A deep neural network for lake ice detection with Sentinel-1 data

Anneloes Noordhoff

July 1, 2022

Universiteit Utrecht

Abstract

Ice cover of lakes is an indicator of climate conditions and possible changes thereof. It is therefore identified as an essential climate variable, and tracking its worldwide timing, duration and extent is important. Due to the vast number of lakes on Earth however, it can be difficult to find efficient ways to continuously monitor the formation, duration and break-up of lake ice. Remote sensing can be a useful tool in that regard, but optical passive remote sensing can be hindered by the presence of clouds or night-time. In this study, the use of synthetic aperture radar (SAR) imagery is therefore proposed, an active system that can penetrate clouds and works both day and night. Because ice conditions can vary strongly through space and time, a fully convolutional network (FCN) is constructed. This deep learning network is specifically designed for semantic image segmentation: learning patterns from large amounts of data and assigning labels to each pixel in the imagery. The model is trained on four study areas from different parts of the world, and overall results show a mean accuracy of >80%. Predictions are better for non-frozen conditions (~90%) compared to frozen conditions (~72%). Slight overfitting of the data indicates that the use of additional study areas may be required to optimize model performance, but the overall results are promising and demonstrate the usefulness of its application in worldwide lake ice monitoring.

Contents

List of figures

List of tables

1. Introduction

With a changing climate come rising temperatures, not only of air but also of water. Freshwater lakes around the globe have experienced a rise in average surface water temperatures, which together with increasing ambient temperatures can lead to important changes in the lakes' ecosystems (O'Reilly et al., 2015). One directly observable effect is seen in the ice seasonality of lakes, with strong reduction of yearly ice cover due to later ice-on and earlier ice-off trends (Sharma et al., 2021). In the last 25 years, these trends have also been six times faster than in the 75 years before (Sharma et al., 2021). Because of the sensitivity of lake ice cover to changes in atmospheric fluxes, it has been identified as an essential climate variable, asserting the importance to track worldwide changes in timing and extent of lake ice (Ma et al., 2021).

1.1 Background

Studies that have looked into lake ice cover trends have generally used data based on visual ground observations and/or optical (satellite) imagery (e.g. Heinilä et al., 2021; Tom et al., 2018). Although useful, ground observations can be costly and time-consuming, especially considering the many lakes in remote areas. Worldwide there are an estimated 117 million lakes, of which for example about 880.000 are located in the high latitudes of Canada (Messager et al., 2016; Verpoorter et al., 2014). Optical remote sensing is more efficient in that regard, but is often hindered by the presence of clouds, low temporal/spatial resolution, or due to long winter nights and low sun elevation at high latitudes (Barbieux, Charitsi, & Merminod, 2018; Surdu et al., 2015; Tom et al., 2020).

To overcome these limitations, active remote sensing systems using synthetic aperture radar (SAR) can be used. Microwave SAR works day and night and can penetrate clouds, thus providing year-round, all weather images with a high resolution. Different types of sensors have been or are currently carried on multiple spaceborne SAR systems, such as Sentinel-1, RADARSAT or ALOS-PALSAR (Singha et al., 2018), creating a large available dataset.

The classification of ice from satellite imagery can be a challenging one. Although the availability of SAR data greatly enhances the possibilities due to its independence of cloud cover, properties of ice and water can vary through time and space, making linear classification often difficult (Lindenschmidt & Li, 2019; Tom et al., 2018). Ice forms differently under different circumstances, thickening and changing throughout the winter. Water can contain various amounts of sediments of plants, or can have ripples due to wind. These factors all affect how a radar signal is backscattered. However, the application of deep learning models in this field has been promising. Studies implementing convolutional neural networks for both sea ice detection (Boulze & Korosov, 2020; Wang & Li, 2020) and lake ice detection (Dirscherl et al., 2021; Ma et al., 2021; Scott et al., 2020; Tom et al., 2020) have shown good results in making a distinction between water and ice.

Convolutional neural networks (CNNs) are artificial neural networks that focus on pattern recognition within images (O'Shea & Nash, 2015). Initially developed to mimic human vision, they simulate processing of visual input in order to assign a label to an image. For the case of ice classification, a prediction for each pixel is required, meaning that the output size is the same as the size of the input image. One type of CNN that is able to assign labels per pixel is a Fully Convolutional Network (FCN). This type of network often uses the encoder-decoder architecture (EDA), where first the convolution network is applied and then a transpose-convolution network to upsample back to

the original map size (Long et al., 2015; Ronneberger et al., 2015; Xing et al., 2020). This type of architectures allows to take into consideration small and large features.

1.2 Objective and research question

Research performed using SAR data has often been on local scales and methods or models are therefore not (yet) suited for global applications (Dirscherl et al., 2021; Sobiech & Dierking, 2013; Tom et al., 2020). Furthermore, many of the lake ice detection studies have been performed using classical CNN models. Therefore, this paper aims to determine lake ice cover at multiple locations around the world using a fully convolutional network and basic Sentinel-1 SAR data. The main research question is: *how can lake ice cover be determined from SAR imagery using a deep learning fully convolutional network?*

2. Data and methods

2.1 Study area

This study will attempt to detect lake ice in various regions of the world using a deep learning model. Therefore, several study areas were selected that were used in training of the model. This selection was done following several criteria:

- Study areas are 10x10 km.
- Study areas contain at least 1 complete lake.
- Study areas contain a lake of at least 2x2 km.
- Study areas must contain lakes that are frozen continuously for more than 2 months per year.

Based on these criteria, four study areas were selected in Alaska (US), Canada, Finland and Russia (Figure 1). Besides the given criteria, the selection was done arbitrarily.

