DISS. ETH NO. 27829

Lake Ice Monitoring from Space and Earth with Machine Learning

A thesis submitted to attain the degree of

Doctor of Sciences of ETH Zurich

(Dr. sc. ETH Zurich)

presented by

Manu Tom

M.Sc. RWTH, RWTH Aachen University, Germany

born on 12.05.1988

citizen of India

accepted on the recommendation of

Prof. Dr. Konrad Schindler, ETH Zurich, Switzerland
Dr. Emmanuel Baltsavias, ETH Zurich, Switzerland
Prof. Dr. Andreas Kääb, University of Oslo, Norway
Prof. Dr. Devis Tuia, EPFL, Switzerland

2021

Abstract

Lake ice monitoring is a topical problem in remote sensing of the environment. The continuous observation of lake ice phenology, including important events such as *ice-on* and *ice-off* dates, along with the estimation of the corresponding long-term trends, provides useful cues on local and global climate, and hence, is beneficial for the research of climate change and global warming. Consequently, the Global Climate Observing System has identified lake ice (as part of *lakes*) as an essential climate variable. Moreover, monitoring lake ice is important for developing winter tourism, planning freshwater transportation and hydro-power generation, and understanding the various physical, chemical, and biological aspects of freshwater systems.

Existing methodologies for lake ice monitoring largely rely on a top-down knowledge-driven design. This thesis approaches the problem from a different perspective. The objective is to monitor lake ice in a data-driven manner, without the need for physics-inspired models, by employing supervised, statistical machine learning. The spatio-temporal extent of ice cover in lakes is estimated from satellite imagery and terrestrial webcams. Lake ice detection is formulated as a pixel-wise semantic segmentation problem. The models are trained using ground truth generated by visual interpretation of images from public webcams and optical satellites that monitor the target lakes.

This thesis investigates the potential of various single- and multi-sensor methodologies and presents a satellite sensor-fusion approach. First, classical machine learning methodologies such as Support Vector Machines are used to survey ice in small-to-medium-sized mountain lakes from low-spatial resolution (250m–1000m) optical satellite imagery such as MODIS and VIIRS. For higher-resolution SAR imagery from Sentinel-1, we turn to deep, convolutional neural networks. Unlike optical satellite imagery, SAR has the advantage that no data loss occurs due to clouds. In addition, the potential of close-range webcam images (again in combination with convolutional networks) is studied as a possible alternative to satellite imagery, particularly as a stand-alone data source to monitor small lakes that cannot be easily observed using MODIS and VIIRS. Furthermore, a new dataset of webcam images with lake ice annotations is presented. Finally, a deep data fusion methodology is proposed to fuse the information from both optical and radar satellite sensors to improve the temporal resolution for lake ice monitoring.

Results are reported for four selected lakes located in the Swiss Alps: Sihl, Sils, Silvaplana and St. Moritz. The primary focus lies on the analysis of two winters, 2016-17 and 2017-18. Moreover, a 20-winter time series using optical satellite imagery (MODIS) is established to derive long-term lake ice trends. Detailed experiments are performed to assess model performance, including spatial as well as temporal generalisation.

Kurzzusammenfassung

Die Beobachtung der Seevereisung ist ein aktuelles Problem der Umweltfernerkundung. Eine kontinuierliche Beobachtung der Eisphänologie, einschliesslich wichtiger Ereignisse wie der *ice-on* und *ice-off* und deren langfristiger Trends, liefert wichtige Hinweise auf das lokale und globale Klima und ist daher nützlich für die Erforschung des Klimawandels. Folglich hat das *Global Climate Observing System* Seeeis (als Unterkategorie von *Lakes*) als eine *Essential Climate Variable* definiert. Darüber hinaus ist die Überwachung von Seeeis auch wichtig für den Wintertourismus, für die Planung von Wassertransport und Wasserkraft, sowie für das Verständnis verschiedener physikalischer, chemischer und biologischer Aspekte von Süsswassersystemen.

Bestehende Methoden für die Überwachung von Seeeis beruhen meist auf einem wissensbasierten *top-down*-Ansatz. Die vorliegende Arbeit behandelt die Aufgabe aus einer anderen Perspektive. Ziel ist es, Seeeis mit einem datengetriebenen Ansatz zu überwachen, d.h., statistisches maschinelles Lernen einzusetzen, ohne auf physikalisch inspirierte Modelle zurückzugreifen. Der räumliche und zeitliche Umfang der Eisdecke von Seen wird aus Satellitenbildern und terrestrischen Webcams geschätzt. Die Erkennung von Seeeis wird dabei als pixelweise semantische Segmentierung formuliert. Die entsprechenden Modelle werden mit Referenzdaten trainiert, die durch visuelle Interpretation von Bildern öffentlicher Webcams und von optischen Satellitenbildern gewonnen wurden.

Die Arbeit untersucht das Potenzial verschiedener Einzel- und Multisensor-Methoden und stellt einen Ansatz zur Datenfusion von Satellitensensoren vor. Zunächst werden klassische maschinelle Lernmethoden wie die Support Vector Machine verwendet, um in optischen Satellitenbildern mit geringer räumlicher Auflösung (MODIS und VIIRS, 250m–1000m GSD) das Eis auf kleinen bis mittelgrossen Bergseen zu detektieren. Als alternative zu optischen Satellitenbildern werden für die Analyse von SAR-Bildern (Sentinel-1) tiefe *convolutional neural networks* eingesetzt, um Datenverlust durch Wolken zu verhindern. Des Weiteren wird das Potential von Nahbereichsbildern öffentlicher Webcams untersucht, insbesondere als Datenquelle zur Überwachung kleiner Seen, die mit MODIS und VIIRS kaum beobachtbar sind (ebenfalls unter Einsatz von *convolutional networks*). Dazu wird auch ein neuer Datensatz von Webcam-Bildern mit annotierten Referenzdaten für die Seevereisung zur Verfügung gestellt. Schliesslich wird eine auf *deep learning* basierende Methode zur Datenfusion vorgeschlagen, um die Informationen aus optischen und Radarsensoren zu kombinieren, mit dem Ziel, die zeitliche Auflösung bei der Beobachtung von Seeeis zu verbessern.

Es werden Ergebnisse für vier ausgewählte Seen in den Schweizer Alpen präsentiert: *Sihlsee, Silsersee, Silvaplanersee* und *St. Moritzersee*. Das Hauptaugenmerk liegt auf der Analyse von zwei Wintern, 2016/17 und 2017/18. Darüber hinaus wurde mit Hilfe von

MODIS eine Zeitreihe über 20 Winter erstellt, um langfristige Vereisungstrends abzuleiten. Die Qualität der entwickelten Modelle wird in detaillierten Experimenten analysiert, inklusive ihrer Fähigkeit zur räumlichen und zeitlichen Generalisierung.

To my parents..
Special thanks to Trisha and Noonu.
To Konrad and Manos for their support and trust in me.
Thanks to all my colleagues and students at ETH and my close friends.

List of acronyms

AI	Artificial Intelligence
ASPP	Atrous Spatial Pyramid Pooling
BUE	Break-Up End
BUS	Break-Up Start
CCI	Climate Change Initiative
CFD	Complete Freeze Duration
CLASS	Comprehensive Large Array-data Stewardship System
CNN	Convolutional Neural Network
DAAC	Distributed Active Archive Center
D/A	Days per Annum
DL	Deep Learning
ECV	Essential Climate Variable
EO	Earth Observation
EOS	Earth Observation System
ESA	European Space Agency
FUE	Freeze-Up End
FUS	Freeze-Up Start
GCOS	Global Climate Observing System
GPU	Graphics Processing Unit
GSD	Ground Sampling Distance
IoU	Intersection-over-Union
LAADS	Level-1 and Atmosphere Archive & Distribution System
LIC	Lake Ice Cover
LIE	Lake Ice Extent
LIP	Lake Ice Phenology
mIoU	Mean Intersection-over-Union

ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
MSP	MODIS Snow Product
MTA	Multi-Temporal Analysis
NASA	National Aeronautics and Space Administration
NDSI	Normalised Difference Snow Index
NDWI	Normalised Difference Water Index
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Centre
OSM	Open Street Map
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
S1-SAR	Sentinel-1 SAR
UAV	Unmanned Aerial Vehicle
VIIRS	Visible Infrared Imaging Radiometer Suite
VSP	VIIRS Snow Product
XGB	XGBoost

Contents

1 Introduction

1.1. Motivation

Climate change is one of the fundamental challenges that humanity encounters today. Understanding the extent of environmental damage occurring due to climate change and global warming is essential to devise efficient mitigation plans to protect the environment for future generations. Lake ice is an important climate change indicator („sentinel"). Investigating the underlying patterns in lake freezing and thawing processes provides valuable insights into the regional and global climate system. Hence, as part of *Lakes*, the Global Climate Observing System (GCOS) recognised *lake ice cover* as an Essential Climate Variable (ECV, `https://gcos.wmo.int/en/essential-climate-variables/lakes`).

Past study (Hendricks Franssen and Scherrer, 2008) on eleven lakes (Biel, Murten, Sempach, Hallwil, Baldegg, Sarnen, Aegeri, upper lake Zurich, Greifen, Pfaeffikon, Untersee) in Switzerland (with an area of 3.2 - 39.8 km^2 and altitude of 397 - 724 m) has reported freezing in fewer lakes and for shorter periods. Earlier attempts to record and document lake ice were largely independent and not necessarily coordinated and primarily came from sporadic sources such as onshore observers, regional authorities, hobby enthusiasts etc. Additionally, the lack of large-scale automated monitoring capabilities limited these observations to regional levels. Furthermore, such ground-based observations reduced significantly in the past few decades, which effectively increased the relevance of satellite-based remote sensing of lake ice (Duguay et al., 2015). In this context, taking advantage of the available data sources to develop an automated visual monitoring system can be helpful, especially to extend the lake ice documentation to country- or world-level. This thesis proposes to use and integrate various sensors and methodologies to move closer to this long-term goal.

Regarding the monitoring of lake ice, several initiatives are underway. An example is the National Snow and Ice Data Centre (NSIDC) database in Colorado (`https://nsidc.org`) that consolidates data on various Lake Ice Phenological (LIP) events for 865 lakes from across the globe. Several operational lake ice products already exist, such as the Climate Change Initiative (CCI) Lake Ice Cover (LIC) product (Crétaux et al., 2020), and Copernicus Lake Ice Extent (LIE) product (`https://land.copernicus.eu/global/products/lie`). However, these dedicated lake ice products do not cover the target lakes. The MODIS snow product (Hall and Riggs, 2016) and VIIRS snow product (`https://nsidc.org/sites/nsidc.org/files/technical-references/VIIRS-snow-products-user-guide-final.pdf`) are relevant too. All these products differ from this work in that they primarily rely on various indices (NDSI, NDWI etc.) and threshold-based algorithms. On the other hand, this thesis attempts to learn lake

ice monitoring from data by utilising contemporary developments in the field of artificial intelligence. Though many studies estimated the spatio-temporal extent of ice in big lakes across the globe, smaller mid-latitude mountain lakes that freeze have not been sufficiently investigated. This thesis aims to fill that research void.

In the last few decades, satellite data analysis has made several breakthroughs for environment monitoring applications. Machine Learning (ML) has also progressed in recent decades, especially the advancements in Deep Learning (DL) in the past decade. This thesis tries to leverage the best of these two worlds for the specific application of lake ice monitoring. Several satellite data sources are explored, both optical and radar. Additionally, this dissertation delves into the possibilities of non-satellite imagery such as publicly available webcams for the geospatial data analysis task at hand.

The algorithms presented in this thesis have been developed as part of two projects (2016-2020) funded by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss in the framework of GCOS Switzerland. The first project was a feasibility study that analysed and compared the effectiveness of optical satellite imagery, webcam data and in-situ temperature measurements for monitoring lake ice (Tom et al., 2019). Targeting systematic and steady observation, the second project integrated data from several sensors and methodologies for lake ice observation (Tom et al., 2020b). Inspired by these projects, this thesis focuses on lake ice estimation over longer time periods in the past and methods that will enable reliable and long time series monitoring in the future.

1.2. Research objectives and challenges

Though there have been some key research developments in the area of lake ice monitoring in the past few decades in the form of various index- and threshold-based approaches, the potential of artificial intelligence has not been explored much. The main goal of this thesis is to fill this research gap by approaching the problem from a different perspective by unfolding the possibilities of ML, especially DL and Convolutional Neural Networks (CNNs). State-of-the-art approaches in computer vision and ML are applied to address the remote sensing challenges of monitoring ice in selected lakes located in the Swiss Alps. Three lakes (Sils, Silvaplana, St.Moritz) which freeze consistently every year are chosen in addition to lake Sihl which freezes during most of the winters. The target is to detect the spatio-temporal extent of ice cover and the crucial LIP events such as Freeze-Up Start (FUS), Freeze-Up End (FUE), Break-Up Start (BUS), and Break-Up End (BUE), with a focus on the integration of various input data and processing methods. The long-term vision is to put forward an automated visual lake ice monitoring system that meets the GCOS requirement: daily observation with an accuracy of +/-2 days in estimating the critical LIP events. Regarding sensors, in addition to close-range webcams, the data from optical and radar satellites are used. Besides developing various single-sensor approaches, multi-modal satellite data fusion for robust lake ice observation is also proposed.

In essence, this thesis explores answers to the following overarching research questions:

1. What role can ML, including the recent advances in DL, play for single-and multi-

sensor lake ice observation?

2. What are the trade-offs between spatial and temporal resolution of satellite imagery when monitoring ice in small- and medium-sized mountain lakes?

3. Can terrestrial webcams complement or even replace satellite observations, especially to survey ice in small lakes?

4. Can lake ice detection in SAR imagery be learnt from data using DL?

5. Is it viable to construct a unified, joint representation for multiple (optical and SAR) sensors to support lake ice monitoring with higher temporal frequency?

6. To what extent do local lake ice trends correlate with climate indicators such as temperature, precipitation, sunshine etc.?

In the quest for answers to these questions, several challenges have been identified. For the target regions, no custom datasets (and corresponding ground truth) exist that are tailored to solve the problem of interest. One of the major bottlenecks in any learning-based system is the time and effort needed to consolidate the dataset and gather reliable ground truth (usually done manually) at an acceptable granularity level (image-wise, instance-wise, pixel-wise etc. depending on the problem). This challenge is even more intense if the models are fully supervised and/or if the processing pipeline relies on data-hungry deep neural networks. The ground truth labels needed are primarily generated by interactive visual interpretation of RGB webcam imagery. This annotation process is labour intensive, and pixel-wise labelling is difficult, especially for webcams with very high spatial resolution. While delineating the ground truth using the freely available webcam images, the human operator should deal with obstacles like low image quality, compression artefacts, poor radiometric and spectral resolution. Additional bottlenecks during the annotation process are inter-class appearance similarities due to oblique viewing angles, adverse lighting conditions like fog, sun reflections, shadows from the nearby mountains and/or clouds etc. Specifically for the problem of lake ice labelling, *water vs thin ice* inter-class similarity could confuse even the experts.

The primary task of monitoring lake freezing patterns is to detect lake ice from optical (MODIS and VIIRS) and radar (Sentinel-1 SAR, S1-SAR) satellite imagery, and webcam images. One big challenge with satellite imagery analysis is the spatio-temporal-spectral resolution trade-off. Since the GCOS requirement is daily observation with an accuracy of +/-2 days for the LIP events, the optimum optical satellite choices for the problem are MODIS and VIIRS imagery with daily temporal resolution. The spectral resolution is also good for these two satellite sources. However, their spatial resolution (250-1000m) is satisfactory, making it difficult to survey the relatively small lakes. Note that, the lakes that freeze in Switzerland are generally small in area. However, this thesis investigates how these two satellite images are useful for detecting lake ice in such small lakes. When these low spatial resolution (large Ground Sampling Distance, GSD) satellite images are used to analyse the small- or medium-sized lakes, a new problem arises: the scarcity of pixels, which makes it difficult to train a robust data analysis system. Additionally, in those scenarios, absolute geolocation errors of the satellite sensors become critical, making it essential to correct them. However, even the most efficient correction methodologies

still leave behind an error of up to half a pixel post-correction, which is critical for the target lakes with a relatively small area ($0.78 - 11.3\ km^2$). Other freely available optical satellite imagery options such as Sentinel-2 and Landsat-8 have a better spatial resolution but worse temporal resolution, which is insufficient to meet the GCOS criteria. Hence, such options are not considered.

An inherent vulnerability of optical satellite sensors is the inability to penetrate clouds, resulting in data loss. More critical is when the clouds occur during/near the crucial freeze-up and/or break-up periods. Compared to the optical counterparts, S1-SAR imagery offers a better spatio-temporal resolution trade-off for the target application, thanks to the microwave sensor's ability to see through clouds and better spatial resolution. However, it is not easy to manually interpret SAR imagery, unlike optical satellite images. Furthermore, the temporal resolution is not enough for the regions of interest (and in general for regions outside the Arctic) to meet the GCOS criteria. Hence, this thesis additionally explores the possibilities of multi-sensor data fusion. However, the huge spatial resolution difference between MODIS/VIIRS and S1-SAR, and the large gap between the optical and radar domains pose additional challenges while performing data fusion.

Two main challenges are expected in webcam-based lake ice monitoring. Firstly, achieving full lake coverage using a single webcam might be nearly impossible, especially for medium-sized (and larger) lakes. In practice, matching the images and outputs from multiple webcams might be needed to circumvent this problem. Secondly, the placement of freely available webcams are typically driven by other considerations and not optimised for ice monitoring and hence does not always deliver the images with settings optimal for lake ice detection. Considering all these challenges, this thesis investigates in detail the possibility of using webcams as an alternative for satellite data.

Semantic segmentation is one of the core research areas in the fields of computer vision and pattern recognition. The underlying task is pixel-wise classification, i.e., to generate an output map by assigning a label to each pixel in an input image from an already known set of classes. Some surveys on semantic segmentation already exist in the literature (Thoma, 2016, Zhu et al., 2016, Garcia-Garcia et al., 2017). Semantic segmentation of remotely sensed images has also been of interest in Earth observation research, see Section 2.3. This thesis applies the same for lake ice monitoring.

1.3. Outline and contributions

This dissertation is presented as a "cumulative doctoral thesis", complying with the Doctoral Ordinance (ODS Art. 26 to 29 and RIPO 10.b) by the Department of Civil, Environmental, and Geomatics Engineering at the ETH Zurich. The key scientific contributions of this thesis are described in Chapters 3, 4, 5 and 6 which each constitute a separate, unaltered scientific paper. The individual related works discussed in these chapters cover in detail all the literature (including the latest publications) relevant to lake ice monitoring. A synopsis of proposed methodologies and contributions are presented in sub-sections 1.3.1 - 1.3.4. An introduction (Chapter 1), some theoretical background

(Chapter 2) and, conclusions and outlook (Chapter 7) are also incorporated.

