3049031.

Trenberth, Kevin E., John T. Fasullo, and Jeffrey Kiehl (2009). "Earth's Global Energy Budget". *Bulletin of the American Meteorological Society* 90.3. DOI: https://doi.org/10.1175/2008BAMS2634.1.

Ullah, Habib, Salman Khaleghian, Thomas Kromer, Torbjørn Eltoft, and Andrea Marinoni (2021). "A Noise-Aware Deep Learning Model for Sea Ice Classification Based on Sentinel-1 Sar Imagery". *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. DOI: 10.1109/IGARSS47720.2021.9553971.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods* 17. DOI: 10.1038/s41592-019-0686-2.

W.F. Weeks and S.F. Ackley, U.S. Army Cold Regions Research and Engineering Laboratory (1982). "The Growth, Structure and Properties of Sea Ice". *CRREL Monograph 82-1*.

Wahlström, Niklas, Thomas B. Schön, and Marc Peter Deisenroth (2014). *Learning deep dynamical models from image pixels*. arXiv: 1410.7550 [stat.ML].

Wang, Lei, K. Andrea Scott, and David A. Clausi (2017). "Sea Ice Concentration Estimation during Freeze-Up from SAR Imagery Using a Convolutional Neural Network". *Remote Sensing* 9.5. DOI: 10.3390/rs9050408.

Wang, Y.-R. and X.-M. Li (2021). "Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning". *Earth System Science Data* 13.6. DOI: 10.5194/essd-13-2723-2021.

Way, M. J. et al. (2016). "Was Venus the first habitable world of our solar system?" *Geophysical Research Letters* 43.16. DOI: https://doi.org/10.1002/2016GL069790. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL069790.

Webster, Melinda A., Ignatius G. Rigor, Donald K. Perovich, Jacqueline A. Richter-Menge, Christopher M. Polashenski, and Bonnie Light (2015). "Seasonal evolution of melt ponds on Arctic sea ice". *Journal of Geophysical Research: Oceans*. DOI: 10.1002/2015JC011030.

Wei, Longhui, Shiliang Zhang, Wen Gao, and Qi Tian (2017). "Person Transfer GAN to Bridge Domain Gap for Person Re-Identification". *CoRR* abs/1711.08565. arXiv: 1711.08565.

Weiss, Karl, Taghi Khoshgoftaar, and DingDing Wang (2016). "A survey of transfer learning". *Journal of Big Data* 3. DOI: 10.1186/s40537-016-0043-6.

Woodhouse, I.H. (2017). *Introduction to Microwave Remote Sensing*. CRC Press. DOI: 10.1201/9781315272573.

Yu, Qiyao, C. Moloney, and F.M. Williams (2002). "SAR sea-ice texture classification using discrete wavelet transform based methods". *IEEE International Geoscience and Remote Sensing Symposium*. Vol. 5. DOI: 10.1109/IGARSS.2002.1026863.

Zakhvatkina, Natalia Yu., Vitaly Yu. Alexandrov, Ola M. Johannessen, Stein Sandven, and Ivan Ye. Frolov (2013). "Classification of Sea Ice Types in ENVISAT Synthetic Aperture Radar Images". *IEEE Transactions on Geoscience and Remote Sensing* 51.5. DOI: 10.1109/TGRS.2012.2212445.

Zhang, Daniel et al. (2022). *The AI Index 2022 Annual Report*. arXiv: 2205.03468 [cs.AI].

Zhou, Kang, Shenghua Gao, et al. (2019). "Sparse-GAN: Sparsity-constrained Generative Adversarial Network for Anomaly Detection in Retinal OCT Image". *CoRR* abs/1911.12527. arXiv: 1911.12527.

Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang (2018). "Unet++: A Nested U-Net Architecture for Medical Image Segmentation". *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer.

— (2019). "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation". *IEEE Transactions on Medical Imaging*.