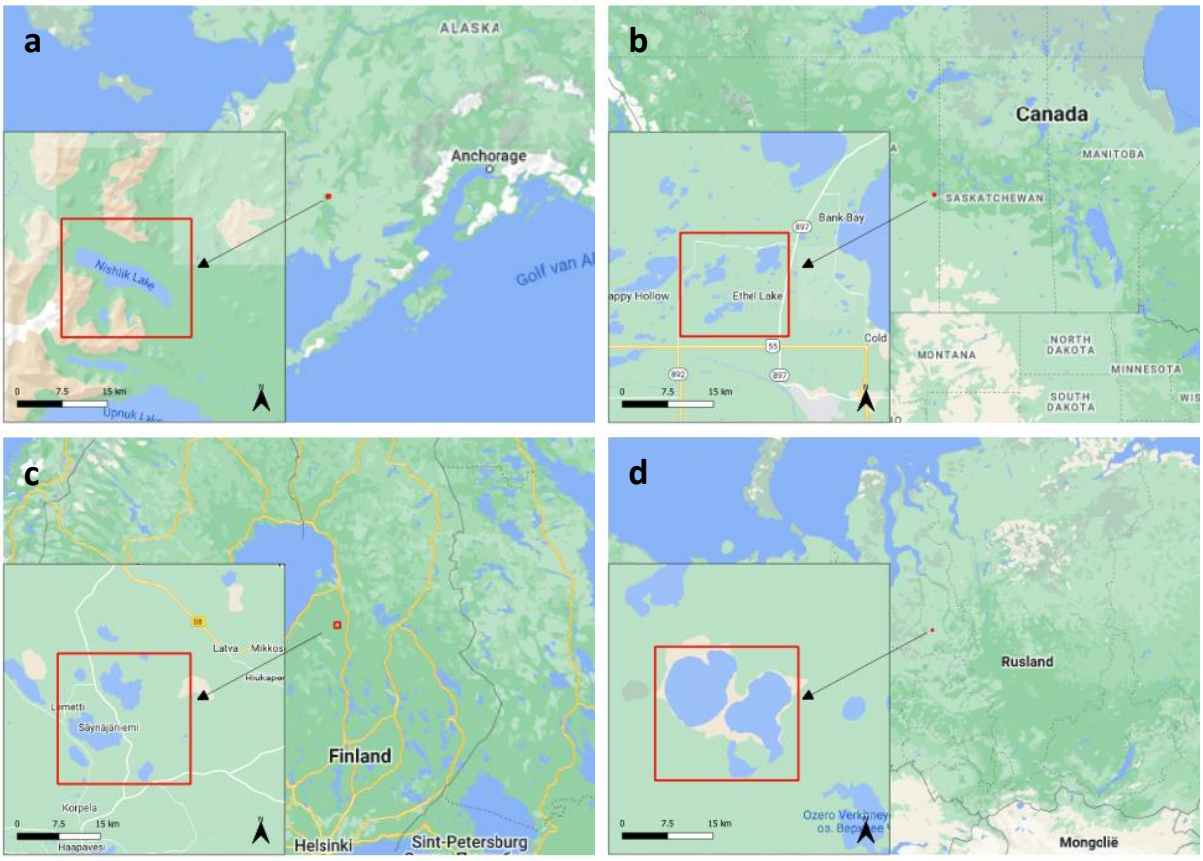

Figure 1. Selected study areas in a) Alaska (US), b) Canada, c) Finland and d) Russia.

2.2 Synthetic aperture radar

A synthetic aperture radar is an active remote sensing system that uses radio- and microwaves for sensing. As opposed to optical passive sensors that operate in the visible and infrared part of the spectrum, SAR can penetrate through clouds and vegetation, and is not affected by the time of day. This makes it a useful instrument for studies where darkness or clouds can hinder acquisition (Podest, 2017).

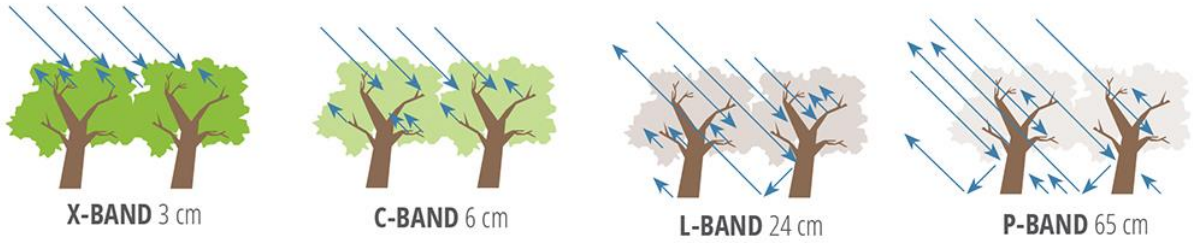

Figure 2. Different wavelengths from different bands give stronger or shallower penetration (from: https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar).

SAR emits electromagnetic waves that are backscattered and collected by the radar antenna. How much of the signal it backscattered depends on the polarization and wavelength of the signal, and the surface it reflects of. Polarizations can be transmitted or received vertically (V), horizontally (H) or a combination of both, with each enabling the detection of different physical properties of an object. Wavelength will determine the depth of penetration into a surface. SAR can use various wavelengths, which are often referred to as letter indicated bands, such as C-band (~6 cm) or L-band (~23 cm). With longer wavelengths, the penetration becomes stronger (Figure 2). How a signal backscatter depends largely on the roughness of the surface. When a surface is smooth, reflection will be specular, meaning that most of the signal is returned away from the sensor, while rough surfaces with result in diffuse scattering, with the signal going into all directions (Figure 3) (Flores-Anderson et al., 2019; Moreira et al., 2013).

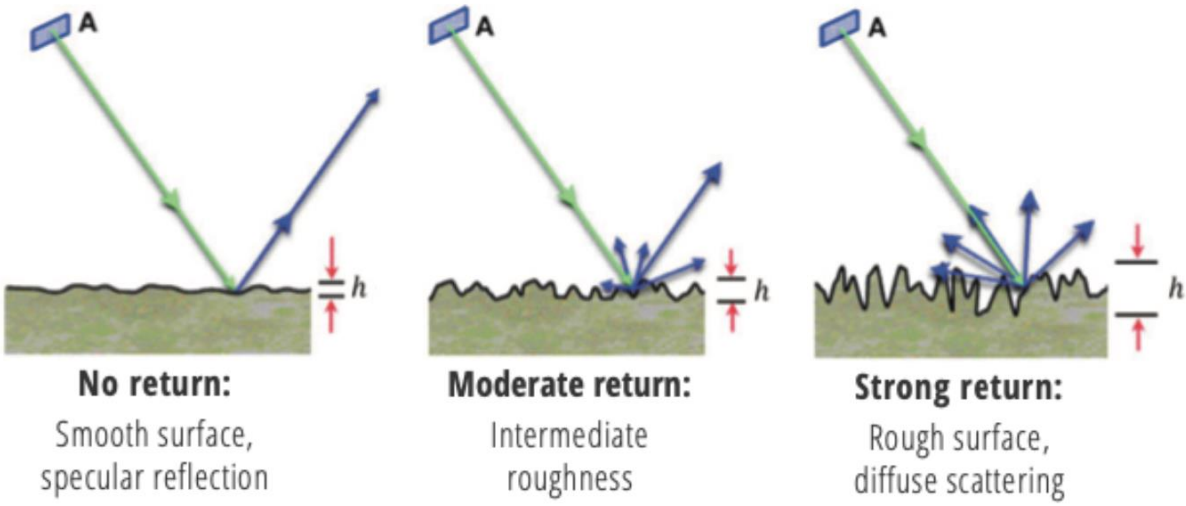

Figure 3. Backscatter mechanisms for different surface types, with h as an indicator for surface roughness (adapted from Flores-Anderson et al., 2019)

2.3 Deep learning model

Deep learning methods for computer vision have been rapidly developing over de last decades and the implementation of convolutional neural networks played an important role in the advancement of object detection and semantic segmentation. Initially developed to mimic human vision, they simulate the processing of visual input in the brain in order assign a label to an image. Semantic segmentation of images is the task of classifying different elements of an image. This makes it useful in lake ice detection, as parts of a lake may be frozen or unfrozen. Convolutional neural

networks are the most common models used for semantic image segmentation. Specifically, a fully convolutional neural network is a deep learning method where prediction is done on a pixel basis, and the output layer has the same size as the input layer (Ronneberger et al., 2015).

An FCN is a CNN with an encoder-decoder structure. Its basic architecture consists of convolutional layers in which kernels (filters) convolve the input data to new values, pooling layers in which the dimensionality of the new values are reduced, and fully-connected layers in which the final calculation for the classification is carried out. The FCN model used in this study is the U-Net model, originally designed by Ronneberger et al. (2015) (Figure 4). The first part of the model has a general CNN architecture, where convolutional layers extract features from the image with 3x3 filters with Rectified Linear Unit (ReLU) activation, while also reducing the amount of data with max pooling layers. The convolution starts with 32 filters, which are doubled after every convolution block. The max pooling has a stride length of 2x2, reducing the output by factor 2 each block. The second part of the model consists of upsampling blocks that double the size of the output at each block, eventually returning to the original image size (Dirscherl et al., 2021; Ronneberger et al., 2015).