1.3.1. Optical satellites and webcams for lake ice monitoring (Tom et al., 2020c)

Chapter 3 presents two independent approaches for lake ice monitoring in selected Swiss lakes using multi-temporal optical satellite images and terrestrial webcams. ML-based image analysis is used as a tool to determine the spatio-temporal extent of ice, including the ice-on and ice-off dates, from both MODIS and VIIRS imagery, and RGB webcam data. Lake ice monitoring is formulated as a per-pixel semantic segmentation problem. Empirical evidence demonstrates that the proposed satellite- and webcam-based methods produce consistently good results when tested on the data from multiple winters and lakes, including the notable spatio-temporal generalisation performance.

Contributions

1. This chapter investigates the potential of ML-based image analysis in combination with various image sensors to retrieve lake ice. So far, such an approach has rarely been explored, especially for the many small mid-latitude lakes on Earth (particularly in mountainous regions). However, it can be a valuable source of information that is largely independent of in-situ observations and models of the freezing/thawing process.
2. An easy-to-use, Support Vector Machine (SVM)-based approach is put forward to detect lake ice in low spatial resolution (250 - 1000 m) but high temporal resolution (1 day) satellite imagery from the optical sensors MODIS and VIIRS.
3. A new state-of-the-art is set for webcam-based lake ice monitoring, using the *Deep-U-Lab* network. Additionally, in that context, the detection of lake outlines is automated as a further step towards operational monitoring with webcams.
4. A new benchmark webcam dataset (*Photi-LakeIce*) is introduced and made available, which includes data from two winters (2016–17, 2017–18) and three cameras that monitor lakes St. Moritz and Sihl, along with pixel-accurate annotations.
5. The DL model pre-trained on the *Photi-LakeIce* dataset is made public to the research community that can be fine-tuned on new webcams with different conditions and for other environmental monitoring applications.

1.3.2. Recent lake ice trends: 20-year analysis of MODIS imagery (Tom et al., 2021b)

In Chapter 4, the critical LIP events (freeze-up, break-up and freeze duration) are observed across two decades in selected Swiss lakes from optical satellite images. The long MODIS time series is analysed and cross-checked with VIIRS data when available. Long-term LIP trends are derived by estimating spatially resolved maps of lake ice with supervised ML. The results are validated against the operational MODIS and VIIRS

snow products and the ground truth derived by visual interpretation of webcam images (whenever available). The estimated freezing and thawing trends are correlated with climate data measured at the nearby meteorological stations.

Contributions

1. A 20 winter (2000-2019) long time series is established for monitoring ice in selected lakes in south-eastern Switzerland, primarily based on MODIS imagery.

2. The limits of MODIS data are pushed for the analysis of small-to-medium-sized high-alpine lakes. This work shows that it is possible to derive meaningful correlations between the 20-winter LIP trends and climate indicators (temperature, sunshine and precipitation), even for such lakes.

3. The dedicated ML scheme maps lake ice more accurately than the classical index- and threshold-based approaches.

4. As expected, the results point towards later freeze-up, earlier break-up and reduced temporal freeze extent. A change in Complete Freeze Duration (CFD) of -0.76 and -0.89 days per annum (d/a) for lakes Sils and Silvaplana, respectively, are found.

1.3.3. Lake ice monitoring from Sentinel-1 SAR data (Tom et al., 2020a)

Chapter 5 proposes to use a deep neural network for lake ice detection using S1-SAR data. Frequent cloud cover was the main bottleneck in the past studies that performed lake ice monitoring from optical satellite data. The presented algorithm overcomes this hurdle, leveraging microwave sensor's potential to penetrate clouds and monitor the lakes irrespective of the weather and illumination conditions. The semantic segmentation problem is solved with a state-of-the-art CNN architecture: Deeplab v3+ (Chen et al., 2018b). The results are reported on two winters (2016-17 and 2017-18) and three Swiss lakes.

Contributions

1. The problem of lake ice detection from S1-SAR data is addressed as an alternative to optical satellite imagery to bypass the problem of missing data due to clouds.

2. At a technical level, the chapter shows that a DL model pre-trained on an optical RGB dataset can nevertheless be reused successfully as initialisation for fine-tuning the network parameters on radar data.

3. The model pre-trained on the proposed SAR dataset is made public to the community and can be fine-tuned for specific target regions and other geoscience applications.

1.3.4. Learning a joint embedding for multiple satellite sensors (Tom et al., 2021a)

The fusion of space-borne SAR and optical data is of special interest for geospatial data analysis and processing. Chapter 6 proposes to use deep neural networks for combining the imagery from heterogeneous satellite sensors such as S1-SAR, MODIS and VIIRS. A deep feature space representation or 'satellite embedding' is learnt to perform the information fusion. Additionally, a case study of lake ice monitoring is performed to confirm the efficacy of the proposed deep satellite data fusion methodology for an application task.

Contributions

1. A DL framework is proposed that learns a joint embedding to fuse MODIS, VIIRS, and S1-SAR satellite data.
2. The proposed methodology can be used to perform both optical-optical and optical-SAR fusion.
3. The presented approach achieves good cross-winter and cross-lake generalisation performance.
4. A new state-of-the-art is set in estimating the ice-on/off dates for many of the target lakes, closely meeting the GCOS requirement.

1.4. Technical relation of the chapters

Optical satellite imagery analysis is performed in Chapters 3, 4 and 6. While Chapter 3 describes the preliminary lake ice detection results using MODIS and VIIRS data from two back to back winters (2016–17, 2017–18), Chapter 4 presents the detailed results of MODIS data processing from all available winters to date. In Chapter 6, in addition to optical imagery, radar satellite data is used for lake ice monitoring. Additionally, Chapter 5 discusses radar satellite data analysis. However, webcam image analysis is only discussed in Chapter 3.

While Chapters 4 and 5 investigate the feasibility of single-sensor approaches, a multi-sensor fusion methodology is presented in Chapter 6. Though it is not the primary focus of Chapter 3, the fusion of MODIS and VIIRS data (at a decision level) is also discussed. While Chapters 3 and 4 describe the use of more traditional ML approaches, Chapters 3, 5 and 6 employ DL.

1.5. Relevance to science and society

This dissertation is the outcome of four years of cooperation with MeteoSwiss, the Federal government agency tasked with implementing and maintaining the GCOS objectives

for Switzerland. The goals of this thesis also align with the following interest of MeteoSwiss: to enhance and strengthen the Swiss climate observing system by promoting the development and integration of existing and emerging observation methods.

Analysing the decreasing LIP trends across several winters and correlating them with the climate patterns is significant for climate change quantification and global warming research. The findings of this thesis in Chapter 4 show that, in the Swiss lakes Sils and Silvaplana, a significant loss of ice happened in the past two decades. As the global temperature is on the rise, these lakes will probably lose further ice in the future years.

Estimating the changing duration of ice cover helps to understand the various physical, chemical, and biological processes of freshwater systems. It has many economic implications as well, such as for hydroelectric power generation, efficient freshwater transportation etc. Furthermore, efficient surveillance of lake ice is useful for the development of winter tourism. E.g. there exists the Snow Polo World Cup (since 1985, worth many hundreds of millions of dollars) and White Turf International Horse Race (since 1907) organised every winter in frozen lake St. Moritz in Switzerland. For such events, automated monitoring of lake ice is useful for better long-term planning. Furthermore, monitoring the duration of lake ice cover is significant for recreational activities such as snowmobiling, ice cricket (St. Moritz), ice skating (e.g., in lake Louise in Canada or lake Baikal in Russia) and ice-fishing (e.g., lakes Sils and Oeschinen in Switzerland). Such activities are of large economic importance to the regional communities, bringing in supplies and/or significant tourism during the winter season.

The methodologies proposed in this thesis can potentially, with little effort, be adapted to other Earth observation tasks. Some options are mapping and monitoring sea ice, river ice, polar ice sheets, snow cover, glacial lakes, glaciers, and other applications that observe the Earth's cryosphere. The webcam images made available as part of the Photi-LakeIce dataset (Chapter 3) can be directly used for computer vision applications such as outdoor scene understanding (especially under bad weather scenarios), semantic segmentation, fog modelling etc., and other environmental monitoring applications such as lake water level prediction. Further applications foreseen for webcam-based analysis are water quality prediction, estimation of lake surface water temperature (using thermal and RGB cameras), monitoring the phenology of blue-green ecosystems (e.g., the EAWAG-WSL blue-green biodiversity research initiative, `https://www.eawag.ch/en/department/surf/projects/phenology-of-blue-green-ecosystems/`), methane hole detection etc. Though Chapter 6 proposes to learn a satellite embedding for lake ice monitoring, this DL-based approach is generic. It can be easily adapted to other Earth science applications that demand multi-sensor satellite data fusion, especially the fusion of optical and SAR data.

Open-source. Relevant source code and the Photi-LakeIce webcam dataset, implemented and developed in the course of this thesis, are made publicly available under an open-source license at `https://github.com/czarmanu/lake-ice-ml` to facilitate the easy reproducibility of this research. The backbone chapters (3, 4, 5 and 6) in this dissertation stem from two publications and two submissions in open access journals/conferences which the research community can access free of cost.

2 Background

2.1. Machine and deep learning

The foundations of ML, a sub-field of AI, fundamentally stem from statistical analysis. This section introduces the underlying concepts and principles of ML methodologies applied for lake ice monitoring and is hence relevant for this thesis, particularly the DL approach using CNNs.

Learning can be designed in supervised, unsupervised or even semi-supervised formats. The former can perform reliable classification as well as regression by training a model from annotated data. On the other hand, in unsupervised learning, data is clustered using a distance measure or similarity score to learn the underlying distribution from an unlabelled dataset. Partly-annotated data is only needed in semi-supervised learning which takes a middle ground between the supervised and unsupervised approaches.

Since supervised learning methodologies have proved to be highly effective for remote sensing image analysis, they are adopted in this thesis (Chapters 3-6) too. However, learning from a limited amount of labelled training data is a perennial problem in Earth observation research because reference data tends to be scarce. Semi-supervised learning could potentially stand out with the possibilities of big data. However, it never did quite fulfil its promise even for generic ML problems, and geospatial data analysis is no exception. Yet another possibility is unsupervised pre-training, which for images still did not bring a breakthrough, but in language models it did, for example, BERT (Devlin et al., 2019), GPT3 (Brown et al., 2020) etc. However, it might at some point in the future also establish itself for images.

In this big data era, massive volumes of data already exist in both offline and online platforms in various formats from different sources and this volume is rapidly increasing each day. This is also true for satellite data, for example, the huge online archives of both optical and radar imagery from various sensors. Hence, training even the data-hungry DL models for Earth observation applications is not really an issue in terms of data volume, these days. However, the bottleneck is that such data are generally unlabelled and are hence not directly usable. Annotation of data at a per-pixel granularity level to train parameter-heavy deep networks is a tedious process that often involves a time-consuming manual visual interpretation process.

2.1.1. Classical learning approaches

Feature engineering-based classical learning methodologies were heavily used for image understanding tasks before DL techniques took over. Extracting the features must be

done with expert knowledge for the former approaches while the latter learns the feature extractor. In some applications the raw data or simple combinations of them may already be good features, refer to Chapters 3 and 4. The classical approaches are still enough to solve many problems in geospatial data analysis either when a huge amount of reliable ground truth data is unavailable to train deep networks or when a sophisticated neural network is not necessary especially in cases when contextual information cannot be learnt. Additionally, the classical methodologies are easily trainable since they need relatively less amount of training data. This sub-section presents some details of the concepts behind three such techniques that were successfully employed for lake ice monitoring.

In both classification and regression, false assumptions during the learning process could induce a bias error while sensitivity to variations in the training set can lead to variance error. High bias could result in underfitting while high variance causes the model to overfit. To train an optimum model which fits well on the training data as well as generalises well on unseen test data, a right trade-off between bias and variance need to be attained.

Support vector machines (Cortes and Vapnik, 1995). SVM is a discriminative ML approach that learns a hyperplane that best separates the data points in a feature space into different classes by the concept of maximising a *margin* (distance from the learnt hyperplane to the closest feature points). Such feature points (support vectors) that lie on/inside the margin are relevant in learning the hyperplane using an optimisation strategy that maximises the margin, formulated as a convex optimisation problem and hence it is possible to find the global optima.

If the input data points are not linearly separable in the original feature space, mapping to a higher dimensional space by kernel trick could make the data linearly separable in the new space. The kernel trick can be applied whenever dot products exist and the kernels used can be non-linear such as Radial Basis Function (RBF), polynomial etc. Choosing the right kernel is however very important and also can be dataset dependent.

SVMs are also applicable for other complex data types such as sequences, graphs etc. However, appropriate kernels need to be tailored to such data. Note that, SVMs can become very slow in case of an increased number of support vectors and/or computationally expensive kernels.

Random forest (Breiman, 2001, RF) is an ensemble learning approach that relies on bagging (bootstrap aggregation) of decision trees constructed from the data.

In a decision tree, each node denotes an attribute and each branch of the tree defines a possible value of the attribute. A tree can be either binary or multi-valued depending on the number of splits possible at each node. Though simple decision trees often produce interpretable outcomes, they generally result in noisy and weak classifiers that do not generalise well. Very shallow trees have high bias and low variance and hence tend to underfit while deep trees have low bias and high variance and overfit on the training data.

In RF, multiple decision trees are produced from a different bootstrapped subset

(sampling with replacement) of the training dataset. In addition, randomness is infused into the trees in many ways which make each tree weaker (depriving it of some data, split variables etc.). However, the ensemble gets stronger in total through randomness. Only a subset of the input features is randomly selected during each split, while creating the decision trees, to minimise the correlation between the trees in the forest. The predictions of the individual trees are averaged to avoid overfitting.

Trees in RF typically have low bias and high variance. Each tree in the forest should have a low bias since the RF classifier effectively has the same bias as that of any individual tree (since the trees are identically distributed). High variance is however acceptable since the averaging step reduces the classifier variance.

RF classifier can be trained relatively fast and also typically generalises very well on unseen data. On the contrary, it has a high memory footprint and is ideal when the training dataset size is not that large.

Extreme gradient boosting (Chen and Guestrin, 2016) or XGBoost (XGB) is an ML methodology fundamentally based on gradient boosted decision trees. Gradient boosting is also an ensemble-based technique that sequentially learns/adds trees to the ensemble which predicts and corrects the errors of previous trees. Many such trees are iteratively combined to create the final predictor. When new trees are added, the loss is minimised using the gradient descent strategy. XGB is highly scalable and is sparsity aware.

Though both XGB and RF are based on decision trees, the depth of the trees involved in each case is very different. Typically, the trees in RF are deeper. On the contrary, XGB trees are shallow with high bias and low variance. This is because, whenever a new tree is added in XGB, though the model complexity rises, the overall bias reduces, irrespective of the tree size. However, the variance remains the same.

More details about SVM, RF and XGB hyperparameters are discussed in Chapter 4.

2.1.2. Deep neural networks

Though DL (LeCun et al., 2015, Goodfellow et al., 2016) research has significantly progressed in the past decade, it is constantly undergoing development and has still not saturated. On the one hand, deep, multi-layer neural networks have redefined the state-of-the-art in various sub-fields of AI research, including but not limited to Earth observation. On the other hand, scientists are still trying to understand how these approaches work so well. There are a set of problems in research and engineering for which such techniques offer excellent solutions. However, deploying DL tools blindly on all sorts of research problems is not a recommendable practice.

One caveat is that training the deep networks is cumbersome especially given the fact that they are extremely data-greedy and consume a huge amount of computational resources and time. However, with the advent of big data and powerful Graphics Processing Units (GPUs), such networks have become unparalleled image recognition tools. Nevertheless, it seems that hyperparameter (optimiser, learning rate, choice of non-linearity,

number of hidden layers etc.) selection matter a lot while applying such networks for real-world problems. In the next paragraphs, this thesis delves deep into the underlying concepts of CNNs.

Convolutional neural networks CNNs (or ConvNets) are specially designed feed-forward neural networks particularly useful to analyse data having a grid-like topology, such as time series (1D), images (2D), videos (3D) etc. Such networks are capable to perform end-to-end learning, for eg., mapping from a raw input image to an outcome which could be a predicted point estimate, semantic segmentation map, learnt feature vector etc., depending on how the problem has been formulated.

CNNs typically consist of multiple convolutional and pooling layers, and non-linear activations. Depending on the problem being solved, there can be fully connected layer(s) too. The first layer is always convolutional. The sequence of convolution, non-linear activation and pooling operations is equivalent to a feature extractor that can "learn" optimal features for the problem and dataset at hand. Deep CNNs stack multiple such feature extractors in such a way that the higher layers learn more global, invariant features such as shapes, object parts, and sometimes even full objects. On the contrary, the lower layers learn simple patterns like edges, blobs, colours etc.

There exist many popular off-the-shelf CNN architectures such as LeNet (Lecun et al., 1998), AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan and Zisserman, 2015), GoogLeNet (Szegedy et al., 2015), ResNet (He et al., 2016), Xception (Chollet, 2017), MobileNetV2 (Sandler et al., 2018) etc. An example LeNet-type CNN architecture with two convolutional and pooling layers each, and three fully connected (dense) layers, is shown in Fig. 2.1. The following paragraphs will discuss some details about different type of layers and activations generally used in CNNs.

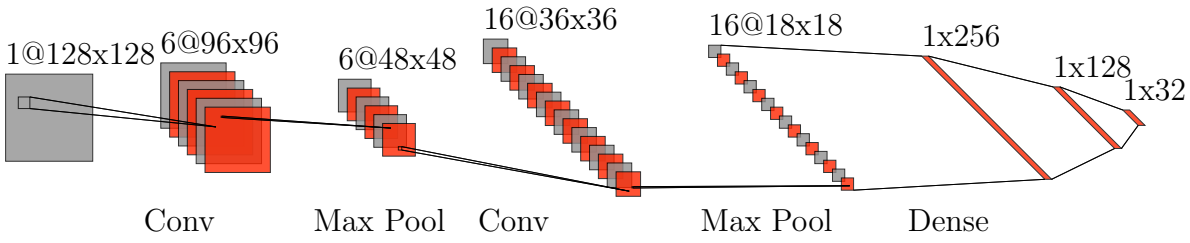


Figure 2.1.: Example LeNet-style CNN architecture. *Conv, Max Pool* and *Dense* represent convolution, max pooling and fully connected layers respectively.

In a convolutional layer, an input image is convolved (discrete) with the filter kernels that are learnt and feature maps (one channel per filter kernel) are generated. For eg., in Fig. 2.1, six and sixteen feature maps are produced in the first and second convolution layers respectively. The convolution weight kernel is a linear filter and the concept of weight sharing applies, i.e., the filter parameters are shared between different spatial locations on the input image, thereby substantially curtailing the number of network parameters.

A pooling layer operates on each feature map separately and reduces the spatial dimensions, targeting to lower the number of parameters. Furthermore, the pooling

layers impart small translation invariance and can additionally model context relations by effectively increasing the receptive field. It is to be mentioned in this context that strided convolutions with stride > 1 (equivalent to downsampling after convolution with unit stride) also increases the receptive field. Different types of standard pooling operations are: *max* and *mean*. A study exists in the literature which deals with the theory of selection of the best pooling strategy for vision problems (Boureau et al., 2010). The most widely used technique is max pooling where the maximum pixel value in a rectangular window is chosen. An example max pooling operation with 2×2 filter kernel and stride 2 between the pools, when applied on a 4×4 image is demonstrated in Fig. 2.2, where the input size is reduced by a factor of two in both x- and y-directions. It can be seen in Fig. 2.1 that two max pooling layers (2×2 kernel with stride 2) exist in the delineated CNN architecture and the spatial dimensions are halved after the input is passed through these layers: 96×96 to 48×48, and 36×36 to 18×18, after the first and second pooling layers respectively.