When upsampling from the smallest resolution, information loss can occur. At every upsampling block there is therefore additional input from its downsampling counterpart to improve information availability throughout the network and counter information loss (so-called skip connections). Furthermore, at every block batch normalization is applied, as well as a dropout of 0.25.

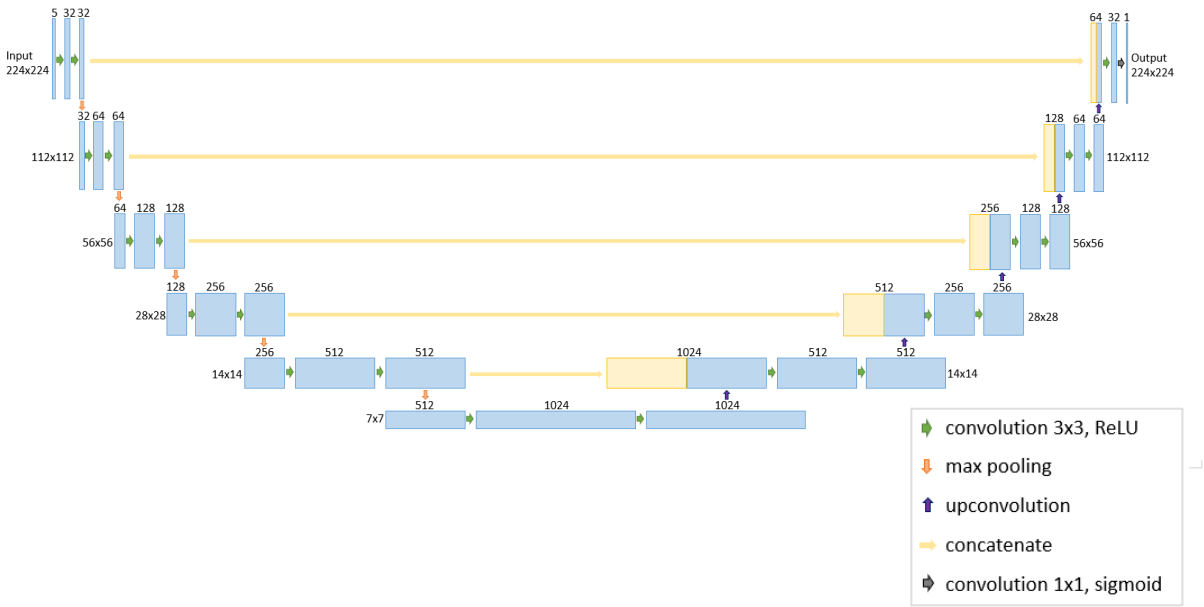

Figure 4. U-Net model architecture as used in this study.

2.4 Data pre-processing

The SAR data used in this study is captured by the two Sentinel-1 satellites, in orbit since April 2016. These satellites have sun-synchronous, near-polar orbit with a 12-day repeat cycle for each satellite. Because they have a 180° orbital phasing difference, both satellites together have a repeat cycle of six days. Sentinel-1 carries a C-band SAR instrument, operating with a wavelength of ~5.5 cm, with dual polarization (HH+HV, VV+VH) (ESA, 2022). It operates in various acquisition modes, of which the Interferometric Wide Swath Mode (IW) is most often used for land cover classification outside of

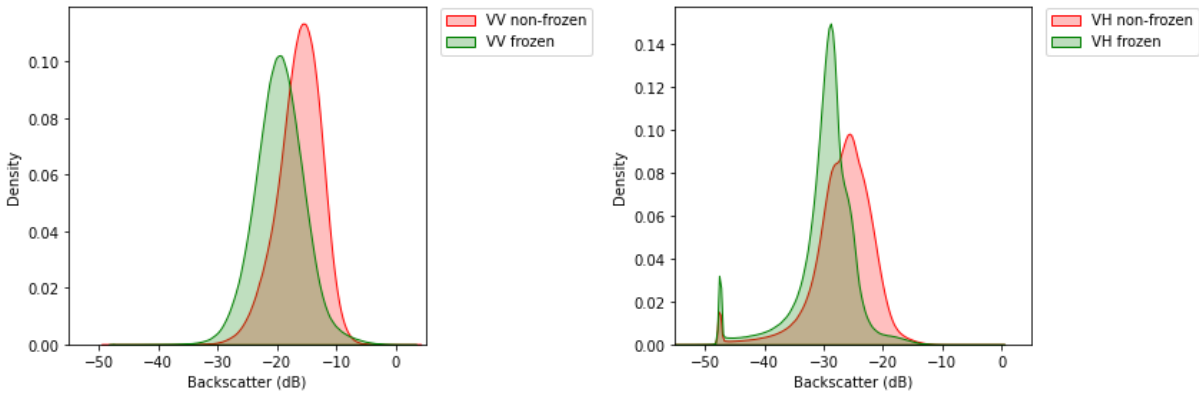

Figure 5. Density plots of VV (left) and VH (right) polarizations for frozen and non-frozen lake conditions (training data).

polar zones, with an incidence angle between 29-46°, a swath width of 250m and a resolution of 5-20m. Flight direction can be ascending or descending (ESA, 2022).

SAR data is freely available and easily accessed in Google Earth Engine (GEE). In this study, the first part of the data pre-processing was therefore done using the GEE platform. Data was collected for the years 2017-2021. The SAR Ground Range Detected data product with the IW swath mode was collected for each study area, with an ascending flight path and VV+VH dual polarization. The VV and VH backscatter signals in the study areas were the foundation of the dataset on which this study was performed. The difference in backscatter between frozen and non-frozen circumstances (Figure 5) form the basis of the hypothesis that machine learning models can detect and classify these classes. Additionally, incidence angle of the sensor was also added, as this can have an effect on the strength of the return signal due to different reflection mechanisms (Ma et al., 2021). In total, 420 images were collected for the four study areas for 5 summer and 5 winter seasons.

To determine whether lakes are frozen or unfrozen during a certain time of year, optical satellite imagery from Planet was used. The satellite constellation of Planet (called PlanetScope) consists of over 130 nanosatellites that cover the entire globe on a daily basis with a resolution of about 3m (Planet, 2022). The PlanetScope 4-band (RGB-NIR) multispectral scenes were used for visual inspection of the study areas. For every year from 2017 to 2021, ice-on and ice-off days were determined. It was assumed that between the first full ice-on day and the first ice-off day, the lakes remained completely frozen, while between the first full ice-off day and the first ice-on day, the lakes are unfrozen. Break-up or freeze-up periods are therefore not considered. Although it was attempted to classify lake ice

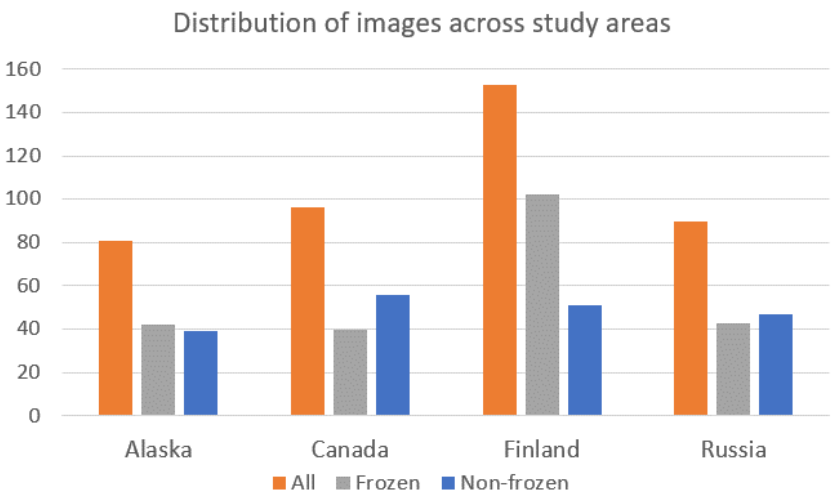

Figure 6. Number of images per category per study area.

based on a near-infrared threshold, this threshold seemed to differ strongly over time and space. Manually identifying frozen and unfrozen parts during transition periods for each study area for each year is too time-intensive, which led to the decision to leave these periods out of the dataset.

Lake polygons were downloaded from HydroSHEDS (Messager et al. , 2016) and uploaded to GEE. These polygons were used to create part of the dataset that made the distinction between lakes and other surfaces (e.g. land or sea). Furthermore, they were used to define the lake areas that were either frozen or unfrozen, in order to create labels for the training, test and validation datasets.

Because the backscatter of the SAR signal on water can be affected by wind (waves), wind speed and direction were also taken into account. The dataset was supplemented with hourly wind speeds in directions u and v (orthogonal) from the ERA5 Reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Muñoz Sabater, 2019), which is also available through the GEE platform. In addition to the VV+VH polarizations and incidence angle, both wind components were added to the dataset for analysis, with a resolution of ~11 km.

The final dataset consisted of 420 images of 5 bands (VV, VH, incidence angle, u-wind, v-wind). In addition, information about lake location and frozen/non-frozen labels were added. The distribution of the images across the study areas is given in Figure 6. For Finland, the number of images is higher, mainly due to a higher number of frozen images. This point will be addressed in the discussion. The entire dataset was exported in GeoTIFF format with dimensions of 2016x2016 pixels to a Google Cloud Storage Bucket. This has the advantage of being accessible from both GEE and Google Colab, where the second part of the study was performed.

2.5 Model training

Google Colab is a Jupyter notebook environment that runs entirely in the cloud, with access to GPU for faster training of machine learning models. Due to a 12GB RAM limitation however, not all 420 images could be downloaded into Colab's virtual environment. Therefore, an arbitrary selection was made, consisting of 130 images that were divided into a 60% training data (78 images, Figure 7), 20% validation data (26 images) and 20% test data (26 images). All datasets had about a 50/50 balance of frozen/non-frozen images. Due to the RAM limitations during model training, each GeoTIFF of 2016x2016 pixels were split into 81 images of 224x224 pixels. This gave a total of about 6300 training images and 2x2100 test or validation images.

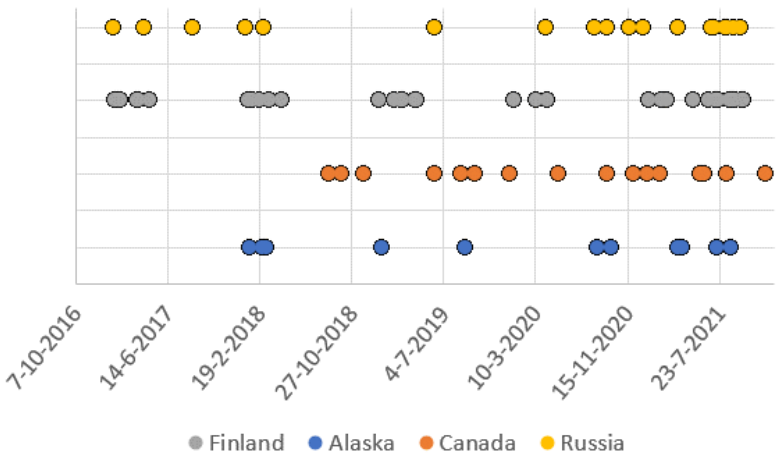

Figure 7. Distribution of selected train imagery over time per study area.

The model was compiled with a binary cross-entropy loss function, an Adam optimizer and a learning rate of 0.0005. Training was done in batches of 32 images at the time, giving an input size of (32x224x224x5), of which 5 denotes the number of layers (variables) in each image. Training was done for 30 epochs at a time, after which the model was saved. This was done to prevent early termination due to GPU time limitation in Google Colab. After 150 epochs the model showed some overfitting and training was terminated.

2.6 Performance measures

To check model performance, the results were analyzed with several metrics, all based solely on pixels within lake polygons. Because the classification is binary and the datasets are assumed to be relatively balanced (about a 50/50 division between frozen and non-frozen pixels), first of all accuracy was determined. Accuracy is a metric that determines the percentage of correctly classified pixels as:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{1}$$

where tp is true positive, tn is true negative, fp is false positive and fn is false negative (Dalianis, 2018). Because accuracy can be biased due to imbalanced data, a confusion matrix will be constructed to inspect the performance for both classes individually. In addition to that, the F1-score will be determined, to assert that a slight imbalance in the class distribution does not alter the performance value. The F1-score is a metric that calculates a weighted combination of precision and recall, determined by (Dalianis, 2018):

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{2}$$

Because a large part (290) of the collected images was not available for model training due to GPU limitations, these were used for an additional analysis. For this, the model was slightly altered to enable the ice prediction on complete images (dimension 2016x2016 pixels). For each image the accuracy was determined and then pooled into an average accuracy over all images. These results were used to check the robustness of the model for each study area and for either of the two lake conditions.

3. Results

In this chapter, model results on both the test dataset and additional imagery is presented. The train dataset consisted of about 146 million pixels, including non-lake pixels. The total number of lake pixels on which Table 1 is based is about 58.5 million. The training dataset shows both an accuracy and F1-score of 99.4%. Comparing these values to the test dataset, the model may be slightly overfitted to the training data. However, also the test data has good performance, as shown in Table 2. The test dataset consisted of about 52 million pixels, with a total number of lake pixels on which the statistical analysis was performed of about 21.5 million, of which 46.7% was frozen and 53.3% non-frozen. Overall, the model prediction on the test dataset has an accuracy of 83.3% and an F1-score of 82.4%.

Table 1.Confusion matrix, accuracy and F1-score for training dataset. All values are in percentage.

Prediction / True	Non-frozen	Frozen
Non-frozen	45.26	0.01
Frozen	0.61	54.12

Accuracy	F1-score
99.38	99.43

Table 2. Confusion matrix, accuracy and F1-score for test dataset. All values are in percentage.