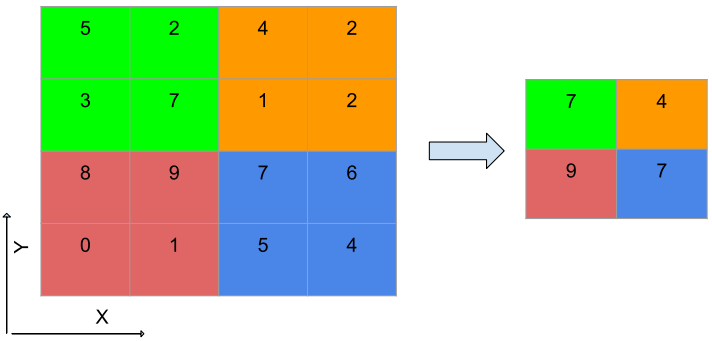


Figure 2.2.: Example max pooling operation on a 2×2 image with 2×2 filter and stride 2 generating a 2×2 image as output.

Within each node in any layer in a neural network, a transfer function (a.k.a. activation function) decides how the output is generated from the weighted sum of the input. Non-linearity is a necessary property for activation functions so that the model could learn complex input-output relationships. Some commonly used non-linear activation functions are: *sigmoid, tanh, Rectified Linear Unit (ReLU)* and *softmax*, see also Fig. 2.3 and equations 2.1-2.4. For an input x, the *sigmoid* transfer function $g(x)$ is defined as:

$$g\left(x\right) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

Since the sigmoid output lies in the range (0,1), it is particularly useful for models that predict probability. Additionally, it is differentiable. On the other hand, the hyperbolic tangent (*tanh*) function produces values between -1 and 1. Furthermore, unlike sigmoid, the tanh output is zero-centred, which makes learning relatively easier. tanh is also differentiable and is defined as follows:

$$g\left(x\right) = tanh(x) = 2\sigma(2x) - 1 \tag{2.2}$$

Though ReLU (Jarrett et al., 2009, Nair and Hinton, 2010) activation is non-linear by definition, see equation 2.3, its properties are pseudo-linear in many aspects. As a re-

sult, the gradients can be propagated easily during the training which results in faster convergence. Since ReLU maps many values to zero, it typically results in sparse representations, i.e., only a subset of neurons will be active and the computations will be linear within this subset (Glorot et al., 2011). Furthermore, the computation overhead is less than sigmoid and tanh. Moreover, the networks with ReLU activation generalise very well. Hence, it became the de-facto standard for deep neural networks. *ReLU* is mathematically expressed as:

$$g(x) = max\{0, x\} \tag{2.3}$$

New advanced varieties of ReLU such as leaky ReLU (Maas et al., 2013), ELU (Clevert et al., 2016) etc., also have been proposed.

Softmax is an interesting activation function since the output lies in the range (0,1) which sum to 1, making it apt to predict probabilities. Equation of *softmax* activation is given by:

$$g(x) = \frac{e^{-x_i}}{\sum_j e^{-x_j}} \tag{2.4}$$

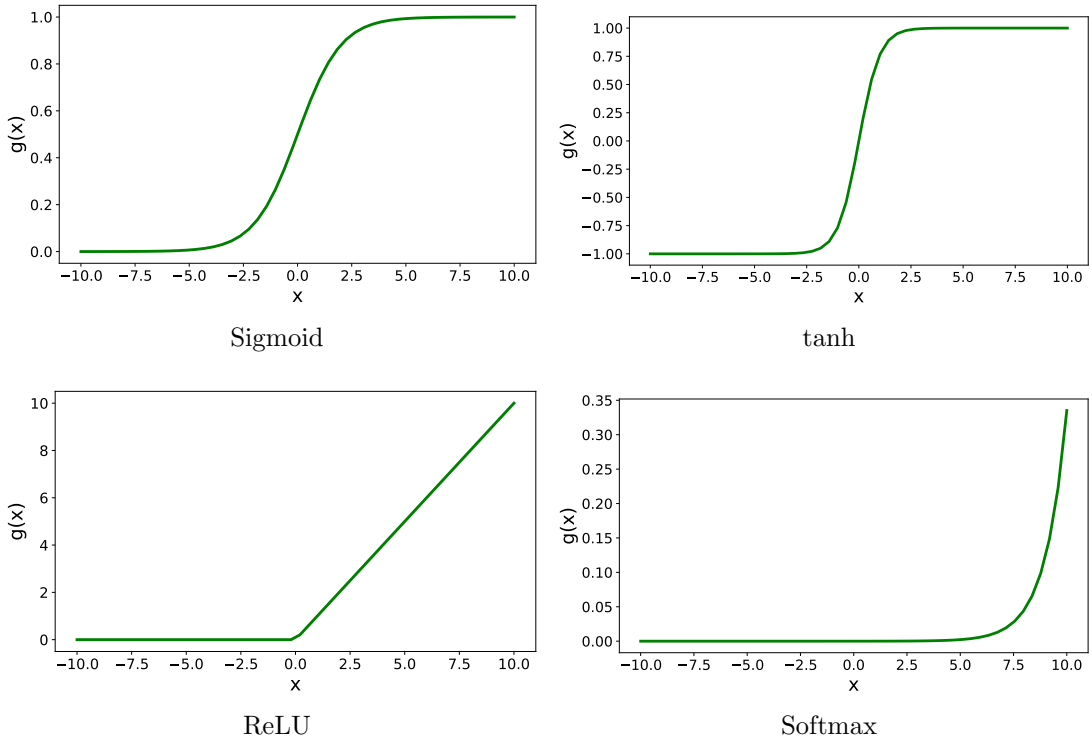


Figure 2.3.: Different non-linear activation functions commonly used in deep neural networks.

In the example CNN displayed in Fig. 2.1, three dense (fully connected) layers that learn 256-, 128- and 32-dimensional feature vectors respectively are shown. In a dense layer, each neuron obtains as input the output of every neuron in the previous layer, see Fig.2.4, making the dense layer similar to a convolution layer with a receptive field as big as the entire input layer. Consequently, the spatial information is lost.

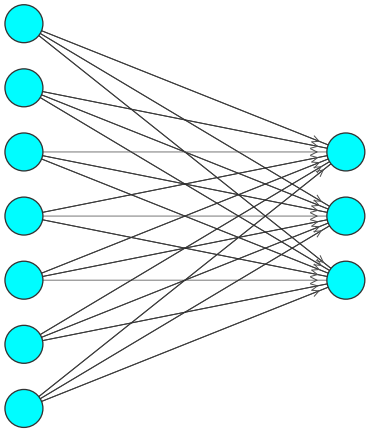


Figure 2.4.: An example fully connected layer with 7 input nodes and 3 output nodes.

Training the network. Neural networks are trained using a procedure called *backpropagation*. For each image in the training dataset, the output is determined in a forward pass which is then compared with the ground truth to estimate the error. The error derivatives (gradients) are then computed in each hidden layer and propagated backwards through the network, one layer at a time. The trainable network weights (parameters) in each layer are then modified according to the individual contribution to the total error. These steps are repeated for all the images in the training set. One complete pass through the whole training dataset is called an *epoch*. Many such epochs are usually needed to train a robust neural network model.

Loss functions. Learning the optimal set of network weights that can make good predictions is done by minimising an objective (loss) function iteratively. However, since the loss function is non-convex, convergence to the global minimum is very unlikely, just because of the sheer number of local minima of the loss landscape. Though there are several other factors involved too, choosing the loss function depends primarily on how the problem is defined, such as classification or regression. Another factor is: whether the dataset has a class imbalance or not? Depending on the loss function, the output layer structure also changes. Some example classification loss functions are cross entropy loss, hinge loss etc., and regression loss functions include mean squared (L_2) loss, mean absolute (L_1) loss etc.

Cross entropy loss for a multi-class problem with M classes is given by:

$$L_{ce} = -\sum_{c=1}^{M} y_{t,c} \log(p_c) \tag{2.5}$$

where $y_{t,c}$ and p_c are the ground truth label and softmax probability of the c^{th} class respectively. For a binary classification problem ($y_t \epsilon \{0, 1\}$), the cross entropy loss boils down to:

$$L_{ce} = -[y_t \log(p) + (1 - y_t) \log(1 - p)] \tag{2.6}$$

Mean squared loss (L_{mse}) is given by:

$$L_{mse} = \frac{1}{N}\sum_{0}^{N-1}(y_p - y_t)^2 \tag{2.7}$$

Learning rate is a hyperparameter that has considerable influence on the model performance, which essentially regulates the extent to which the model parameters are updated based on the computed error. It is hard to find the ideal learning rate since it is problem- and dataset-dependent, and is often fixed by trial and error. A too-large value could overlook the local minima and might "jump" too fast to a sub-optimal solution, or sometimes even delay/prevent the convergence. On the contrary, a too-small value might result in very slow learning and the whole training process could even get stuck to a worse solution. Reducing the learning rate with time is useful for better convergence and hence pre-declared learning rate scheduling (exponential decay, step decay etc.) is often used to adjust the learning rate over time. Some optimisation strategies such as Adam (Kingma and Ba, 2015), Adagrad (Duchi et al., 2011) etc. have shown that adapting the learning rate on the fly during the training process is also effective to achieve rapid convergence.

Optimisation strategies. In neural network training, the loss is minimised using an iterative procedure. The choice of optimiser depends on the problem being solved. Stochastic Gradient Descent (SGD) is the most commonly used optimiser, where the weights are updated as follows:

$$w^{'} = w - l_r * g \tag{2.8}$$

where $w^{'}$, w, l_r, and g are the updated weight, previous weight, learning rate and gradient respectively. Aiming faster convergence, the concept of momentum (Polyak, 1964) is often used which allows to increase the velocity v (movement in the parameter space) in the direction with constant gradient descent. When the momentum $m > 0$, the weight update happens as follows:

$$w^{'} = w + m * v - l_r * g \tag{2.9}$$

Besides SGD, several other iterative minimisation strategies such as Nesterov momentum Sutskever et al. (2013), Adam (Kingma and Ba, 2015), Adagrad (Duchi et al., 2011), Adadelta (Zeiler, 2012) etc. have also been proposed.

Mini-batch learning In batch learning, to estimate the gradient, the entire dataset is collectively processed while in stochastic learning each image in the dataset is processed independently. Mini-batch learning takes a middle ground, where a small batch of images is jointly processed, thereby offering the best of both worlds.

Augmented learning Data augmentation is a standard DL practice in which the images in the original training dataset are tweaked to produce more images. These newly synthesised images are added to the existing training dataset, effectively increasing the volume of data on which a more robust model can be learnt and invariance against all predictable transformations can be ensured. This technique is effective in reducing the

overfitting of the models, especially for the parameter-rich networks, thus achieving better generalisation capability. Example transformations for image data include operations such as rotation, flipping, translation, scaling, cropping, zooming in and out, brightness adjustment, noise addition etc. Augmenting the dataset has proved very useful especially for object recognition problems in computer vision, and for speech recognition.

Transfer learning. It is possible to transfer the knowledge across domains (tasks, datasets etc.) using an ML strategy called transfer learning. An illustration is shown in Fig. 2.5. The model weights trained for one task are reused as the starting point to train a new model for a different task. For example, a deep model pre-trained on a very large (generic) computer vision database like Imagenet (Deng et al., 2009) for an object recognition task (with more than 100 class categories) can be reused for an Earth observation task such as lake ice monitoring (with 2-4 classes, depending on the formulation) from satellite images for which relatively very less labelled data is available to train such a deep network. Essentially, the transfer of learnt knowledge (features, parameters etc.) happens from the source to the target domain.

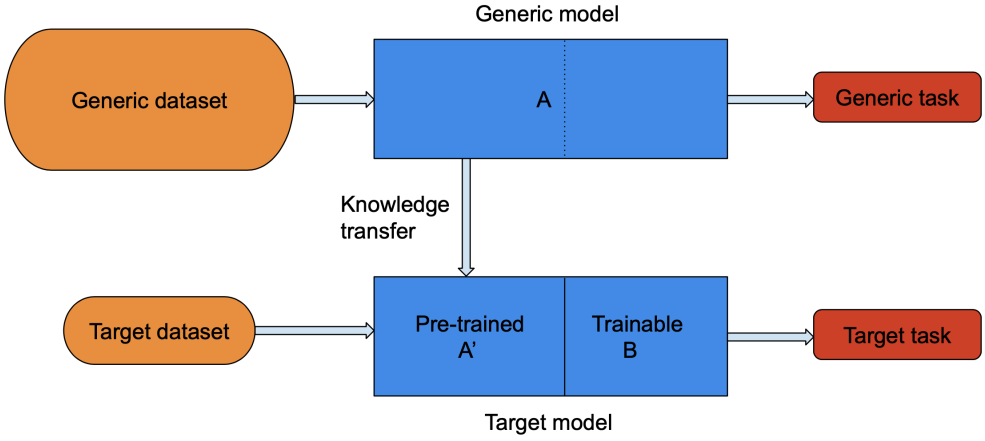


Figure 2.5.: Transfer learning.

Transfer learning is particularly effective when there is significantly more amount of labelled data available to train the first task/setting compared to the second. As opposed to training a network from scratch using the potentially small target dataset, this method substantially reduces the computation overhead by just fine-tuning on the target dataset and usually results in superior performance. The network trained on the big source dataset usually learns very generic image features (especially the initial layers) which can directly be transferred. Fine-tuning of the pre-trained weights in all layers (or a subset) can be done using the target data. However, it is usually recommended to freeze the initial layers which learn very general features and tune just the higher layers.

2.2. Deeplab network

Deeplab v3+ network has been extensively used in Chapters 3, 5 and 6 of this dissertation. Therefore, this section discusses some of the corresponding details.

Deeplab is a state-of-art deep CNN-based semantic segmentation network architecture for which many different versions: v1 (Chen et al., 2015), v2 (Chen et al., 2018a), v3 (Chen et al., 2017), and v3+ (Chen et al., 2018b), were proposed in literature. Deeplab v1 applied the *atrous* (or dilated) convolution, see Fig. 2.6, in order to regulate the learnt feature resolution (output stride). In deeplab v2, efficient multi-scale object segmen-

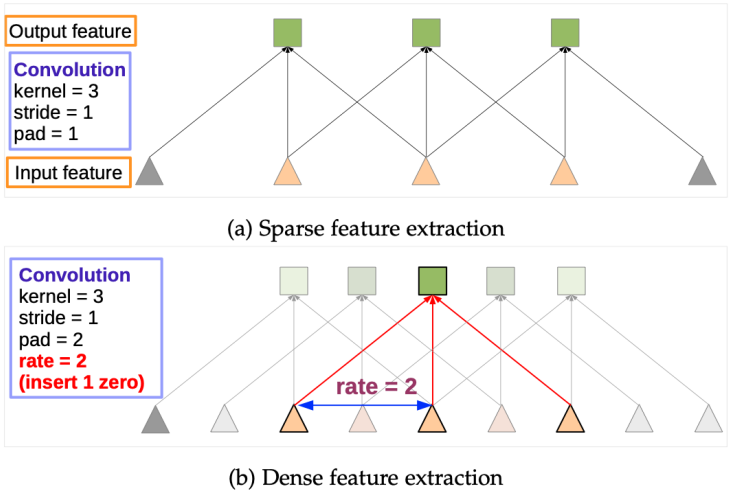


Figure 2.6.: Atrous convolution in 1D. Image courtesy of Chen et al. (2018a).

tation was presented using an Atrous Spatial Pyramid Pooling (ASPP) module which comprised of filters with numerous sampling rates and hence various receptive fields, see Fig. 2.7. Deeplab v3 acquired longer-range knowledge by augmenting the ASPP mod-

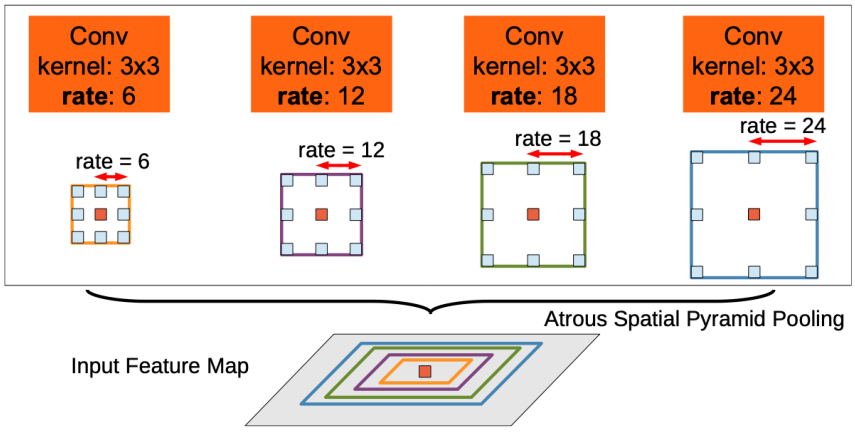
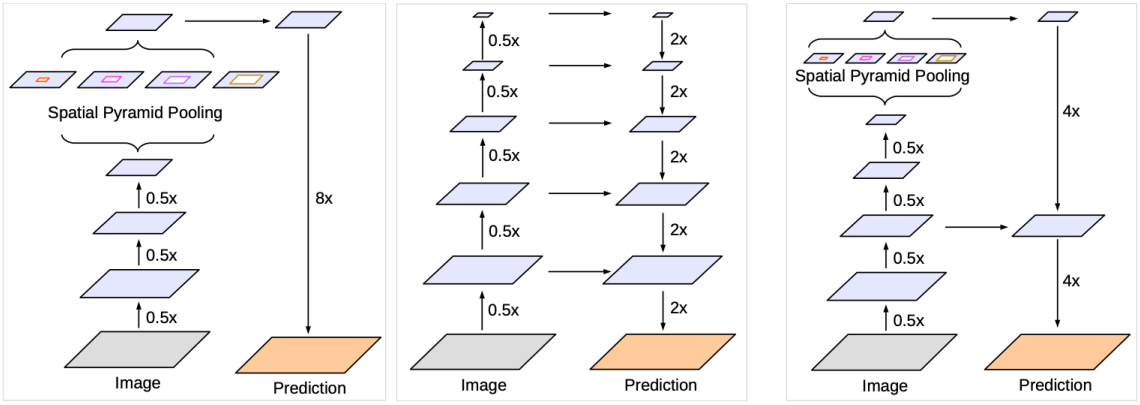


Figure 2.7.: An example ASPP module. Image courtesy of Chen et al. (2018a).

ule with image-level features. The batch normalisation technique was also applied during training. Deeplab v3+ brought further improvement, especially by learning accurate class

boundaries by adding a simple but useful decoder module in an encoder-decoder architecture (see Fig. 2.8b). In such networks, at first, the encoder employs a series of pooling (downsampling) and strided convolution layers to generate low spatial-dimensional feature representations targeting to increase the size of the receptive field and to achieve some translation invariance. Then, the decoder upsamples these features using a sequence of unpooling and deconvolution operations. To preserve higher-resolution information, skip connections are used. Encoder-decoder networks are particularly useful for the task of semantic image segmentation.