Prediction / True	Non-frozen	Frozen
Non-frozen	44.11	9.19
Frozen	7.56	39.14

Accuracy	F1-score
83.25	82.37

To inspect the model performance more extensively, the 290 additional complete images were used for prediction and analysis. For the prediction, the accuracy was calculated per study area for all images and per category (Table 3 and Figure 8). The accuracy ranges from 0 to 100% for every category, but the accuracy of the model on all additional images was 80.3% on average and performs better on the non-frozen imagery (89.5%) than on the frozen imagery (72.4%). Also, the standard deviation of the accuracy for the frozen imagery is higher than for non-frozen. Overall, the model performs best on the Finland study area, although the non-frozen lakes in Canada have the highest

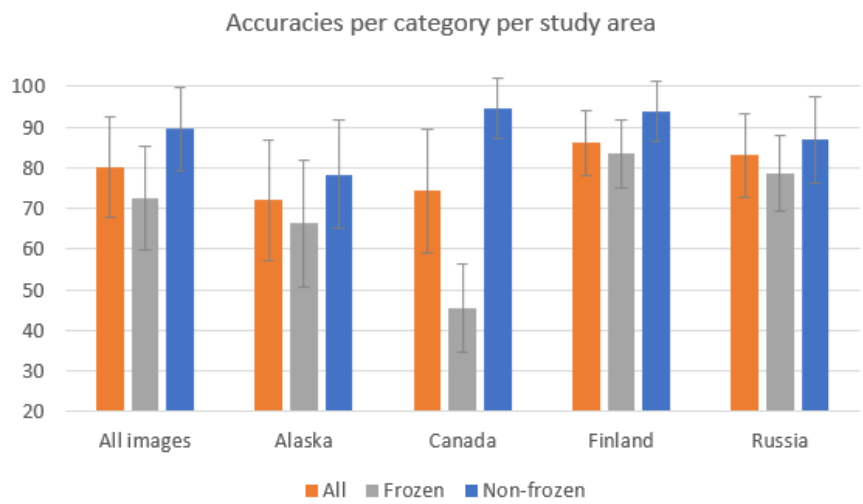

Figure 8. Mean accuracy (%) of all images per category for each study area. Vertical bars indicate standard deviation.

Table 3. Statistics of images per category per study area for 290 additional images. Count and % of all images indicate contribution of given categories to the original training data. Shading indicates lowest (orange) and highest (green) values per column.

		All	Frozen	Non-frozen
All images	count in training data	78	43	35
	% of all images	100	55.1	44.9
	mean accuracy (%)	**80.3**	**72.4**	**89.5**
Alaska	count in training data	11	6	5
	% of all images	14.1	12.8	14.3
	mean accuracy (%)	**72.1**	**66.2**	**78.4**
Canada	count in training data	16	6	10
	% of all images	20.5	12.8	28.6
	mean accuracy (%)	**74.3**	**45.4**	**94.7**
Finland	count in training data	33	23	10
	% of all images	42.3	48.9	28.6
	mean accuracy (%)	**86.2**	**83.5**	**93.8**
Russia	count in training data	18	8	10
	% of all images	23.1	17.0	28.6
	mean accuracy (%)	**83.0**	**78.5**	**87.0**

mean accuracy (94.7%). However, on frozen lakes in Canada the model performs the worst (45.4%), lower than from a random binary draw. Alaska has the lowest mean accuracy for both the non-frozen images (78.4%) and the total set of images (72.1%).

Table 3 also shows the number and percentage of images per category for each study area. It can be seen that low and high mean accuracies often coincide with the lowest and highest contribution of images to the total dataset. This will be further discussed in the next chapter. More detailed statistics can be found in Appendix A.

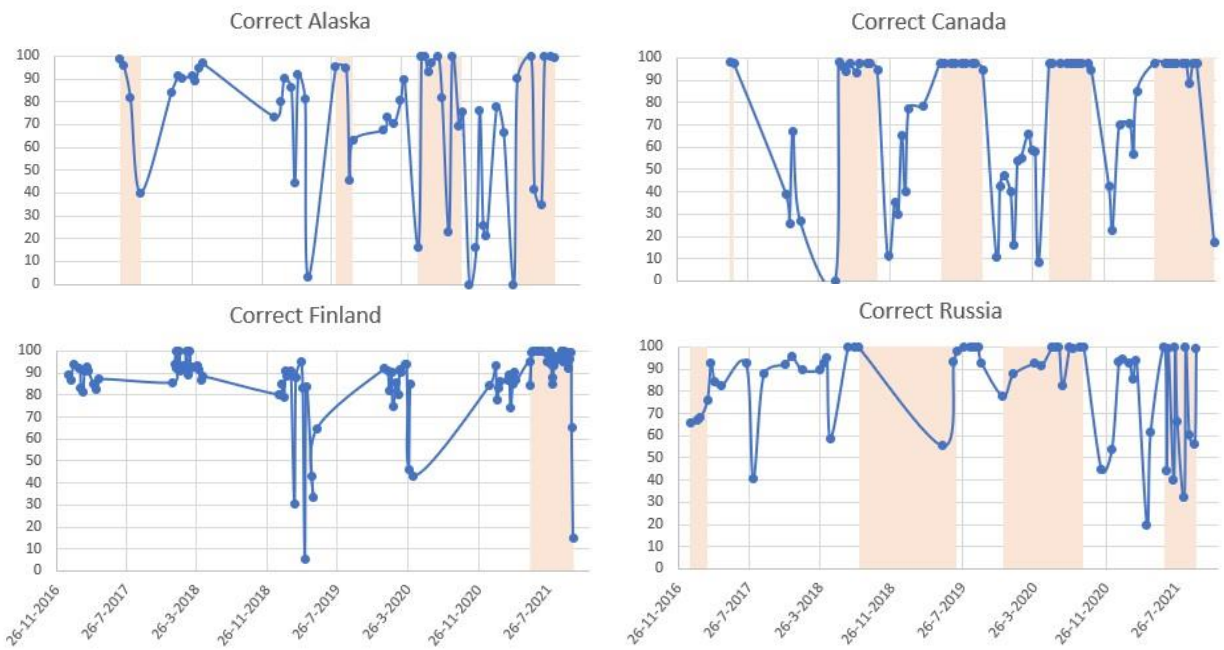

Figure 9. Accuracy for imagery over time for each study area. Shaded areas indicate non-frozen imagery (summer periods).

Accuracies over time per study area for the additional imagery is show in Figure 9. It clearly shows that non-frozen images from the Canada study area have high accuracies. For Finland, accuracy in summer is also high (Table 3), but the figure shows that this all stems from the 2021 summer period. Alaska shows large variation, although summer imagery seems slightly better. Russia has overall quite high accuracies, which also shows from the average accuracies from Table 3.

Examples of the qualitative results are shown in Figure 10 for better interpretation of the model outcome. For each study area a prediction for a frozen lake image is given, with the examples selected based on the accuracy of the prediction being similar to the average accuracy of the total frozen imagery dataset per study area. These accuracies in fact range from 0 to 100%, with the standard deviation shown in Figure 8. The VV and VH band show that that there are patterns visible on the lake surface in frozen conditions that may explain some of the prediction, but they don't seem to dominate it. In Appendix B examples of non-frozen image predictions are given.

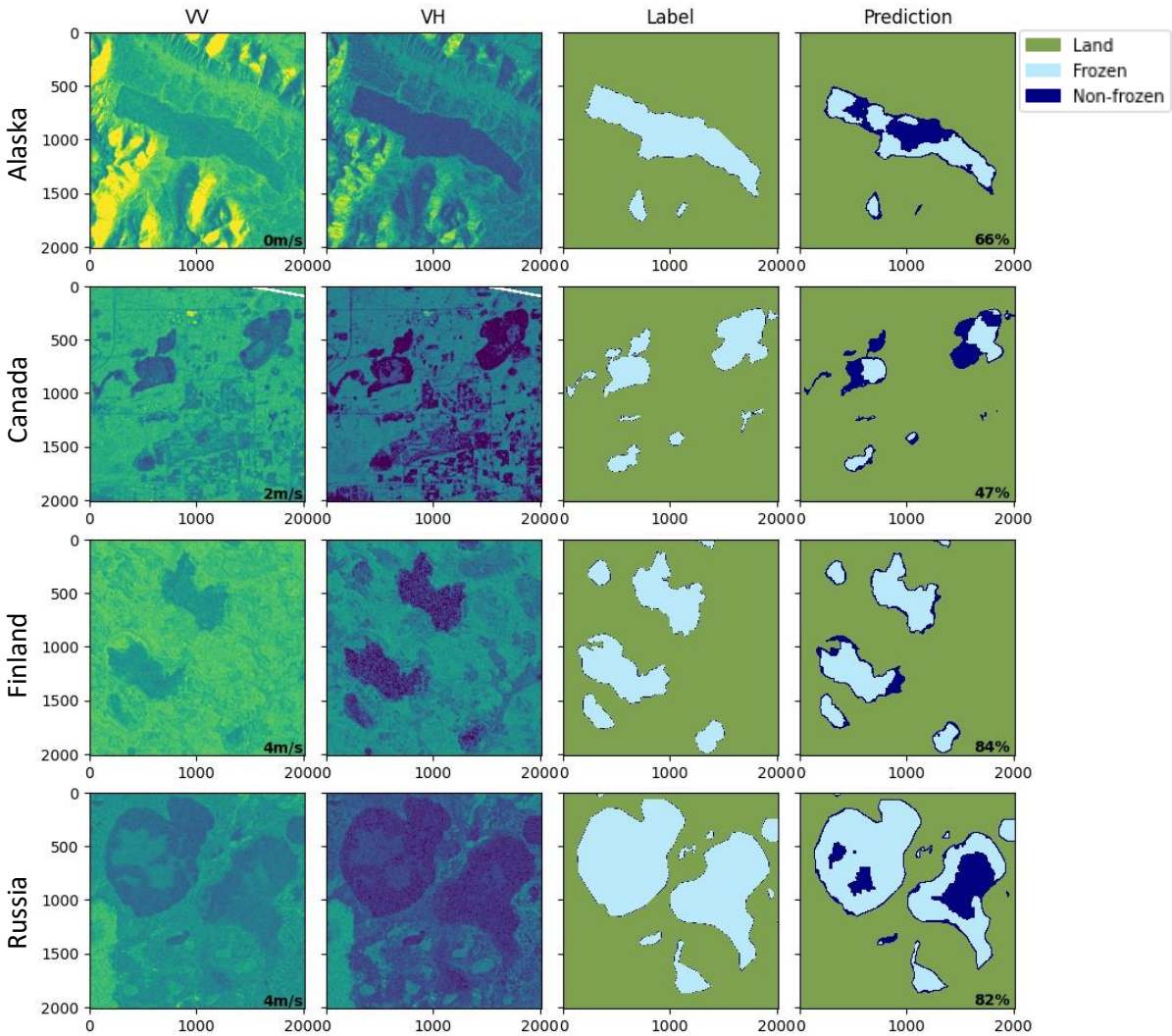

Figure 10. Qualitative results for frozen lake imagery for Alaska (19-03-2021), Canada (20-12-2019), Finland (10-04-2019) and Russia (23-04-2017). The first two columns show Sentinel-1 SAR VV and VH values (color range -30dB – 0 dB) including average windspeed, the third column the label (frozen) for the lake polygons and the fourth column the model prediction including accuracy (%).

4. Discussion

Fully convolutional neural networks have previously been shown to be an excellent tool in semantic image segmentation. Furthermore, Sentinel-1 SAR imagery forms a valuable addition to traditional optical remote sensing data, which can be hindered by cloud presence or nighttime. The use of Sentinel-1 SAR imagery in combination with data-driven deep learning models for image classification therefore seems to have high potential, which was explored in this research. Here, interpretation of the results and corresponding discussion points are addressed.

4.1 Dataset imbalance

The FCN designed in this study for lake ice detection using Sentinel-1 SAR shows promising results, although they do show that there can be a large variation in the model performance both spatially and temporally. There are several factors that contribute to this variation, of which the training dataset distribution may be the most important one. Although there was a balance between the frozen and non-frozen images in the training dataset, the distribution of images between study areas was unproportionate, as can be seen in Figure 11, as well as in Table 3. With 4 study areas and 2 classes, ideally each category covers about 12.5% of the entire training dataset, giving each study area 25% total. The Finland study area is however overrepresented mainly in the frozen domain (24.5%), which caused it to have the highest accuracy for the images with lake ice. However, there was also a high accuracy in the non-frozen domain, while the number of images there were not proportionally high (12.4%). The same goes for the Canada non-frozen category, which had the highest accuracy (94.7%), but contributed relatively equally to the training dataset (14.1%). This indicates that good model performance partly stems from overrepresentation, but that the model also has been able to learn from the equally distributed data in mainly the non-frozen domain. Balancing the dataset better could therefore lead to better results primarily in the now low-scoring categories.

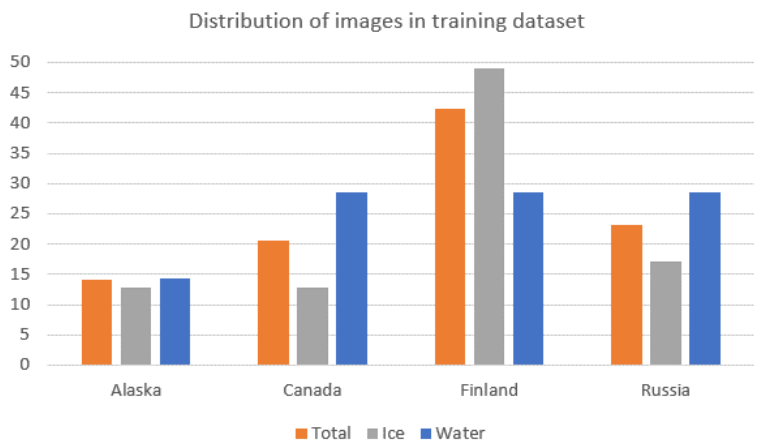

Figure 11. Number of images in the training dataset per category per study area.

4.2 Model set-up

The U-Net FCN model was used with its original architecture and training was done with a single set of hyperparameters. In machine learning models, optimization of hyperparameters is often crucial in creating a well performing model (Yang & Shami, 2020). Due to time limitations however, optimization of for example number of convolution blocks, number of filters, size of filters or the

activation function was not possible. The number of epochs can sometimes also influence model performance, but with the current dataset it was clear it was overfitting after a certain amount of time and further training of the model would not lead to better results on the validation and/or test dataset. Further research could look into improving mentioned hyperparameters, or even the effect of altering the architecture.