(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

Figure 2.8.: Deeplab v3+. Image courtesy of Chen et al. (2018b).

By exploring the incoming features with various stages of filters or pooling operations, a network with inbuilt spatial pyramid pooling modules (refer Fig. 2.8a) encodes multi-scale contextual knowledge. On the other hand, networks with encoder-decoder architecture learn fine class borders by gradually regaining spatial knowledge. Deeplab v3+ blends both these mechanisms, see Fig. 2.8c, and performs semantic image segmentation.

One of the main reasons why CNNs perform well on semantic segmentation tasks is that such networks are capable to learn contextual relationships over large receptive fields. However, high-frequency details are often lost as a byproduct of spatial downsampling occurring due to pooling and convolution operations, which results in inaccurate learning of the class boundaries. Encoder-decoder architecture with built-in skip connections can however curtail this issue to a large extent.

2.3. Deep learning in remote sensing

After the huge success of DL in computer vision, especially for image recognition, deep networks have been applied for solving Earth science problems as well. A handful of comprehensive reviews (Zhu et al., 2017, Ma et al., 2019, Yuan et al., 2020) discussed the challenges and recent advances of DL methodologies applied to remote sensing problems. Additionally, some approaches are reviewed in Chapters 3, 4, 5, and 6. The following two paragraphs summarise few prominent works, not included in those chapters.

Some notable works that used CNNs were proposed for the semantic segmentation of remote sensing images. Volpi and Tuia (2017) learnt hierarchical feature representations using CNNs ("downsample-then-upsample architecture") to perform per-pixel land-cover classification in very high spatial-resolution imagery. To learn the accurate semantic labels, rough downsampled spatial maps were first generated using convolution operations followed by upsampling them back to the original resolution using deconvolution operations. This approach produced excellent results on two tested datasets (Potsdam and Vaihingen with 5cm and 9cm spatial resolution respectively). While Volpi and Tuia (2017) demonstrated their approach using aerial imagery, Maggiori et al. (2017) performed per-pixel classification on Pléiades satellite imagery. In addition to proposing an end-to-end CNN architecture that accurately learnt semantic labels, this work presented an improved training strategy using a two-step procedure to efficiently handle the case with imperfect training data. First, the CNN was trained on inaccurate reference data, followed by fine-tuning on a small dataset with correct labels. Another contribution of this work was that they used a multi-scale neuron module that precisely learnt class boundaries. Later, Marmanis et al. (2018) proposed an end-to-end deep CNN for semantic segmentation of aerial imagery. This approach utilised semantically aware edge detection to improve the segmentation performance on Vaihingen and Potsdam datasets and demonstrated experiments using two different CNN architectures.

Marmanis et al. (2016) proposed to use transfer learning for aerial scene recognition. A CNN model pre-trained on the Imagenet database was transferred and fine-tuned in a supervised manner on the relatively smaller UCML benchmark dataset and convincing results were achieved. Xiao et al. (2018) used fully convolutional Densenets (Jégou et al., 2016) to perform lake ice monitoring with freely available terrestrial webcam imagery and produced excellent results including good cross-winter generalisation. However, the cross-camera generalisation performance was just satisfactory. This thesis proposes a solution in Chapter 3 to mitigate this shortcoming.

2.4. Lake ice physics and remote sensing

According to Michel and Ramseier (1971), three layers (primary, secondary and superimposed) can be spotted in lake ice. The unique structure of the crystals and the impurity content (gas bubbles) makes each layer different. The secondary layer develops below the primary layer and comprises relatively big crystals (Leppäranta and Kosloff, 2000). On the other hand, the superimposed layer appears above the primary layer and mainly contains a snow-water mixture (Palosuo, 1965, Michel and Ramseier, 1971).

Except for minor fractures and cracks, ice cover remains static in small lakes. In large lakes, ice displacements triggered by wind might occur. In medium-sized lakes, the ice is normally static in cold winters and mobile in mild winters (Burda, 1999, Kondratyev and Filatov, 1999, Wang et al., 2006). Kirillin et al. (2012) discussed the physical properties of lake ice cover, including the dynamics of ice freeze-up and break-up periods. Various factors such as air temperature, lake inflows, internal waves, depth, heat flow and storage etc., have an impact on the ice formation process during freeze-up. On the other hand, ice thawing during the break-up period is primarily steered by air temperature (Brown

and Duguay, 2010). For more details on lake ice physics, refer to Kirillin et al. (2012).

Lake ice remote sensing can be roughly classified into two categories: optical and radar remote sensing. The latter can be further subdivided into active and passive remote sensing. For a general overview of remote sensing of lake ice, refer to Duguay et al. (2015).

As opposed to open water, lake ice (and snow on ice) reflects back a large amount of radiation at the near-infrared and visible wavelengths. Hence, the corresponding spectral bands have cues to distinguish frozen and non-frozen lake pixels (Oke, 1987, Kirillin et al., 2012, Svacina et al., 2014, Wu et al., 2021a). Additionally, parameters such as impurity content in ice, surface coarseness, snow wetness and snow age affect the amount of energy reflected back (Wynne and Lillesand, 1993). Existing approaches for optical remote sensing of lake ice are discussed in Sections 3.1, 4.2, 5.2 and 6.1.

Though passive optical satellite sensors are useful for lake ice detection, they are not effective in cloudy scenarios and during the nighttime as they depend on the Sun's energy reflected back from the Earth's surface (Arp et al., 2013, Murfitt and Duguay, 2021). For the high-latitude lakes, low solar elevation during the winter months poses an additional challenge to detect lake ice using optical sensors (Latifovic and Pouliot, 2007). The cloud issue can be counteracted by remotely sensing lake ice using active and passive microwave sensors. Typically, L, C, X and Ku radar bands (with wavelengths $\approx$ 25cm, $\approx$ 5.6cm, $\approx$ 3.1cm, and $\approx$ 2.2cm respectively) are used for lake ice observation. Vital cues on brightness temperatures are significant to monitor lake ice using passive microwave sensors. (Lemmetyinen et al., 2009). However, according to a recent survey (Murfitt and Duguay, 2021), active microwave remote sensing (including Synthetic Aperture Radar [SAR]) is the most widely adopted technology to monitor lake ice. Details on SAR backscatter properties of lake ice, water and snow are discussed in Section 5.3. Some previous studies for lake ice observation using radar data are reviewed in Sections 5.2 and 6.1.

3 Ice Monitoring in Swiss Lakes from Optical Satellites and Webcams using Machine Learning

Manu Tom, Rajanie Prabha, Tianyu Wu, Emmanuel Baltsavias, Laura Leal-Taixe, Konrad Schindler
MDPI Remote Sensing, 12(21), 3555, 2020
(Author version; for typeset version please refer to the original journal paper.)

Abstract

Continuous observation of climate indicators, such as trends in lake freezing, is important to understand the dynamics of the local and global climate system. Consequently, lake ice has been included among the Essential Climate Variables (ECVs) of the Global Climate Observing System (GCOS), and there is a need to set up operational monitoring capabilities. Multi-temporal satellite images and publicly available webcam streams are among the viable data sources capable of monitoring lake ice. In this work we investigate machine learning-based image analysis as a tool to determine the spatio-temporal extent of ice on Swiss Alpine lakes as well as the ice-on and ice-off dates, from both multispectral optical satellite images (VIIRS and MODIS) and RGB webcam images. We model lake ice monitoring as a pixel-wise semantic segmentation problem, i.e., each pixel on the lake surface is classified to obtain a spatially explicit map of ice cover. We show experimentally that the proposed system produces consistently good results when tested on data from multiple winters and lakes. Our satellite-based method obtains mean Intersection-over-Union (mIoU) scores $> 93\%$, for both sensors. It also generalises well across lakes and winters with mIoU scores $> 78\%$ and $>80\%$ respectively. On average, our webcam approach achieves mIoU values of $\approx 87\%$ and generalisation scores of $\approx 71\%$ and $\approx 69\%$ across different cameras and winters respectively. Additionally, we generate and make available a new benchmark dataset of webcam images (*Photi-LakeIce*) which includes data from two winters and three cameras.

3.1. Introduction

Climate change is one of the main challenges for humanity today and there is a great necessity to observe and understand the climate dynamics and quantify its past, present, and future state (Rolnick et al., 2019, Sharma et al., 2019). Lake observables such as ice duration, freeze-up, and break-up dynamics etc. play an important role in understanding climate change and provide a good opportunity for long-term monitoring. Lake ice (as part of lakes) is therefore considered an Essential Climate Variable (ECV, `https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables`) of the Global Climate Observing System (GCOS). In addition, the European Space Agency (ESA) encourages climate research and long-term trend analysis through the Climate Change Initiative (CCI, `https://www.esa.int/Applications/Observing_the_Earth/Space_for_our_climate/ESA_s_Climate_Change_Initiative`). This consortium recently addressed the following variables: Lake water level, lake water extent, lake surface water temperature, lake ice, and lake water reflectance. Recent research also emphasises the socio-economic and biological role of lake ice (Knoll et al., 2019). Moreover, according to an analysis of data from 513 lakes, winter ice in lakes is depleting at a record pace due to global warming (Sharma et al., 2019). That study also underlined the importance of lake ice monitoring, observing that a comprehensive, large-scale assessment of lake ice loss is still missing. The vanishing ice affects winter tourism, cold-water ecosystems, hydroelectric power generation, water transportation, freshwater fishing, etc., which further emphasises the need to monitor lake ice in an efficient and repeatable manner (Schindler et al., 1990). Interestingly, an investigation of the long-term ice phenological patterns in Northern Hemisphere lakes (Magnuson et al., 2000) observed trends towards later freeze-up (average shift of 5.8 days per 100 years) and earlier break-up (average shift of 6.5 days per 100 years), which was also confirmed by another overview (Duguay et al., 2015). However, a previous 50-year (1951–2000) study (Duguay et al., 2006) based on Canadian lakes confirmed the earlier break-up trend but reported less of a clear trend for freeze-up dates.

The idea of monitoring lake ice for climate studies is not new in the cryosphere research community. A main requirement for monitoring lake ice is high temporal resolution (daily) with an accuracy of ± 2 days for phenological events such as ice-on/off dates (according to GCOS). Among the data sources that fulfil this requirement are optical satellite images such as MODIS and VIIRS. In the following, we delve deeper into the literature on using optical satellite images and webcams for monitoring lake ice.

3.1.1. Optical satellite images for lake ice monitoring

At present, satellite images are the only means for systematic, dense, large-scale monitoring applications. Satellite observations with good temporal as well as spatial resolution are ideal for the remote sensing of lake ice phenology. Optical satellite imagery such as MODIS and VIIRS offer very good temporal resolution and satisfactory spatial resolution, making them a good choice. On the other hand, although sensors such as Landsat-8 and Sentinel-2 have a good spatial resolution, the insufficient temporal resolution rules them out as main sources for monitoring lake ice. Some literature exists which uses op-

tical satellite data for lake ice analysis. Inter-annual changes in the temporal extent and intensity of lake ice and snow cover in the Alaska region have been studied using MODIS imagery (Spencer et al., 2008). In addition, studies by Brown and Duguay (2012) and Kropáček et al. (2013) demonstrated that MODIS data is effective for surveying lake ice. The former approach used MODIS and Interactive Multi-sensor Snow and Ice Mapping System (IMS) snow products to monitor daily ice cover changes. The latter derived ice phenology of 59 lakes (area larger than 100 km^2) on the Tibetan Plateau from MODIS 8-day composite data for the period 2001–2010. The estimated area of open water was compared against the area extracted from high-resolution satellite images (Landsat, Envisat/ASAR, TerraSAR-X and SPOT) and achieved a Root Mean Square (RMS) error of 9.6 days. Recently, Qiu et al. (2019) derived the daily lake ice extent from MODIS data by employing the snowmap algorithm (Hall and Riggs, 2016). The results of this approach were consistent with the reference observations from passive microwave data (AMSR-E and AMSR2, average correlation coefficient of 0.91). Additionally, the MODIS daily snow product was used to derive the lake ice phenology of more than 20 lakes in China (Xinjiang territory) using a threshold-based method (Cai et al., 2020). On average, the estimated freeze-up start and break-up end dates were respectively 7.33 and 4.73 days different (mean absolute error) compared to the reference dates derived from passive microwave data (AMSR-E and AMSR2). Very recently, another threshold-based technique (Zhang and Pavelsky, 2019) was also proposed using MODIS data which achieved a mean absolute error of 5.54 days and 7.31 days for break-up and freeze-up dates respectively.

Lake Ice Cover (LIC), a sub-product of the newly released CCI Lakes (Crétaux et al., 2020) product, provides the spatial cover (spatial resolution of 250 m) of lake ice and the associated uncertainty at a daily temporal resolution. At present, LIC is only available for 250 lakes spread across the globe. However, none of the target lakes in Switzerland are included in this list. Hence, a direct comparison with our results is not feasible. Lake Ice Extent (LIE, `https://land.copernicus.eu/global/products/lie`) is one of the Copernicus Global Land Service near-real-time products derived by thresholding the Top-of-Atmosphere reflectances from Level 1B calibrated radiances of Terra MODIS (Collection 6) for snow-covered ice, snow-free ice, and water. The 250 m resolution product has been validated against ice break-up observations over 34 Finnish lakes spanning four years (2010–2013). However, the LIE product has high uncertainty during the lake freezing period due to low light conditions in the higher latitudes as well as uncertainty in cloud cover detection. In addition, the LIE differs by an average of 3.3 days compared to the in-situ ground truth, not quite meeting the GCOS specification. MODIS snow product (Hall and Riggs, 2016, version 6, `https://modis-snow-ice.gsfc.nasa.gov/uploads/snow_user_guide_C6.1_final_revised_april.pdf`, version 6.1) maps snow and ice cover (including ice on large, inland lakes) at a relatively coarse spatial resolution of 500 m and daily temporal resolution using Earth Observation System (EOS) MODIS data. A comparison of specifications of our machine learning-based product with the operational products is shown in Table 3.1.

Though in many aspects VIIRS and MODIS imagery are similar (Trishchenko and Ungureanu, 2017, Trishchenko, 2019), the former has not been leveraged much to study lake ice. Previously, Sütterlin et al. (2017) proposed to use VIIRS data to retrieve lake ice phenology in Swiss lakes using a threshold approach. Another algorithm (Liu et al.,

Table 3.1.: Comparison of specifications of our machine learning-based product with the operational products such as CCI Lake Ice Cover (CCI LIC), Lake Ice Extent (LIE), MODIS Snow Product (MODIS SP), and VIIRS Snow Product (VIIRS SP).

	CCI LIC	**LIE**	**MODIS SP**	**VIIRS SP**	**Ours**
Temporal resolution	1 day	1 day	1 day	1 day	1 day
Spatial resolution (GSD)	250 m	250 m	500 m	375 m	250 m
Input data	MODIS	MODIS	MODIS	VIIRS	MODIS, VIIRS

2016) used VIIRS to detect inland lake ice in Canada. Using VIIRS as well as MODIS, Trishchenko and Ungureanu (2018) constructed a long time series over Canada and neighbouring regions. They also developed ice probability maps using both sensors. Various approaches have been proposed using the Landsat-8 and/or Sentinel-2 optical satellite images (Miles et al., 2017, Barbieux et al., 2018, Pointner et al., 2018, Williamson et al., 2018). However, we do not go into the details since our work is focused on sensors with at least daily coverage.

3.1.2. Webcams for lake ice monitoring

To some extent, satellite remote sensing can be substituted by close-range webcams (Jacobs et al., 2009), especially in cloudy scenarios. As far as we know, the FC-DenseNet (Tiramisu) model (Jégou et al., 2016) of Xiao et al. (2018) used terrestrial webcam images for the first time for lake ice monitoring application, followed by a joint approach (Tom et al., 2019) which used in-situ temperature and pressure observations, and a satellite-based technique in addition to webcams. We note that these two works presented results only on cameras that capture a single lake (St. Moritz) and the generalisation performance was poor, especially for cross-camera predictions. In this work we achieve better prediction performance using webcams compared to such approaches. In addition, we report results on data from two lakes (St. Moritz, Sihl) and two winters (2016–2017, 2017–2018).

3.1.3. Machine (deep) learning approaches for lake ice monitoring

The literature on lake ice monitoring is vast. However, most works make use of elementary threshold-based or index-based methodologies. While, machine learning approaches have been successfully leveraged for various remote sensing and environment monitoring applications, their use for lake ice detection remains under explored. We intend to fill this research gap in our paper. To our knowledge, the previous version of our satellite-based method (Tom et al., 2018) and Xiao et al. (2018) applied machine learning techniques for the first time to detect ice in lakes. Very recently, we also proposed a preliminary version of our webcam-based methodology (Prabha et al., 2020). In this paper, we extend our works (Tom et al., 2018, 2019, Prabha et al., 2020) and perform thorough experimentation, targeting an operational system for lake ice monitoring. For completeness, we

mention that, very recently, we have also explored the possibility to detect lake ice using Sentinel-1 SAR data with deep learning (Tom et al., 2020a).

3.1.4. Motivation and contributions

Existing observations and data on lake ice from local authorities, fishermen, ice-skaters, police, internet media, publications, etc. are not well documented. Additionally, there has been a significant decrease in the number of such field observations in the past two decades (Lenormand et al., 2002, Duguay et al., 2015). At the same time, the potential of different remote sensing sensors has been demonstrated to measure the occurrence of lake ice. In this context we note that, for our target region of Switzerland, the database at the National Snow and Ice Data Centre (NSIDC) currently includes only the ice-on/off dates of a single lake (St. Moritz), and only until 2012. Given the need for automated, continuous monitoring of lake ice, we propose to explore the potential of artificial intelligence to support an operational system. In this paper, we put forward a system which monitors selected Alpine lakes in Switzerland and detects the spatio-temporal extent of ice and in particular the ice-on/off dates. Though satellite data is the best operational input for global coverage, close-range webcam data can be very valuable in regions with large enough camera networks (including Switzerland). Firstly, we use low spatial resolution (250–1000 m) but high temporal resolution (1 day) multispectral satellite images from two optical satellite sensors (Suomi NPP VIIRS, `https://ncc.nesdis.noaa.gov/VIIRS/`, Terra MODIS, `https://terra.nasa.gov/about/terra-instruments/modis`). Here, we tackle lake ice detection using XGBoost (Chen and Guestrin, 2016) and Support Vector Machines (Cortes and Vapnik, 1995, SVM). Secondly, we investigate the potential of images from freely available webcams using Convolutional Neural Networks (CNN), for the independent estimation of lake ice. Given an input webcam image, such networks are designed to predict pixel-wise class probabilities. Additionally, we use webcam data for the validation of results from satellite data.