4.3 Training and prediction

Overall, the accuracies from the model predictions were much higher (89.5%) than the frozen domain (72.4%). This can indicate that it is easier for the model to distinguish water using SAR imagery than ice, a result often seen in ice classification with data-driven models (Ma et al., 2021; Tom et al., 2020). This can be explained by the fact that the frozen state of lakes does not exist of a single type of surface cover. The Canadian Ice Service (CIS, 2005) reports over 10 different types of ice, with additional changes over time due to age and deformations. Furthermore, ice can be covered by (various types of) snow, further changing its reflective properties. Although not feasible within the scope of this study, this limitation may be addressed by changing the image classification from a binary one to a multi-label one, including the different ice categories. For sea ice classification such studies have been done (e.g. Boulze & Korosov, 2020; De Gelis et al., 2021; Khaleghian et al., 2021), but for lake ice types they have yet to be carried out.

The five variables used for prediction (VV, VH, incidence angle, u-wind component and v-wind component) were considered to be the main influencers of backscatter under frozen or non-frozen conditions. Other studies using SAR to classify ice sometimes only use 2 polarizations for training (Boulze & Korosov, 2020; Stonevicius et al., 2022; Tom et al., 2020) or have added derivates such as mean values or ratios (Dabiri et al., 2021). Most comparable studies have better model results, independent of the number of variables used. It is therefore thought that the variable selection is not the main driver of the current results, but that the difference in data and model set-up is. However, including additional variables that assert certain spatial or temporal relationships may improve model performance.

Further corrections to the data, such as speckle removal or radiometric corrections, were not applied, because those would require additional assumptions about the data, which was not desired. Under the objective of the study, it was implied that basic SAR data would be used, from which the model would learn the relevant features. However, applying some corrections may benefit model training, as noise may be filtered out and relevant features would become more pronounced.

4.4 Study areas

In this study it was attempted to create a deep learning model that could predict the frozen or non-frozen state of different lakes around the world based on SAR imagery. Four studies areas from four different places were therefore used for training and analysis of model performance. However, it is unknown how the model performs on lakes outside the selected study areas. Lower model performance on underrepresented areas in the training data might indicate that overfitting occurs not only on the training data, but on the four selected study sites in general. Further research could therefore look into adding training data from additional locations around the world and testing on another variety of different places. This may improve the model robustness and the usefulness of its application in lake ice and climate research all around the world.

5. Conclusion

This study created a deep learning FCN model for lake ice prediction using Sentinel-1 SAR data. For this, a model with U-Net architecture was trained on four study areas in Alaska, Canada, Finland and Russia, that had continuous ice cover for several months each year. The performance of the model was assessed using accuracy and F1 metrics, and because the dataset was relatively balanced, accuracy was considered a good indicator. On a test dataset the model reached an accuracy of 83%, while for the train dataset it reached over 99%. This is an indication that the model was overfitted on the training data. However, performance on an additional dataset still shows relatively high accuracies ranging from 72.4% (frozen) to 89.5% (non-frozen) for the two different classes. Between study areas, there was some variation, mainly due to an imbalance in the distribution of the data between different study areas.

Overall, the FCN seems like an adequate tool for semantic segmentation of lake ice imagery. Although the model results in this study are not as high as comparable studies, there are several improvements that are believed would increase model performance, such as better distribution of the training data and incorporating data of different study areas. Furthermore, model hyperparameter optimization might also improve the predictive performance. Further research can look into the application of these enhancements and possibly expand the workflow for easier predictions in new study areas.

References

Barbieux, K., Charitsi, A., & Merminod, B. (2018). Icy lakes extraction and water-ice classification using Landsat 8 OLI multispectral data. *International Journal of Remote Sensing*, *39*(11), 3646–3678. https://doi.org/10.1080/01431161.2018.1447165

Boulze, H., & Korosov, A. (2020). Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks, 1–20. https://doi.org/10.3390/rs12132165

CIS. (2005). *Manual of Standard Procedures for Observing and Reporting Ice Conditions."*.

Dabiri, Z., Hölbling, D., Abad, L., & Guðmundsson, S. (2021). Comparing the applicability of sentinel-1 and sentinel-2 for mapping the evolution of ice-marginal lakes in Southeast Iceland. *GI_Forum*, *9*(1), 46–52. https://doi.org/10.1553/GISCIENCE2021_01_S46

Dalianis, H. (2018). *Clinical text mining: Secondary use of electronic patient records*. *Clinical Text Mining: Secondary Use of Electronic Patient Records*. https://doi.org/10.1007/978-3-319-78503-5

De Gelis, I., Colin, A., & Longepe, N. (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*(Dl), 5831–5841. https://doi.org/10.1109/JSTARS.2021.3074068

Dirscherl, M., Dietz, A. J., Kneisel, C., & Kuenzer, C. (2021). A novel method for automated supraglacial lake mapping in antarctica using sentinel-1 sar imagery and deep learning. *Remote Sensing*, *13*(2), 1–27. https://doi.org/10.3390/rs13020197

ESA. (n.d.). Sentinel-1. Retrieved June 18, 2022, from https://sentinel.esa.int/web/sentinel/missions/sentinel-1

Flores-Anderson, A. I., Herndon, K. E., Thapa, R. B., & Cherrington, E. (Eds.). (2019). *The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. NASA. https://doi.org/10.25966/nr2c-s697

Heinilä, K., Mattila, O., Mets, S., Sakari, V., Schwaizer, G., & Koponen, S. (2021). A novel method for detecting lake ice cover using optical satellite data. *International Journal of Applied Earth Observations and Geoinformation*, *104*(October). https://doi.org/10.1016/j.jag.2021.102566

Khaleghian, S., Ullah, H., Kræmer, T., Hughes, N., Eltoft, T., & Marinoni, A. (2021). Sea ice classification of sar imagery based on convolution neural networks. *Remote Sensing*, *13*(9), 1–20. https://doi.org/10.3390/rs13091734

Lindenschmidt, K. E., & Li, Z. (2019). Radar scatter decomposition to differentiate between running ice accumulations and intact ice covers along rivers. *Remote Sensing*, *11*(3). https://doi.org/10.3390/rs11030307

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. https://doi.org/10.1109/ICCVW.2019.00113

Ma, Z., Liu, Z., Pu, J., Xu, L., Li, K., Wangqu, L., … Duguay, C. (2021). Deep convolutional neural network with random field model for lake ice mapping from Sentinel-1 imagery. *International Journal of Remote Sensing*, *42*(24), 9343–9367. https://doi.org/10.1080/01431161.2021.1995074

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 1–11. https://doi.org/10.1038/ncomms13603

Moreira, A., Prats-iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). SAR-Tutorial-March-2013. *IEEE Geoscience and Remote Sensing Magazine*, *1*(1), 6–43.

Muñoz Sabater, J. (2019). ERA5-Land hourly data from 1981 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., … Allan, M. G. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, *42*(10), 773–781. https://doi.org/10.1002/2015GL066235.Received

O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks, 1–11. Retrieved from http://arxiv.org/abs/1511.08458

Planet. (2022). PlanetScope. Retrieved June 18, 2022, from https://developers.planet.com/docs/data/planetscope/

Podest, E. (2017). Basics of Synthetic Aperture Radar. NASA Applied Remote Sensing Training (ARSET) Program.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1–8.