3.2. Materials and methods

3.2.1. Definitions used

By definition, ice-on date is the first day a lake is totally or in great majority frozen (≈90%), with a similar day after it [this is the same definition as in Hendricks Franssen and Scherrer (2008), i.e., end of freeze-up]. Ice-off is used here as the symmetric of ice-on, i.e., the first day after having all or almost all lake frozen, when any significant amount of clear water appears and in the subsequent days this water area increases. We point out that our ice-off date marks the start of melting (break-up), such that the two dates symmetrically delimit the fully frozen period. As far as we know there is no universally accepted definition of ice-on/off dates. Hence, in the scientific cooperation we had with MeteoSwiss we adopted the above definition which is consistent throughout this work. Ice thickness plays no role in this definition. In very rare cases, in Switzerland, there may be more than one such date. In those cases we use the latest ice-on and the earliest ice-off dates. Some researchers, especially in North America and in the NSIDC database,

define ice-off as the end of break-up, when almost everything is water (Duguay et al., 2015). That date can also be retrieved with our scheme, without any changes to the methodology. Clean pixels are those that are totally within the lake outline. In all subsequent investigations with satellite image data, only the cloud-free clean pixels are used. Additionally, non-transition dates are the days when a lake is mostly (≈90% or above) frozen (ice, snow) or non-frozen (water) while the partially frozen days are termed as transition dates.

3.2.2. Target lakes and winters

Using satellite images, we processed the Swiss Alpine lakes: Sihl, Sils, Silvaplana, and St. Moritz, see Fig. 3.1. To assess the performance, the data from two full winters (16–17 and 17–18) are used, including the relatively short but challenging freeze-up and break-up periods. In each winter, we processed all available dates from the beginning of September until the end of May. The target lakes exhibited moderate to high difficulty, with a variable area (very small to mid-sized), altitude (low to high), and surrounding topography (flat/hilly to mountainous), and they freeze more or less often. More details of the target lakes are shown in Table 3.2, including the details of the nearest meteorological stations. For completeness, the temperature and precipitation data near the observed lakes were also plotted (see Fig. 3.2 for 2016–2017 winter months). Additionally, we processed three different webcams monitoring lakes St. Moritz and Sihl from the same two winters. For satellite images, the lake outlines are digitised from Open Street Map (OSM) and have an accuracy of ≈25–50 m. For webcams, our algorithm automatically determines the lake outline.

Table 3.2.: Characteristics of the target lakes (data mainly from Wikipedia). Latitude (lat, °North), longitude (lon, °East), altitude (alt, m), area (km^2), volume (vol, Mm^3), and the maximum and average depth [depth(M,A)] in m are shown. Additionally, for each lake, the nearest meteorological station (MS) is shown together with the corresponding latitude, longitude, and altitude.

Lake	Lat, Lon, Alt	Area, Vol	Depth (M, A)	Remarks	MS, Lat, Lon, Alt
Sihl	47.14, 8.78, 889	11.3, 96	23, 17	frozen most years	Einsiedeln, 47.13, 8.75, 910
Sils	46.42, 9.74, 1797	4.1, 137	71, 35	freezes every year	Segl-Maria, 46.43, 9.77, 1804
Silvaplana	46.45, 9.79, 1791	2.7, 140	77, 48	freezes every year	Segl-Maria, 46.43, 9.77, 1804
St. Moritz	46.49, 9.85, 1768	0.78, 20	42, 26	freezes every year	Samedan, 46.53, 9.88, 1708

3.2.3. Data

We use data from three different type of sensors for lake ice monitoring. Parameters of all these data types are shown in Table 3.3.

Optical satellite images

Both MODIS (aboard Terra [`https://terra.nasa.gov`] and Aqua [`https://aqua.nasa.gov/modis`] satellites) and VIIRS (Suomi NPP, `https://www.nasa.gov/mission_pag`

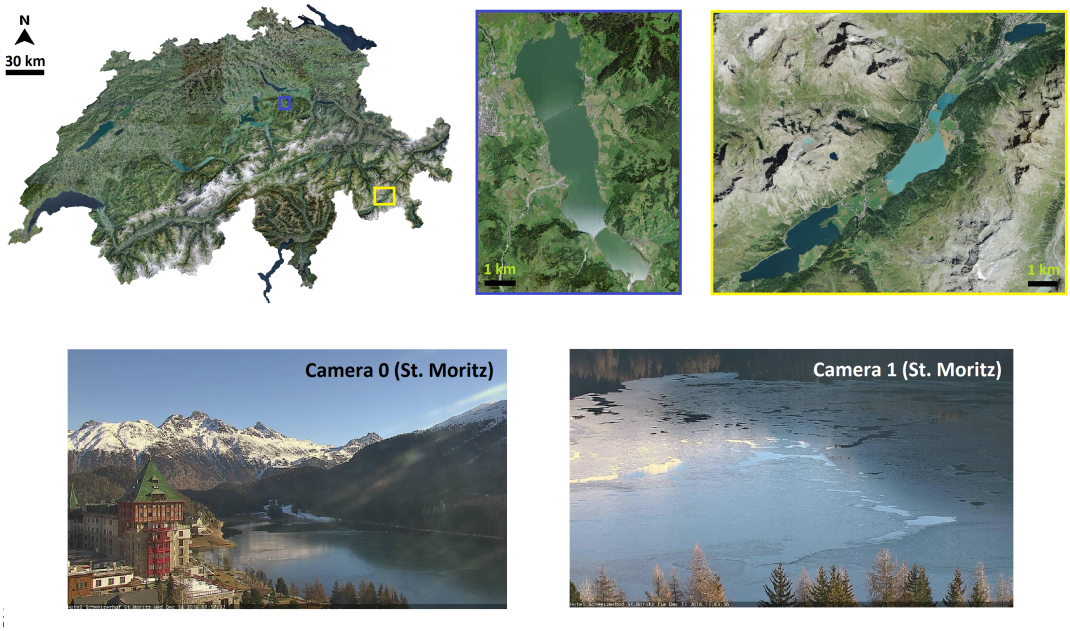


Figure 3.1.: On the first row, left image shows the orthophoto map of Switzerland (source: Swisstopo (`https://www.swisstopo.admin.ch/`)). Regions around the four target lakes (shown as blue and yellow rectangles on the map) are zoomed in and shown on the right side of the map (lake Sihl on the left, region around lakes Sils, Silvaplana, St. Moritz on the right). On the second row, the image footprints of two webcams monitoring lake St. Moritz are displayed (Camera 0 and 1 images were captured on 14 December 2016 and 13 December 2016 respectively when the lake was partially frozen). Best if viewed on screen.

Table 3.3.: Parameters of the used data (GSD = Ground Sampling Distance).

	MODIS	**VIIRS**	**Webcams**
Temporal resolution	1 day	1 day	1 hour (typically)
Spatial resolution (GSD)	250–1000 m	375–750 m	ca. 4 mm to 4 m
Spectral resolution	36 bands (0.41–14.24 μm)	22 bands (0.41–12.01 μm)	RGB
Radiometric resolution	12 bits	12 bits	8 bits
Costs	free	free	free
Availability	very good	very good	depending on location
Cloud mask issues	slight	slight	NA
Cloud problems	severe	severe	negligible

`es/NPP` satellite) images are freely available and have high temporal resolution. Due to the lower quality of Aqua imagery we used Terra images in our analysis. Additionally, following the approach of Tom et al. (2018), we used only 12 MODIS bands and discarded the rest. For our MODIS analysis, we downloaded the following products: MOD02 (calibrated and geolocated radiance, level 1B), MOD03 (geolocation), and MOD35 (48-bit cloud mask) from the LAADS DAAC (Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center, `https://ladsweb.modaps.eosdis.nasa.g`

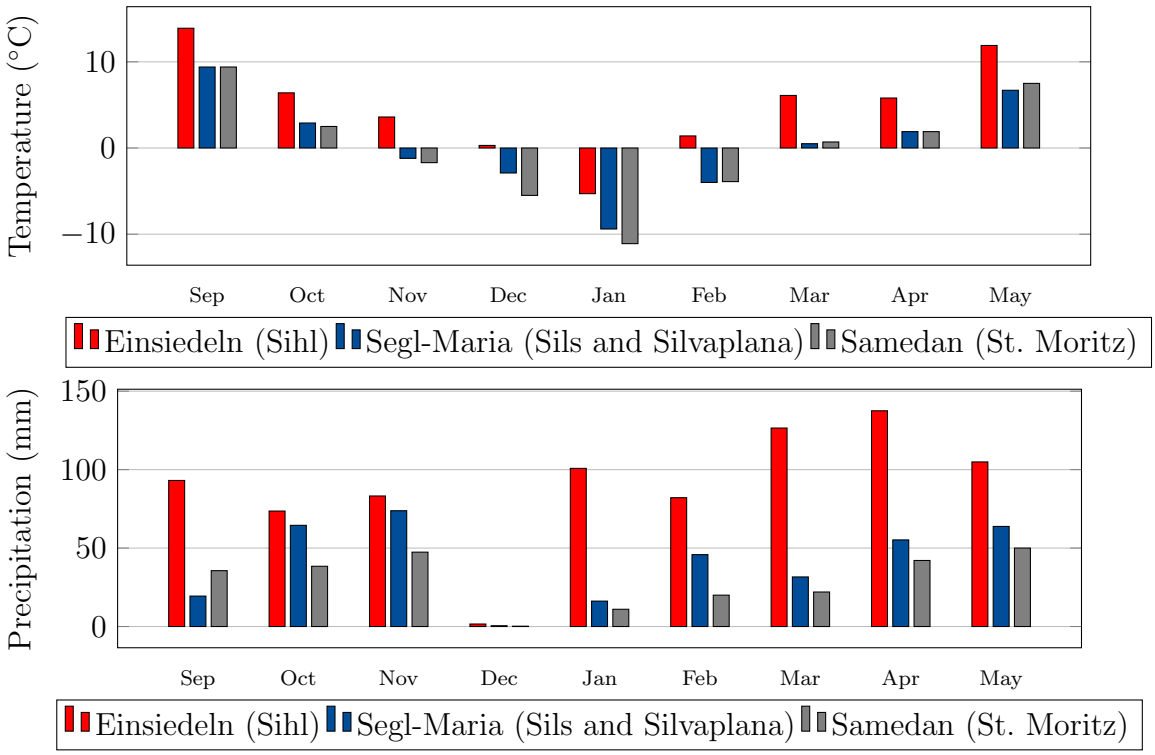


Figure 3.2.: Bar graphs of mean monthly air temperature 2 m above ground (**top**) and total monthly precipitation (**bottom**) in winter 2016–2017, recorded at the meteorological stations closest to the respective lakes. Data courtesy of MeteoSwiss.

ov/) archive. Note that, only the I-bands are used in our VIIRS analysis. See Fig. 3.3 for the spectral range of MODIS and VIIRS bands used in our approach.

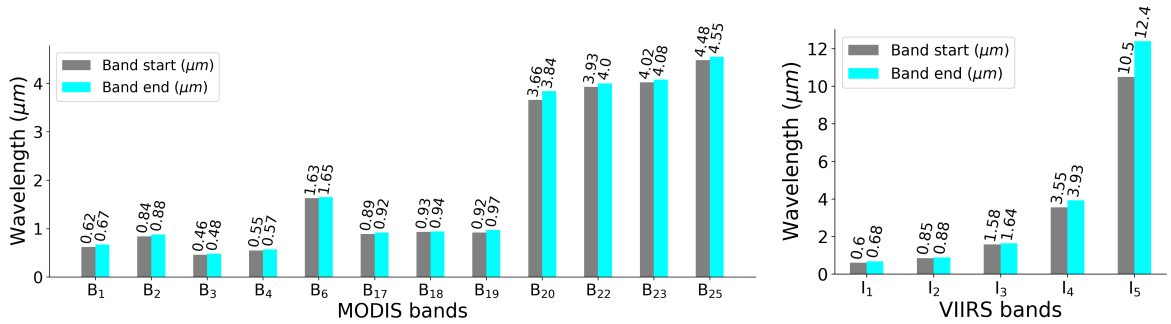


Figure 3.3.: Spectral range of MODIS (**left**) and VIIRS (**right**) bands used in our analysis. The start and end wavelengths are shown for each band.

Technicalities about the processed satellite data are shown in Tables 3.4 and 3.5. It can be seen that we analysed relatively less data in winter of 17–18 as opposed to the previous winter, due to the fact that winter 17–18 was more cloudy than 16–17 in the regions of interest. We only processed the dates when a lake was at least 30% cloud-free, which effectively lowered the temporal resolution from 1 day to approximately 2

days. The effective temporal resolution varies across sensors and winters (see Table 3.5). Additionally, for lakes Sihl and St. Moritz, there were more transition days in winter 17–18. Throughout, we used the non-transition dates for training the SVM model as referred to in Section 3.3.1. This factor along with class-imbalance explains why the decrease in data is more evident for the class frozen. Note also that the transition dates are more likely to occur near the freezing and thawing periods. One can note class imbalance in the dataset of both winters. In each winter, we processed all available acquisitions during the period from September till May, while the lakes were typically fully (or mostly) frozen during only a small subset of these dates. Moreover, the class imbalance was alarmingly high for lake Sihl. This is because Sihl had a moderate freezing frequency compared to the other three lakes, because of its lower altitude and larger area, see Table 3.2. Note also that the presence of a dam near the northern part of lake Sihl makes its freezing pattern relatively less natural.

Even after collecting data from a full winter, very few pixels are available to train a machine learning-based system using MODIS imagery, due to the low spatial resolution. For instance, in every acquisition, there exist only four clean pixels for lake St. Moritz, refer to Table 3.5. This problem is even worse for VIIRS where there is no clean pixel at all for the same lake (see Table 3.5). Note that from Table 3.4, the total number of VIIRS (clean) pixels processed is significantly less compared to MODIS mainly due to the lower spatial resolution (see also Table 3.5). A challenge for machine learning is the scarcity of lake pixels. Note also that the small number of pixels per lake makes a correction of the lake outlines' absolute geolocation a necessity (refer Section 3.2.4). Furthermore, it is highly probable during the transition periods that both frozen and non-frozen classes coexist within a single clean pixel (mixels). For this reason, we also generate the probability for each pixel to be frozen as an end result, especially targeting such mixels during the transition periods. Note also that data hungry deep learning approaches cannot be deployed, as they cannot be reliably trained with such small datasets.

Table 3.4.: Total number of clean, cloud-free pixels on non-transition dates from MODIS (M) and VIIRS (V) sensors (at least 30% cloud-free days) used in our experiments. Fr and N-Fr represent frozen and non-frozen respectively.

	Winter	**Sihl**		**Sils**		**Silvaplana**		**St. Moritz**		**Total**	
		M	**V**	**M**	**V**	**M**	**V**	**M**	**V**	**M**	**V**
Fr	16–17	4137	1919	2345	894	1736	739	157	—	8375	3552
N-Fr	16–17	13568	4598	3019	1051	1965	765	191	—	18743	6414
Fr	17–18	1005	198	1858	722	1169	591	124	—	4156	1511
N-Fr	17–18	11804	4311	2435	784	1574	621	140	—	15953	5716
Total		30514	11026	9657	3451	6444	2716	612	—	47227	17193

Webcam images

We reported our results on various cameras with different intrinsic and extrinsic parameters, which are freely available. For the experiments, we manually removed some unusable

Table 3.5.: Dataset statistics. M/V format displays the respective numbers of MODIS/VIIRS. Both transition (trans) and non-transition (non-trans) days are shown. Fr and N-Fr represent frozen and non-frozen respectively. The effective temporal resolution (shown as 'resolution') and fraction of transition dates over all processed dates (Trans fraction) are also shown. #Pixels (clean) displays the number of clean pixels per acquisition.

Lake	**#Pixels (Clean)**	**Winter**	**Non-Trans Days**		**Trans Days**	**Resolution (Days)**	**Trans Fraction**
			N-Fr	**Fr**			
Sihl	115/45	16–17	98/87	32/33	12/11	1.9/2.1	0.09/0.08
		17–18	90/88	8/6	24/22	2.2/2.4	0.20/0.19
Sils	33/11	16-17	70/73	57/59	33/30	1.7/1.7	0.21/0.19
		17–18	60/57	49/48	25/32	2.0/2.0	0.19/0.23
Silvaplana	21/9	16–17	66/66	63/59	33/34	1.7/1.7	0.20/0.21
		17–18	58/58	43/54	27/31	2.1/1.9	0.21/0.22
St. Moritz	4/0	16–17	79/—	65/—	14/—	1.7/—	0.09/—
		17–18	64/—	58/—	16/—	2.0/—	0.12/—

images, examples are shown in Fig. 3.4. We point out that for oblique webcam viewpoints, the GSD varied greatly between nearby and distant parts of a lake, as does the angle between the viewing rays and lake surface. As a consequence, webcam images are hard to interpret in the far field, in practice usable distances tend to be up to ≈1 km. We note that the usable distance also depends on the surface material, e.g., snow on ice can be detected at further distances where it is already impossible (for humans as well as machines) to distinguish black ice from water.

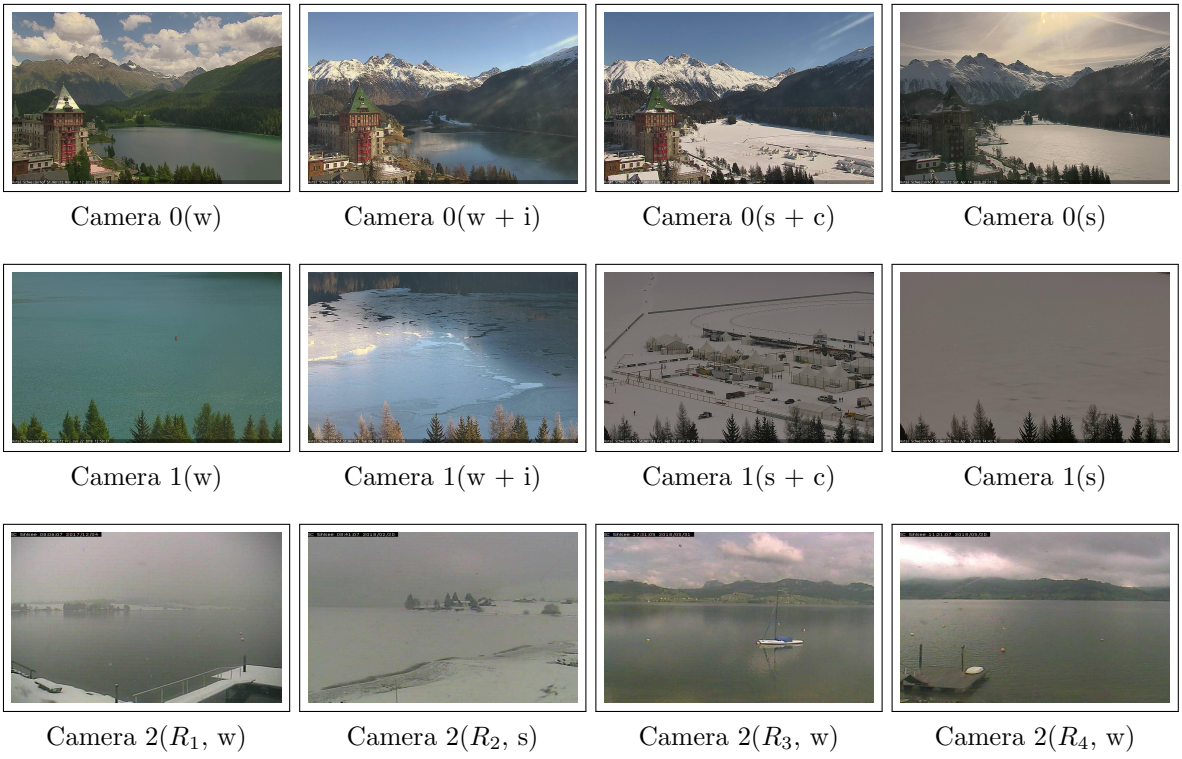
Figure 3.4.: Example images that were discarded from the dataset due to bad illumination (**left**), sun over-exposure (**middle**), and thick fog (**right**).

We make available (`https://github.com/czarmanu/photi-lakeice-dataset`) a new webcam dataset, termed *Photi-LakeIce*, for lake ice monitoring and report our results on it. Sample images and details of the dataset are presented in Fig. 3.5 and Table 3.6 respectively. RGB images (and the corresponding ground truth annotations) from two lakes (Sihl and St. Moritz) and two winters (2016–2017, 2017–2018) are included in the dataset. Though the camera mounted at Hotel Schweizerhof in St. Moritz is rotating, in our analysis we consider it as two different fixed cameras (camera 0 and camera 1, see Fig. 3.6). The major difference between these two streams is image scale: Camera 0 captures images with larger GSD compared to camera 1. Another camera that monitors Sihl is non-stationary, but captures the lake at the same scale (refer Fig. 3.5 row 3). Hence, we consider it as a single rotating camera (camera 2). Our dataset is not limited to but includes images with different lighting conditions (due to the sun's angle, time of

the year, presence of clouds, etc.), shadows (from both clouds and nearby mountains), fog conditions (we remove the extreme cases but keep the images from slightly foggy days), wind scenarios, etc.

Table 3.6.: Details of the *Photi-LakeIce* dataset. Lat and Long respectively denote latitude (°North) and longitude (°East) of the approximate camera location. #16–17 and #17–18 represent the total number of images from winters 16–17 and 17–18 respectively. Res stands for resolution and H and W represent height and width of the image in pixels (after cropping).

Name	**Lake (Lat, Long)**	**Model**	**#16–17**	**#17–18**	**Res (H × W)**
Camera 0	St. Moritz (46.50, 9.84)	AXIS Q6128-E	820	474	324 × 1209
Camera 1	St. Moritz (46.50, 9.84)	AXIS Q6128-E	1180	443	324 × 1209
Camera 2	Sihl (47.13, 8.74)	unknown	500	600	344 × 420



Camera 0(w) | Camera 0(w + i) | Camera 0(s + c) | Camera 0(s)

Camera 1(w) | Camera 1(w + i) | Camera 1(s + c) | Camera 1(s)

Camera 2(R_1, w) | Camera 2(R_2, s) | Camera 2(R_3, w) | Camera 2(R_4, w)

Figure 3.5.: Photi-LakeIce dataset. Rows 1 and 2 display sample images from cameras 0 and 1 (St. Moritz) respectively. Row 3 shows example images of camera 2 (Sihl, non-stationary, and some rotations [R_1, R_2, etc.] are also displayed). State of the lake: water(w), ice(i), snow(s), and clutter(c) is also displayed in brackets.

To study the class imbalance in our dataset, we plot the class distribution, individually for each camera and winter, see Fig. 3.7. It can be inferred that the classes are highly imbalanced in most of the sub-datasets, where ice and clutter classes suffer the most. In Fig. 3.7, we show the percentage of the class background in addition to the four main classes. Note that the percentage of clutter in camera 0 is less compared to camera

1. Note also, camera 1 has almost zero background while the lake area (foreground) to background ratio for Sihl is too low, making it a very challenging case. Additionally, the number of ice pixels is consistently low in all the cameras across all years. It will not be surprising if the performance of classes clutter and ice are not good in a relative sense. Note that the background class frequencies differ from one year to another even for the same camera, since in each year the foreground-background separation was done by different human experts. The difference is even more so for camera 2 (Sihl), since it is rotating.

Figure 3.6.: Two webcams monitoring lake St. Moritz along with their approximate coverage. Image courtesy of Google.

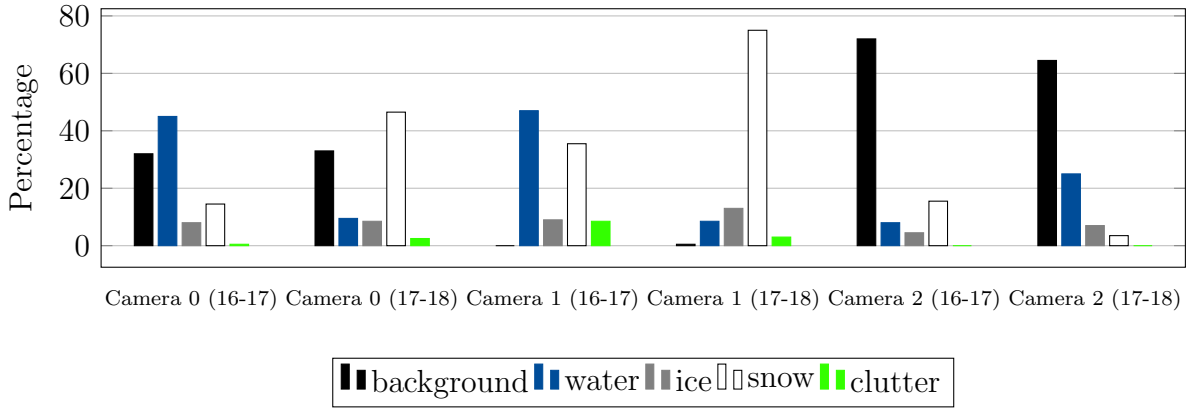


Figure 3.7.: Bar graphs displaying class imbalance (including the class *background*) in our dataset. Ice and clutter are the under-represented classes.

Ground truth generation for webcam analysis

The main difficulty in designing a machine (deep) learning system is the requirement for accurately labelled data. However, to generate pixel-wise labels, the interpretation of webcam images is challenging for several reasons. Image quality is limited and off-the-shelf webcams only offer poor radiometric and spectral resolution and are subject to adverse lighting conditions like fog, which makes the image interpretation process difficult even for humans (see Fig. 3.8). Besides the limitations of the sensor itself, the cameras are

mounted with a rather horizontal viewing angle such that large parts of the water body can be observed. As a result, large differences in GSD within a single image are present. Significant intra-class appearance differences exist throughout image sequences. This is caused by different ice structures, partly frozen water surfaces, waves, varying illumination conditions, reflections, and shadows. Furthermore, inter-class appearance similarities exist, which impedes automatic interpretation. In fact, even manual interpretation for some examples is impossible without using additional temporal cues. Pixel-wise ground truth annotations are produced by human operators by labelling polygons within the input images using the *LabelMe* tool (`https://github.com/wkentaro/labelme`). For the lake detection task, each pixel is either labelled as foreground (lake) or background. Foreground pixels are further annotated as water, ice, snow, and clutter for lake ice segmentation.

Figure 3.8.: Inter-class similarities and intra-class differences of states snow (s), ice (i), and water (w) in our webcam data.

Ground truth generation for MODIS and VIIRS analysis

To generate the ground truth for our satellite image analysis, for each day, a human expert visually interpreted the state of a lake (completely frozen, partly frozen, completely non-frozen, or partly non-frozen) using webcam images of the same. Note also that most of the freely-available webcams are not optimally installed to monitor lakes. Hence, besides webcam images, we interpret cloud-free Sentinel-2 images whenever available. Additionally, we attempted to use online media reports to enrich the generated ground truth, which however only provided limited information. In our analysis, webcams have ground truth at a better granularity level (hourly, per pixel label) compared to satellite images (daily, global label). Accurately registering webcam pixels with satellite image pixels is beyond the scope of this work, hence we did not transfer the webcam-based per-pixel ground truth to the satellite images.

3.2.4. Methodology

Satellite image analysis

Pre-processing of MODIS data (re-sampling to *UTM32N* coordinate system, re-projection) is done using MRTSWATH (`https://lpdaac.usgs.gov/tools/modis_reprojection_tool_swath`) software. Similarly, VIIRS imagery data is pre-processed (assembling of data granules, re-sampling to *UTM32N*, and mapping) with the SatPy (`https://satpy.readthedocs.io/`) package. VIIRS cloud masks are extracted with H5py (`https://www.h5py.org`) and re-sampled using Pyresample (`https://resample.readthedocs.io`)

and GDAL (`https://gdal.org/`) libraries. Among the 12 selected MODIS bands (refer Section 3.2.3), the lower resolution bands (500 m and 1000 m GSD) were upsampled to 250 m resolution using bilinear interpolation. This step is not necessary for VIIRS analysis, since we use only the I-bands (≈375 m GSD). For both VIIRS and MODIS, we only analysed the images with at least 30% cloud-free clean pixels. In MODIS images, there are also some pixels marked as invalid, which were masked out. For MODIS, we merged the cloudy and uncertain clear bits to construct a binary cloud-mask from the standard cloud-mask product. Similarly, a VIIRS pixel is considered non-cloudy only if it is either confidently clear or probably clear. After Douglas–Peucker generalisation (Douglas and Peucker, 1973), the outlines were further backprojected from the ground coordinate system onto the satellite images to steer the estimation of lake ice. In addition, just the clean pixels were analysed, after rectifying the outlines for absolute geolocation shifts, and backprojecting onto the VIIRS band I_2 (≈375 m GSD), respectively MODIS band B_2 (250 m GSD) as in Tom et al. (2018). For MODIS, the mean offsets in x and y direction were 0.75 and 0.85 pixels, respectively. For VIIRS, the mean offsets were 0 and 0.3 pixels in x, respectively y direction.

Fig. 3.9 displays the block diagram of the proposed lake ice monitoring system using satellite images. Our semantic segmentation methodology is generic and is applicable to both VIIRS and MODIS imagery. Here, we formulated ice detection as a supervised pixel-wise classification problem (two classes: *Frozen* and *non-frozen*). To assess the inter-class separability of different bands, we carried out a supervised variable importance analysis using the *XGBoost* feature learning system (Chen and Guestrin, 2016). The training of that method, a gradient boosting approach based on ensembles of decision trees, makes explicit variable importance conditioned on the class labels. The outcome (F-score) indicates how valuable each feature is in the formation of the boosted (shallow) decision trees within the model. The more a feature (in our case a band) is used to make correct predictions with the decision trees, the higher its relative importance. Though XGBoost is also a classifier by default, we only used the built-in variable selection to automatically determine the most informative bands. For the actual classification based on the selected channels we employed a support vector machine (Cortes and Vapnik, 1995, SVM) classifier, mainly because with SVM it is straight-forward to compare a linear and a non-linear variant.

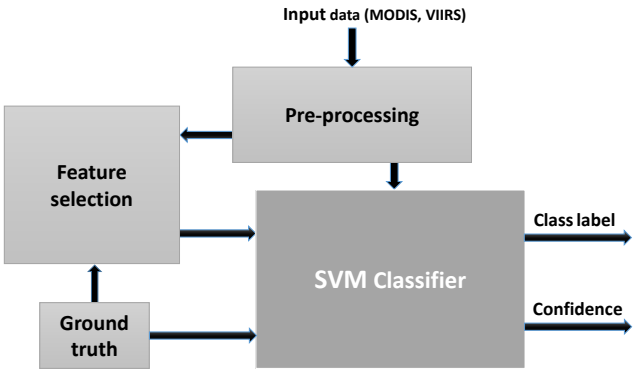


Figure 3.9.: Block diagram of the proposed lake ice detection approach using satellite data.

The 12 usable MODIS bands and 5 I-bands of VIIRS were independently analysed with *XGBoost*. All data from winter 16–17 (see Table 3.4) was used to perform this analysis and the results for both MODIS and VIIRS are shown in Fig. 3.10 (left and right respectively). Bands I_1 and B_1 attained the best scores among the analysed MODIS and VIIRS bands respectively. Furthermore, we plotted the gray-value histograms (see Fig. 3.11) in order to double-check the results generated by *XGBoost*. Due to space limitations, only the histograms for VIIRS are shown. Similar histograms for MODIS can be found in Tom et al. (2017). It can be judged from Fig. 3.11 that the two classes are almost similarly separable in the two near-infrared bands I_1 and I_2. Since those two bands have a similar spectrum and are highly correlated, XGBoost picks only one among them. The same holds for the two near-infrared MODIS bands B_1 and B_2.

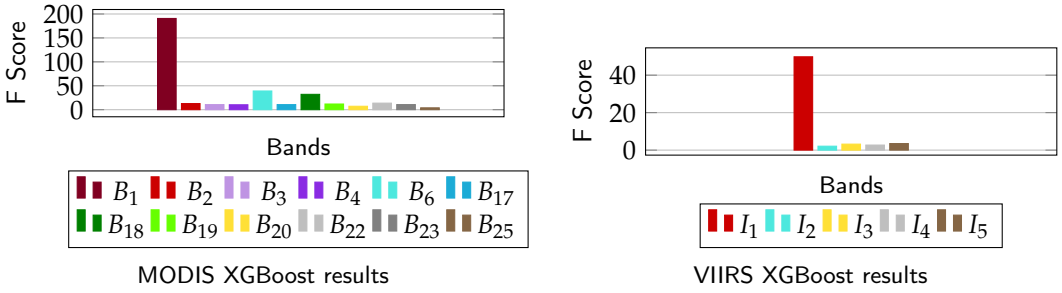


Figure 3.10.: Bar graphs for MODIS (**left**) and VIIRS (**right**) showing the significance of each of the selected bands (12 for MODIS, 5 for VIIRS) for frozen vs. non-frozen pixel separation using the *XGBoost* algorithm (Chen and Guestrin, 2016). All non-transition days from winter 16–17 are included in the analysis.

It is very likely that the physical state of a lake pixel is the same on subsequent days (except during the highly dynamic freezing and thawing periods). Hence, as a post-processing step, multi-temporal analysis (MTA) is applied. For each pixel, a moving average of the SVM class scores is computed along the time dimension. The average is computed within a fixed window length (smoothing parameter) that is determined empirically. Choosing the smoothing parameter is critical, as too-high values can easily wash out the critical dynamics during the transition days. Since the pixels from each MODIS (or VIIRS) acquisition are predicted independently by the trained SVM model, MTA is expected to improve the SVM results by leveraging on the temporal relationships. We test three different averaging schemes: Mean, median, and Gaussian.

Webcam image analysis

Similarly to satellite analysis, we formulated our webcam approach as supervised semantic segmentation problem. Here, we made use of the prominent high-performance deep learning architecture, *Deeplab v3+* (Chen et al., 2018b), with the *Xception65* encoder branch, see Fig. 3.12 (left). That method has a proven track record on other semantic segmentation benchmarks such as PASCAL VOC (Everingham et al., 2015) and Cityscapes (Cordts et al., 2016). Our network classified each pixel in the RGB webcam image as water, ice, snow, or clutter. The clutter class incorporates man-made objects on the lake,

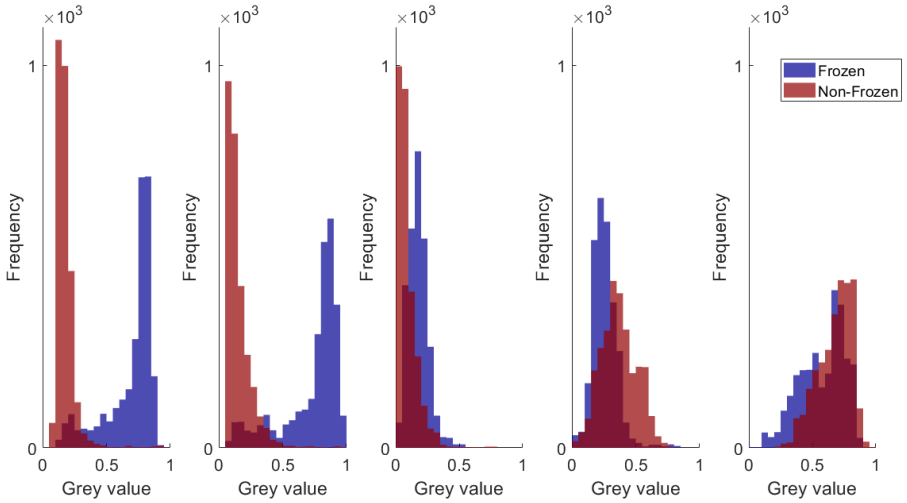


Figure 3.11.: VIIRS grey-value histograms for sanity check (Bands I_1, I_2, I_3, I_4, I_5 are respectively shown from left to right).

e.g., structures built for sport events (such as tents in St. Moritz), boats, etc. Note that, as for satellite images, lake ice segmentation is done only for foreground (lake) pixels.

By integrating the spatial pyramid pooling technique as well as atrous convolution into the standard encoder-decoder architecture, the *Deeplab v3+* network encodes rich contextual information at arbitrary scales and retrieves segment boundaries more precisely. Spatial convolution was applied independently to each channel, followed by 1 × 1 (point-wise) convolution to combine the per-channel outputs. This markedly reduces the computational complexity without any noticeable performance drop. Where needed, these depthwise separable convolutions employ stride 2 in the spatial component, making separate pooling operations obsolete. Note that the atrous (dilated) convolution effectively increases the receptive field without blurring the signal. Using multiple atrous rates makes sure that features are extracted at various spatial scales.

Inspired from *U-net* (Ronneberger et al., 2015), and with an aim to sharpen the class boundaries, in addition to the single skip connection used in *Deeplab v3+* per default, we introduced three more from different flow blocks (entry and mid-level flow) of the *Xception65* encoder to the decoder. We call this new variant *Deep-U-Lab*, see Fig. 3.12 (right). The new feature maps thus generated are concatenated along with the existing ones, for better preservation of high frequency information at the class boundaries.

In order to deal with the shortage (in deep learning terms) of labelled data, we made use of transfer-learning that allows one to re-use knowledge gained from other similar tasks. To do so, we employed a model pre-trained on the PASCAL VOC benchmark dataset (for both lake detection and ice segmentation tasks) rather than starting from scratch. The weights in all layers were fine-tuned since tuning just the final layer did not work, emphasising the fact that even low-level texture characteristics differ between our webcam images and the PASCAL dataset and must be adapted.

In previous works (Xiao et al., 2018, Tom et al., 2019), the lake pixels were manually delineated before inferring their class. Locating the lake pixels makes the job easier for

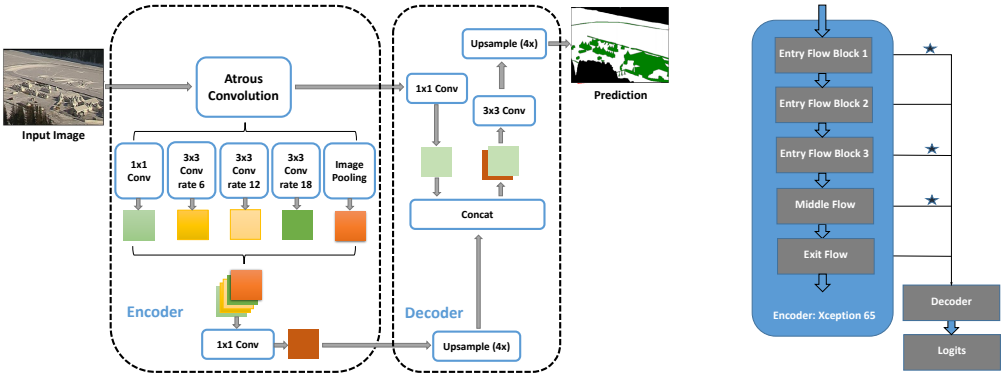


Figure 3.12.: *Deeplab v3+* (**left**) and *Deep-U-Lab* (**right**) architectures. The "⋆" symbol indicates the additional skip connections for *Deep-U-Lab*.

the classifier as it does not have to deal with the spectral appearance outside of the lake. We propose to automate this step, in order to make lake ice observation more practical in operational scenarios. Automated lake detection is very useful especially when scaling up the webcam network to also include non-stationary cameras. Hence, we formulate lake detection as a two-class (background, foreground) segmentation problem and train yet another instance of our *Deep-U-Lab* model. Then, a fine-grained classifier predicts the state (water, ice, snow, and clutter) of lake pixels.