Scott, K. A., Xu, L., & Pour, H. K. (2020). Retrieval of ice/water observations from synthetic aperture radar imagery for use in lake ice data assimilation. *Journal of Great Lakes Research*, *46*(6), 1521–1532. https://doi.org/10.1016/j.jglr.2020.08.018

Sharma, S., Richardson, D. C., Woolway, R. I., & Imrit, M. A. (2021). Loss of Ice Cover , Shifting Phenology , and More Extreme Events in Northern Hemisphere Lakes, 1–12. https://doi.org/10.1029/2021JG006348

Singha, S., Johansson, M., Hughes, N., Hvidegaard, S. M., & Skourup, H. (2018). Arctic Sea Ice Characterization Using Spaceborne Validation by Airborne Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 1–20.

Sobiech, J., & Dierking, W. (2013). Observing lake- and river-ice decay with SAR : advantages and limitations of the unsupervised k -means classification approach. *Annals of Glaciology*, *54*(62), 65–72. https://doi.org/10.3189/2013AoG62A037

Stonevicius, E., Uselis, G., & Grendaite, D. (2022). Ice Detection with Sentinel-1 SAR Backscatter Threshold in Long Sections of Temperate Climate Rivers. *Remote Sensing*, *14*(7), 1627. https://doi.org/10.3390/rs14071627

Surdu, C. M., Duguay, C. R., Pour, H. K., & Brown, L. C. (2015). Ice freeze-up and break-up detection of shallow lakes in Northern Alaska with spaceborne SAR. *Remote Sensing*, *7*(5), 6133–6159. https://doi.org/10.3390/rs70506133

Tom, M., Aguilar, R., Imhof, P., Leinss, S., Baltsavias, E., & Schindler, K. (2020). Lake ice detection from sentinel-1 sar with deep learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*(3), 409–416. https://doi.org/10.5194/isprs-Annals-V-3-2020-409-2020

Tom, Manu, Kälin, U., Sütterlin, M., Baltsavias, E., & Schindler, K. (2018). Lake ice detection in low- resolution optical satellite images.

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, *41*(18), 6396–6402. https://doi.org/10.1002/2014GL060641

Wang, Y.-R., & Li, X. (2020). Arctic sea ice cover data from spaceborne SAR by deep learning. *Earth System Science Data Discussions*, (December), 1–30.

Xing, Y., Zhong, L., & Zhong, X. (2020). An Encoder-Decoder Network Based FCN Architecture for Semantic Segmentation. *Wireless Communications and Mobile Computing*, *2020*. https://doi.org/10.1155/2020/8861886

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Appendix A

Table 4. Statistics on additional images dataset model predictions.

Alaska	All	Ice	Water
count	54	28	26
mean	72.09	66.23	78.39
st. dev.	29.80	31.13	26.90
max	100.00	97.00	100.00
min	0.00	0.00	16.10

Canada	All	Ice	Water
count	70	29	41
mean	74.26	45.43	94.66
st. dev.	30.32	21.89	15.07
max	98.00	85.00	98.00
min	0.00	8.10	0.00

Finland	All	Ice	Water
count	107	71	36
mean	86.17	83.54	93.83
st. dev.	16.06	16.81	14.92
max	100.00	100.00	100.00
min	5.50	5.50	14.70

Russia	All	Ice	Water
count	60	28	32
mean	83.04	78.51	87.00
st. dev.	20.60	18.64	21.40
max	100.00	96.00	100.00
min	19.60	19.60	32.00

Appendix B

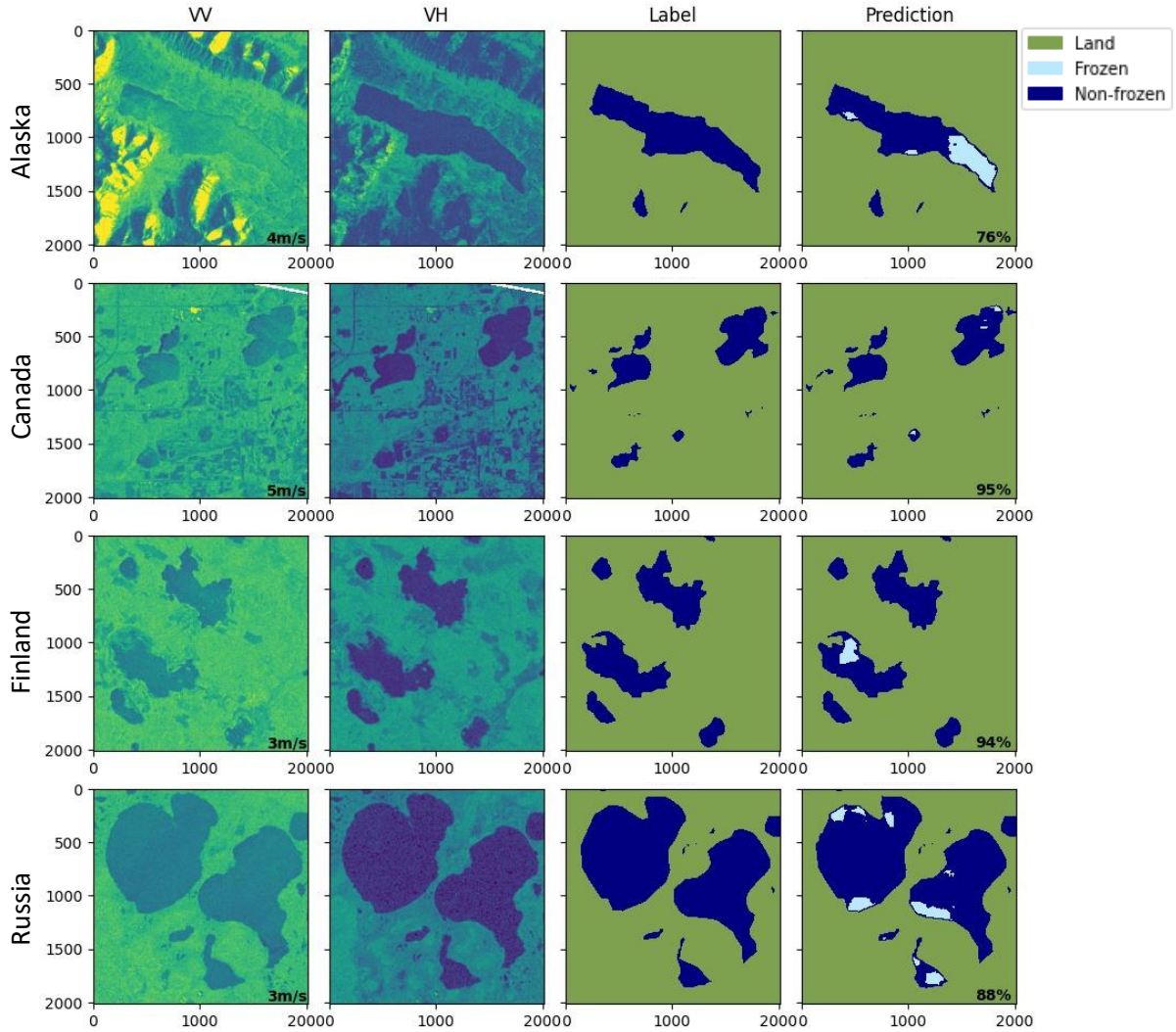

Figure 12. Qualitative results for non-frozen lake imagery for Alaska (26-10-2020), Canada (15-10-2020), Finland (04-08-2021) and Russia (14-09-2017). The first two columns show Sentinel-1 SAR VV and VH values including average windspeed, the third column the label (non-frozen) for the lake polygons and the fourth column the model prediction including accuracy (%).