With the intention to minimise overfitting of the model, we performed data augmentation, i.e., more synthetically generated variations were added to the training dataset. This was done by applying random rotations, zooming, up-down, and left-right flipping.

3.3. Results

For our evaluation we use the following error metrics: Recall, precision, overall accuracy, Intersection over Union (IoU, a.k.a. Jaccard index), and mean IoU across classes (mIoU).

3.3.1. Experiments with satellite images

Unless explicitly specified, the experimental settings are the same (but independent) for both MODIS and VIIRS. The time series is divided into two parts: Transition and non-transition dates. During the partially frozen transition days, ground truth annotation was very challenging, as one has to discriminate thin transparent ice from water. Hence, the quantitative results are reported only on non-transition dates. However, qualitative analysis was done on all the available dates. Furthermore, with VIIRS only three lakes (Sihl, Sils, Silvaplana) were processed since there exists no clean pixel for lake St. Moritz (see Table 3.5).

Four-fold cross validation

As a first experiment, the data of all the lakes were combined from the two winters (16–17 and 17–18) and *4-fold cross validation* was performed in order to figure out the optimum SVM parameter settings for detailed experimentation. We did a grid-search for the two main hyperparamaters of SVM (the cost of a misclassification and the kernel width gamma) and found that, for both sensors, the best results with Radial Basis Function (RBF) kernel are obtained with values 10 and 1 for cost and gamma, respectively. With linear kernel, value 0.1 as cost works best for both MODIS and VIIRS. We notice that classification of our dataset using the RBF kernel is fairly sensitive to the selection of hyperparameters, while the linear kernel provides consistent results. Note also, optimum hyperparameters might vary from one dataset to another. Quantitative results of 4-fold cross validation experiments with the optimum parameters are displayed in Table 3.7. For MODIS, the best results were obtained when all 12 bands are used as feature vector (and for VIIRS with all 5 bands). In addition, our results show that the performance of RBF kernel is a bit better compared to the linear counterpart. Additionally, we tested variants that use fewer bands, down to a single band with the highest F-score as selected by XGBoost (B_1 for MODIS and I_1 for VIIRS, refer Fig. 3.10). Since it does not make sense to run an RBF kernel with a single input band, only the linear kernel was tested for this experiment. Even with a single band and a simple linear kernel the results are fairly decent. The results are even better when using the top-five bands of MODIS, but slightly worse than the full 12-band feature vector. Multi-temporal analysis (MTA) improves the results by a very small margin. For both MODIS and VIIRS, MTA with Gaussian kernel (smoothing parameter 3) gives the best results and is therefore used in all further experiments. For the best setting, we show the results in more detail in Table 3.8.

Table 3.7.: The 4-fold cross validation results on MODIS and VIIRS data from two winters (16–17 and 17–18). For the same SVM setup, results without and with multi-temporal analysis (MTA) are shown.

Sensor	Feature Vector	SVM Kernel	with MTA	Overall Accuracy	mIoU
MODIS	B_1	Linear	No	0.91	0.78
MODIS	Top 5 bands	Linear	No	0.93	0.83
MODIS	All 12 bands	Linear	No	0.93	0.84
MODIS	All 12 bands	Linear	Yes	0.93	0.84
MODIS	Top 5 bands	RBF	No	0.96	0.90
MODIS	All 12 bands	RBF	No	0.99	0.98
MODIS	All 12 bands	RBF	Yes	0.99	0.99
VIIRS	I_1	Linear	No	0.93	0.84
VIIRS	All 5 I-bands	Linear	No	0.95	0.88
VIIRS	All 5 I-bands	Linear	Yes	0.95	0.88
VIIRS	All 5 I-bands	RBF	No	0.97	0.93
VIIRS	All 5 I-bands	RBF	Yes	0.97	0.93

We note that feature selection may be beneficial especially with very small training sets. Ideally, SVM automatically prioritizes the more important dimensions in the feature vector, but when only few examples are available, the danger of spurious correlations in

Table 3.8.: Detailed results on MODIS (left) and VIIRS (right) data for the best cases of 4-fold cross validation on combined data from two winters. Acc denotes accuracy. Fr and N-Fr represent classes frozen and non-frozen respectively.

	Recall	**Precision**	**IoU**
Fr	0.99	0.99	0.98
N-Fr	0.99	0.99	0.99
Acc		0.99	
mIoU		0.99	

	Recall	**Precision**	**IoU**
Fr	0.93	0.97	0.90
N-Fr	0.99	0.97	0.96
Acc		0.97	
mIoU		0.93	

less discriminative bands increases. For lake ice detection, where few channels carry most of the information, we recommend the use of automatic feature selection in case the SVM over-fits.

For a practically useful and efficient learning-based monitoring framework, a model should be trained using annotated data from a handful of lakes as well as a few winters, but should be able to predict for lakes and winters not seen during training. To test the performance of our approach in such scenarios, we perform the leave one lake out, respectively leave one winter out cross validation experiments. In all the following experiments, we used all 12 (5) bands of MODIS (VIIRS), optimum hyperparameters chosen by grid-search (cost 10 and gamma 1 for RBF kernel, cost 0.1 for linear kernel), and MTA with Gaussian kernel (smoothing parameter 3).

Leave one lake out cross validation

This experiment evaluates the across-lake generalisation capability of the classifier. We used the SVM model trained on pixels of all but one lake (from both winters) and test on the pixels from the remaining lake, in round-robin mode. MODIS and VIIRS results are presented in Tables 3.9 and 3.10, respectively. As per the results, our models fair well even when trained using only pixels from different lakes. Using both RBF and linear kernels, our algorithm gives excellent results on lakes Sils and Silvaplana consistently with both VIIRS and MODIS data. Table 3.2 shows that both these lakes are similar in many aspects. It is expected that for a learning-based system, predictions are better if the test data is more similar to the training data. The performance of RBF kernel on lakes St. Moritz and Sihl is also good, but not as good as Sils and Silvaplana. Recall that St. Moritz has just four clean pixels per MODIS acquisition (see Table 3.5) and that the absolute geolocation accuracy could be critical for such a small lake. It appears that for St. Moritz, the linear kernel does not generalise unlike other lakes, but we do not draw any firm conclusions based on these results as the lake is too small. Lake Sihl is slightly different compared to the other lakes (altitude, area etc., refer Table 3.2) and hence the SVM encounters a more generalisation loss. Still a mean IoU of 0.78 (corresponding to 93% correctly classified pixels) for MODIS, respectively 0.85 (95%) for VIIRS is a rather good result. For both sensors, the performance of the linear kernel on Sihl is better compared to RBF. Given the fact that Sihl has a somewhat different geography than the other lakes, it appears that the linear kernel generalises better.

Table 3.9.: MODIS *leave one lake out* results. Numbers are in A/B format where A and B represent the results using Radial Basis Function (RBF) and linear kernels, respectively. The better kernel for a given experiment is shown in black, worse kernel in grey.

	Lake Sihl				Lake Sils		
	Recall	**Precision**	**IoU**		**Recall**	**Precision**	**IoU**
Fr	0.82/0.79	0.63/0.78	0.55/0.65	Fr	0.89/0.88	0.97/0.95	0.86/0.85
N-Fr	0.90/0.95	0.96/0.96	0.87/0.92	N-Fr	0.98/0.97	0.92/0.92	0.90/0.89
Acc		0.89/0.93		**Acc**		0.94/0.93	
mIoU		0.71/0.78		**mIoU**		0.88/0.87	
	Lake Silvaplana				**Lake St. Moritz**		
	Recall	**Precision**	**IoU**		**Recall**	**Precision**	**IoU**
Fr	0.91/0.81	0.96/0.97	0.88/0.79	Fr	0.85/0.64	0.93/0.96	0.80/0.63
N-Fr	0.97/0.98	0.93/0.86	0.90/0.85	N-Fr	0.95/0.98	0.88/0.76	0.84/0.75
Acc		0.94/0.91		**Acc**		0.90/0.83	
mIoU		0.89/0.82		**mIoU**		0.82/0.69	

Table 3.10.: VIIRS *leave one lake out* results. Numbers are in A/B format where A and B represent the results using RBF and linear kernels, respectively. The better kernel for a given experiment is shown in black, worse kernel in grey.

	Lake Sihl				Lake Sils		
	Recall	**Precision**	**IoU**		**Recall**	**Precision**	**IoU**
Fr	0.87/0.87	0.73/0.85	0.66/0.76	Fr	0.93/0.89	0.97/0.99	0.90/0.88
N-Fr	0.92/0.97	0.97/0.97	0.90/0.94	N-Fr	0.98/0.99	0.94/0.91	0.92/0.90
Acc		0.91/0.95		**Acc**		0.95/0.94	
mIoU		0.78/0.85		**mIoU**		0.91/0.89	

	Lake Silvaplana		
	Recall	**Precision**	**IoU**
Fr	0.91/0.87	0.97/0.98	0.88/0.86
N-Fr	0.97/0.98	0.92/0.89	0.90/0.88
Acc		0.94/0.93	
mIoU		0.90/0.87	

Leave one winter out cross validation

To investigate the adaptability of a model to the potentially different conditions of an unseen winter, we trained the classifier using pixels from one of the two available winters (from all lakes), and tested on the data from the other winter. The results for MODIS and VIIRS are shown in Tables 3.11 and 3.12, respectively. Comparing these results with Table 3.7, it can be inferred that, across winters, the SVM does encounter a generalisation loss, especially with the RBF kernel. The loss with the linear kernel is minimal. Apparently, the RBF overfitted to the data characteristics of the specific winter and did not generalise as well as its linear counterpart. Note also, it is possible that freezing patterns could vary across winters even for the same lake, and learning-based systems might fail in case a pattern appears while testing that was not encountered during training. It is

encouraging that the linear kernel does not seem to overfit much, owing to its relatively lower capacity. Still, the results hint that the annotated data from more than one winter should be present in the training set when setting up an operational system.

Table 3.11.: MODIS *leave one winter out* results. The numbers are shown in A/B format where A and B represent the outcomes using RBF and linear kernels, respectively. The better kernel for a given experiment is shown in black, worse kernel in grey. Left is winter 16–17, right is winter 17–18.

	Recall	**Precision**	**IoU**		**Recall**	**Precision**	**IoU**
Fr	0.72/0.77	0.90/0.91	0.67/0.72	Fr	0.73/0.84	0.64/0.85	0.52/0.73
N-Fr	0.96/0.97	0.89/0.91	0.86/0.88	N-Fr	0.89/0.96	0.93/0.96	0.83/0.92
Acc		0.89/0.91		**Acc**		0.86/0.94	
mIoU		0.76/0.80		**mIoU**		0.68/0.83	

Table 3.12.: VIIRS *leave one winter out* results. The numbers are shown in A/B format where A and B represent the outcomes using RBF and linear kernels respectively. The better kernel for a given experiment is shown in black, worse kernel in grey. Left is winter 16–17, right is winter 17–18.

	Recall	**Precision**	**IoU**		**Recall**	**Precision**	**IoU**
Fr	0.83/0.79	0.92/0.99	0.77/0.78	Fr	0.87/0.90	0.77/0.79	0.69/0.72
N-Fr	0.96/0.99	0.91/0.89	0.88/0.89	N-Fr	0.93/0.94	0.97/0.97	0.90/0.91
Acc		0.91/0.92		**Acc**		0.92/0.93	
mIoU		0.82/0.84		**mIoU**		0.79/0.82	

Timeline plots and qualitative results

Fig. 3.13 shows the results of lake Sihl from a full winter (September 2016 till May 2017), listed in chronological order on the x-axis. For each cloud-free day (at least 30% of the lake pixels non-cloudy), the SVM result is shown on the y-axis (in the top and middle timelines) as the percentage of cloud-free clean pixels that are classified as non-frozen. In the bottom timeline, we display the MODIS snow product (100 means no snow and 0 means fully snow covered). The webcam-based ground truth is shown as a cyan colour line in all timeline plots, with four levels (100 for fully non-frozen, 75 for more snow or more ice days, 25 for more water and 0 for fully-frozen). For each sensor, the combined training data of all available lakes from two winters (except Sihl from 16–17) is used for these timeline plots. It can be seen from both MODIS and VIIRS timelines that thin ice vs.water confusion exists for both MODIS and VIIRS. This is because during the freeze-up period (late December), the model classifies a set of consecutive days as completely non-frozen, while the ground truth asserts more ice, probably thin ice floating on water.

In this paper, we compare our results of lake Sihl from winter 2016–2017 with the MODIS snow product (Hall and Riggs, 2016). It can be inferred from Fig. 3.13 that except for very few days, our MODIS results are in agreement with the MODIS snow

product. Although the newly added MOD10A1F (collection 6.1) seems to be a better option with the 'cloud gap filled' feature, we use the MOD10A1 product (Hall and Riggs, 2016, collection 6, daily cloud-free snow cover derived from the Terra MODIS) since the former product is not yet available for winter 2016–2017. Note that the MODIS product has a relatively coarse spatial resolution of 500 m as opposed to our results at 250 m resolution.

Fig. 3.14 displays exemplary qualitative results (lake Sihl, MODIS data, winter 16–17). Three non-transition dates (27 September 2016, 3 January 2017, and 28 March 2017) and a transition date (14 March 2017) are shown. The first and second rows portray the classification results and the confidence of the classifier (soft probability maps) respectively. In row 1, a clean pixel is shown as blue if the classifier estimates it as frozen, and red if non-frozen. In the second row, more blue/less red colour denotes a higher probability of being frozen. A pixel is not processed if it is cloudy. All except the fourth column show successfully classified days. In column 4, we present the results of an actually fully non-frozen day (27 September 2016) that was detected as almost fully frozen. Note the missing cloudy pixels in this image. This example shows that erroneous cloud masks (especially the false negatives) also induce errors in our predictions. Similar effects can be observed for the end of April (MODIS) and early October (VIIRS). Confusion due to undetected clouds is also the reason why a few days were estimated as non-frozen during mid-winter (see VIIRS timeline, February).

3.3.2. Experiments with webcam images

The neural network is implemented in *Tensorflow* (`https://www.tensorflow.org/`). We extracted square patches (crop size, see Table 3.13) from the images and train the network by minimising the weighted cross-entropy loss, giving more attention to the under-represented classes to compensate imbalances in the training data. Testing is performed at full image resolution without cropping. More details about the hyperparameters used are shown in Table 3.13.

Table 3.13.: Hyperparameters for the *Deep-U-Lab* model.

Name	**Lake Detection**	**Lake Ice Segmentation**
Crop size	500, 500	321, 321
Optimiser	stochastic gradient descent	stochastic gradient descent
Atrous rates (dilation)	6, 12, 18	6, 12, 18
Output stride	16	16 (training), 8 (testing)
Base learning rate	1×10^{-5}	1×10^{-5}
Batch size	4	8
Epochs	100	100

For the task of lake detection, we collected image streams from four different lakes, see Table 3.14. The cameras near lakes Sihl and St. Moritz are rotating while the others are stationary. Performance of the network for lake detection is assessed only on summer images in order to sidestep the complications in winter due to the presence of snow in

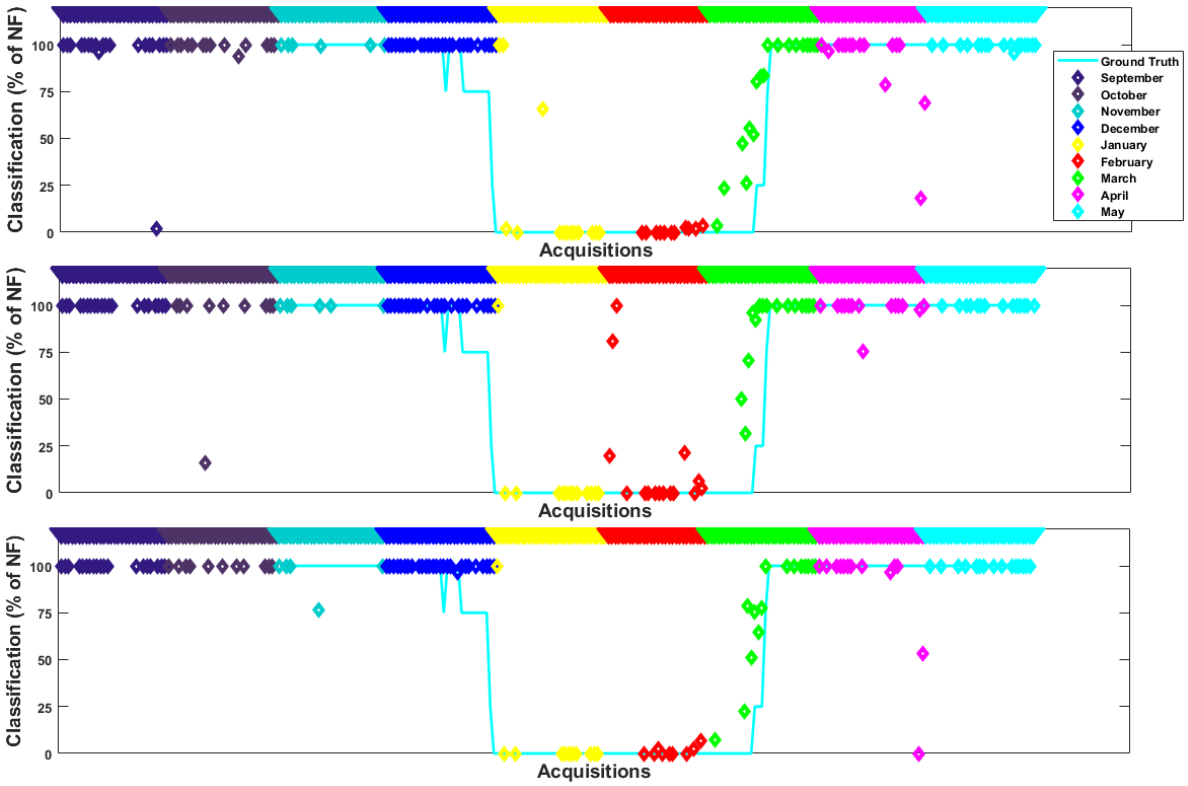
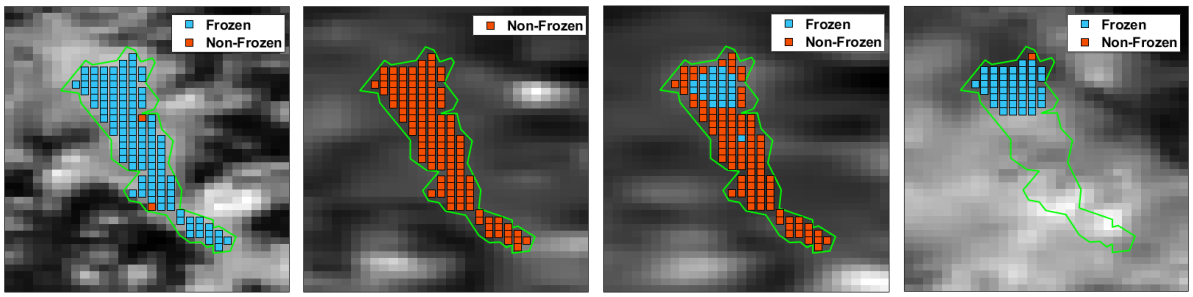


Figure 3.13.: MODIS (**top**) and VIIRS (**middle**) timeline results of lake Sihl for full winter 16–17 using linear kernel. A timeline of the MODIS snow and ice product (**bottom**) is also plotted for comparison with our results and the webcam-based ground truth. In all timelines, the x-axis shows all dates that are at least 30% cloud-free in chronological order and the respective results [% of Non-Frozen (NF) pixels] are plotted on the y-axis.

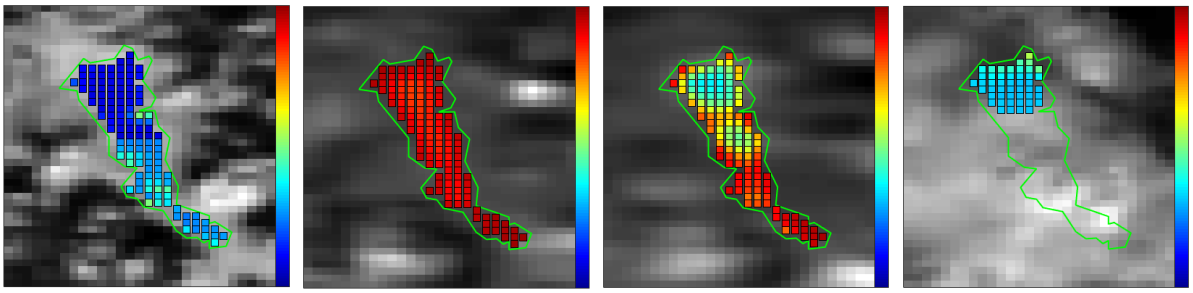
and around the lake. A score of $\geq$0.9 mIoU is achieved, see Table 3.14. However, the IoU for the lake (class foreground, FG) is somewhat lower because of the severe class imbalance. Note that IoU is a rather strict measure, e.g., detection with 80%, recall at 80%, and precision results in an IoU of 60%. Qualitative results are displayed in Fig. 3.15. The first three rows show successful cases while the last row displays a case with some misclassification. Note that on such a low visibility day, even humans find it difficult to spot the transition from lake to sky. Whereas our network detected the lake even in the presence of challenging sun reflections (row 2) and when the foreground lake area is very small (row 3).

To assess lake ice segmentation, we experimented exhaustively for the two lakes (St. Moritz, Sihl) and two winters (16–17, 17–18) annotated in the Photi-LakeIce dataset. The evaluation includes experiments for segmentation within the same camera, across cameras, across winters, and across lakes.

For same camera experiments, we employed a 75%/25% train-test split. Corresponding quantitative results are presented in Table 3.15. Note that, in all comparable experiments, we surpass previous state-of-the-art (Tom et al., 2019) by a significant margin. They produced results only on two cameras monitoring St. Moritz. We demonstrate



3 January 2017 | 28 March 2017 | 14 March 2017 | 27 September 2016



Frozen | Non-Frozen | Transition | Non-Frozen

Figure 3.14.: MODIS qualitative results using the linear kernel. Top and bottom rows show classification results and corresponding confidence respectively. Results of cloudy pixels are not displayed. First, second, and third columns show success cases while the fourth column displays a failure case. In the second row, more red means more non-frozen and more blue means more frozen. The dates and ground truth labels are shown below each sub-figure in the first and second rows respectively.

Table 3.14.: Lake detection results (mIoU). The four cameras that monitor lake Silvaplana are indicated as S_0, S_1, S_2, and S_3. BG and FG denote background and foreground (lake area) respectively. #Im represents the number of images.

Training Set		**Test Set**		**BG**	**FG**	**mIoU**
Lakes	**#Im**	**Lake**	**#Im**	**(IoU)**	**(IoU)**	
S_0, S_1, S_2, S_3, Sils, St. Moritz	9104	Sihl	448	0.95	0.60	0.93
S_0, S_1, S_2, S_3, St. Moritz, Sihl	7477	Sils	2075	0.95	0.60	0.93
S_1, S_2, S_3, Sils, St. Moritz, Sihl	8041	S_0	1511	0.96	0.59	0.94
S_0, S_2, S_3, Sils, St. Moritz, Sihl	8676	S_1	876	0.92	0.58	0.90
S_0, S_1, S_3, Sils, St. Moritz, Sihl	7906	S_2	1646	0.98	0.44	0.95
S_0, S_1, S_2, Sils, St. Moritz, Sihl	7652	S_3	1900	0.98	0.55	0.95
S_0, S_1, S_2, S_3, Sils, Sihl	8456	St. Moritz	1096	0.93	0.80	0.92

our system also on a new lake (Sihl, camera 2) with images that are significantly harder to classify (see Fig. 3.5 and Table 3.6) due to poor spatial resolution, image compression artefacts, frequent unfavourable lighting, etc. Additionally, the foreground to background pixel ratios in Sihl images are very low, which poses an additional challenge, and magnifies the influence of very small misclassified areas on the quantitative error metrics. As a result, the performance on lake Sihl is not as good as for St. Moritz. Nevertheless,

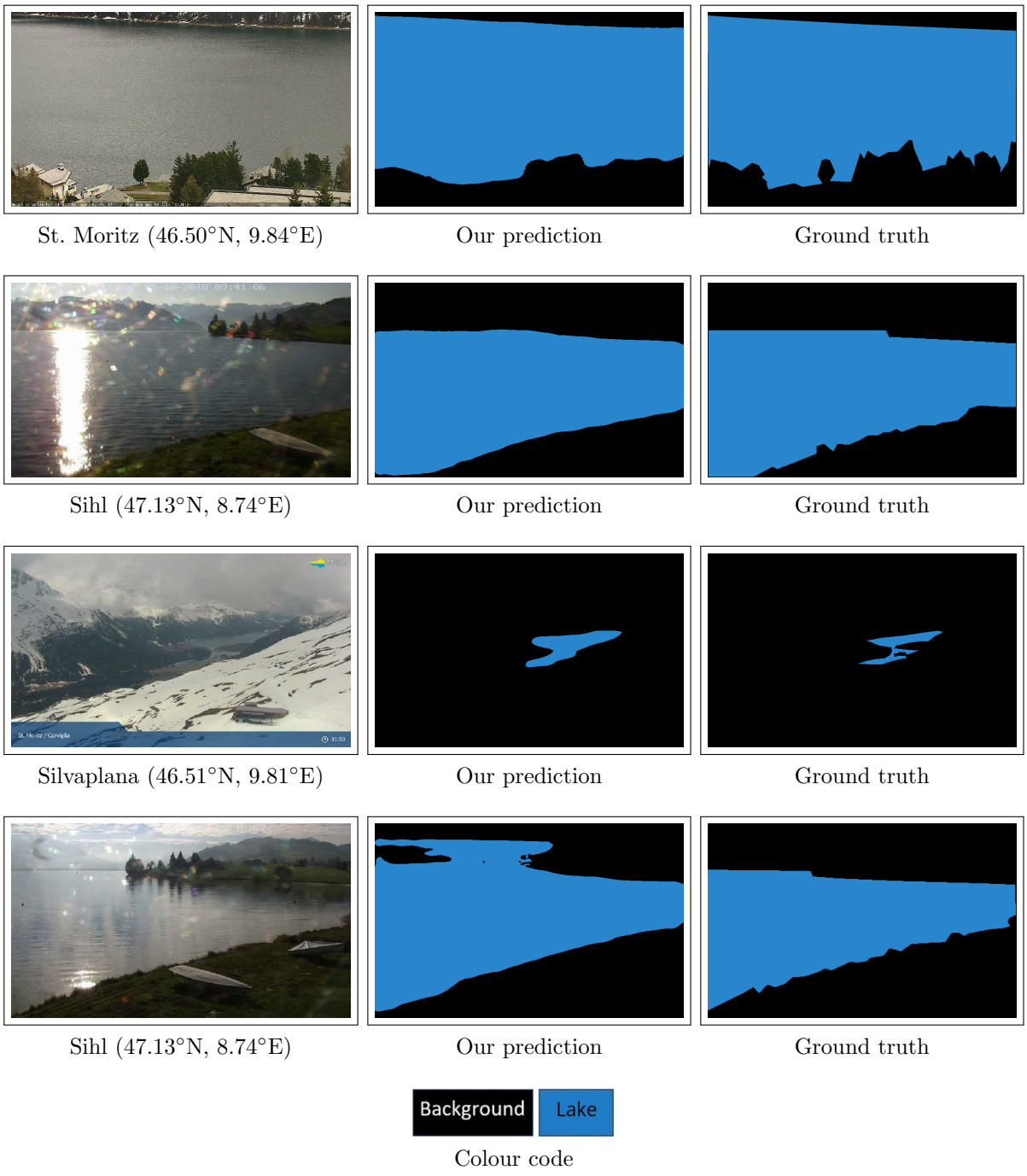


Figure 3.15.: Lake detection results. Both success (rows 1,2,3) and failure (row 4) cases are shown. The colour code used to visualise the results is also displayed. The first column shows the lakes being monitored, along with the approximate location (latitude, longitude) of the webcam.

the predictions have a mean IoU $> 74\%$. The images with a mix of classes, like water with some ice or partially snow-covered ice are the most difficult ones to classify in part due to the fact that the ice class is especially rare and therefore under-represented in the training data, as snow that falls on the ice does not melt away for a long time.

All results shown so far are for networks trained with data augmentation. To quantify

Table 3.15.: Results (IoU) of *same camera train/test* experiments. We compare our results with *Tiramisu Network* (Tom et al., 2019, shown in grey). Cameras 0 and 1 monitor lake St. Moritz while camera 2 captures lake Sihl.

Training Set		Test Set		Water	Ice	Snow	Clutter	mIoU
Camera	Winter	Camera	Winter					
Camera 0	16–17	Camera 0	16–17	0.98/0.70	0.95/0.87	0.95/0.89	0.97/0.63	0.96/0.77
Camera 0	17–18	Camera 0	17–18	0.97	0.88	0.96	0.87	0.93
Camera 1	16–17	Camera 1	16–17	0.99/0.90	0.96/0.92	0.95/0.94	0.79/0.62	0.92/0.85
Camera 1	17–18	Camera 1	17–18	0.93	0.84	0.92	0.84	0.89
Camera 2	16–17	Camera 2	16–17	0.79	0.62	0.81	—	0.74
Camera 2	17–18	Camera 2	17–18	0.81	0.69	0.86	—	0.79

the influence of this common practice, we also report results without augmentation for camera 0, which are 2 percent points lower, see Table 3.16. Additionally, in order to study

Table 3.16.: Effect of data augmentation (IoU values) on the *same camera train/test* experiment (camera 0).

Experiment	Water	Ice	Snow	Clutter	mIoU
Without augmentation	0.97	0.93	0.91	0.96	0.94
With augmentation	0.98	0.95	0.95	0.97	0.96

how quickly the network learns, the mIoU is plotted on the training set against the number of training iterations. For that study, we use the example of lake St. Moritz (camera 0) from winter 16–17. Results are shown in Fig. 3.16. The (smoothed) learning curve is very steep initially (<10 k steps) but does not completely saturate, which indicates that more training data could probably improve the results further.

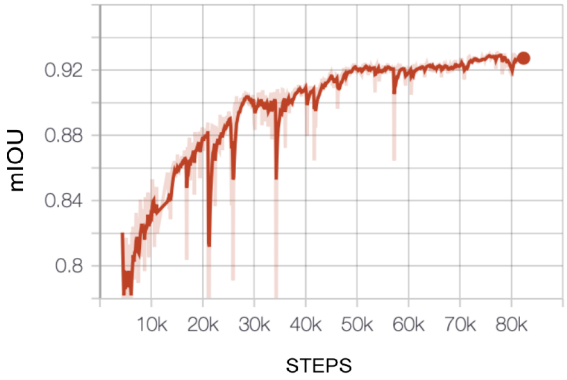


Figure 3.16.: Evolution of mean IoU (mIoU) against the number of training steps (camera 0, St. Moritz, winter 2016–2017). Dark red curve represents a smoothed version of the original (light red) curve.

The generalisation performance (across cameras and winters) of the best webcam model reported in previous work (Tom et al., 2019) is still unsatisfactory, especially for

the cross-camera case. As can be seen from our cross-camera results (within St. Moritz cameras, refer Table 3.17), the *Deep-U-Lab* model trained using data from one camera works well on a different camera, meaning that our method generalises well across cameras with totally different viewpoints, image scales, and lighting conditions. Note that, we indeed improve over prior state-of-the-art (Tom et al., 2019) significantly (gain of 35–40 percent points), which implies that *Deep-U-Lab* has the capacity to learn generalisable class appearance, without overfitting to a specific camera geometry or viewpoint. Our results for winter 17–18 are not as good as 16–17, primarily due to complicated lighting and ice patterns (e.g., black ice) which appeared only in that winter. In addition, the scores for the ice and clutter classes are low, primarily due to lower sample numbers. A comparison to prior work is not possible for winter 17–18, as that season has not been processed before.

Table 3.17.: Results (IoU) of *cross-camera* experiments. We compare our results with the *Tiramisu Network* (Tom et al., 2019, shown in grey). Both cameras 0 and 1 monitor lake St. Moritz.

Training Set		**Test Set**		**Water**	**Ice**	**Snow**	**Clutter**	**mIoU**
Camera	**Winter**	**Camera**	**Winter**					
Camera 0	16–17	Camera 1	16–17	0.76/0.36	0.75/0.57	0.84/0.37	0.61/0.27	0.74/0.39
Camera 0	17–18	Camera 1	17–18	0.62	0.66	0.89	0.42	0.64
Camera 1	16–17	Camera 0	16–17	0.94/0.32	0.75/0.41	0.92/0.33	0.48/0.43	0.77/0.37
Camera 1	17–18	Camera 0	17–18	0.59	0.67	0.91	0.51	0.67

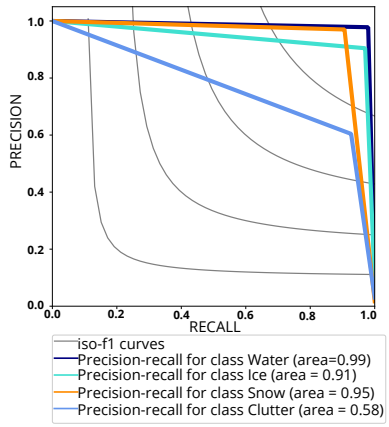
Deep-U-Lab performs superior to prior state-of-the-art in *cross-winter* experiments, too (Table 3.18), outperforming Tom et al. (2019) by about 14–20 percent points. However, it does not generalise across winters as well on lake Sihl.

Table 3.18.: Results (IoU) of *cross-winter* experiments. We compare our results with the *Tiramisu Network* (Tom et al., 2019, shown in grey). Cameras 0 and 1 monitor lake St. Moritz while camera 2 captures lake Sihl.

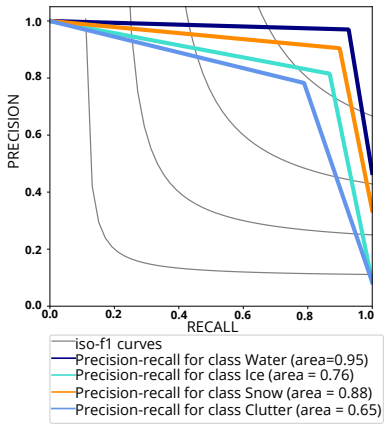
Training Set		**Test Set**		**Water**	**Ice**	**Snow**	**Clutter**	**mIoU**
Camera	**Winter**	**Camera**	**Winter**					
Camera 0	16–17	Camera 0	17–18	0.64/0.45	0.58/0.44	0.87/0.83	0.59/0.40	0.67/0.53
Camera 0	17–18	Camera 0	16–17	0.98	0.91	0.94	0.58	0.87
Camera 1	16–17	Camera 1	17–18	0.86/0.80	0.71/0.58	0.93/0.92	0.57/0.33	0.77/0.57
Camera 1	17–18	Camera 1	16–17	0.93	0.76	0.86	0.65	0.80
Camera 2	16–17	Camera 2	17–18	0.61	0.14	0.35	—	0.51
Camera 2	17–18	Camera 2	16–17	0.41	0.18	0.45	—	0.50

For a more complete picture of the cross-winter generalisation experiment, we also plot precision-recall curves (refer Fig. 3.17). Similar curves for same camera and cross-camera experiments can be found in Prabha et al. (2020). As expected, segmentation of the two under-represented classes (clutter and ice) is less correct. Additionally, for the class clutter, a considerable amount of the deviations from ground truth occur due to improper annotations rather than erroneous predictions. As drawing pixel-accurate ground truth boundaries around narrow man-made items placed on frozen lakes such as

tents, poles, etc. is time-consuming and tedious, the clutter objects are often annotated with rough summary masks that include considerable snow/ice background. This greatly exaggerates the (relative) number of clutter pixels in the annotations, thus increasing the relative error. According to Fig. 3.17, operating points around 85% recall are a good



Tested on camera 0 (16–17), trained on camera 0 (17–18).



Tested on camera 1 (16–17), trained on camera 1 (17–18).

Figure 3.17.: Precision-recall plots (St. Moritz) of *cross-winter* experiments. Best if viewed on screen.

trade-off for *cross-winter* segmentation, if not every single pixel must be labelled. A more extreme test is generalisation across lakes, with different spatial resolution, image quality, reflection, and lighting patterns, shadows, etc. To our knowledge our work is the first one to try this. See Table 3.19 for results. Before training the models, we remove the clutter pixels from camera 0, since camera 2 does not have any clutter that could serve as training data. Classifying images from a lake with different characteristics and acquired with a different type of camera proves challenging. In one case, the results are acceptable for the more frequent classes despite a noticeable drop, as for the case camera 2→camera 0. In the other case camera 0→camera 2 the attempt largely fails. The images of lake Sihl (camera 2) are of clearly lower quality and more difficult to classify, challenging even human annotators. Consequently, training on St. Moritz does not equip the classifier to deal with them.

Table 3.19.: Results (IoU) of *cross-lake* experiments. Cameras 0 and 2 monitor lakes St. Moritz and Sihl respectively.

Training Set	**Test Set**	**Water**	**Ice**	**Snow**	**mIoU**
Camera 0 (16–17)	Camera 2 (16–17)	0.40	0.23	0.42	0.35
Camera 2 (16–17)	Camera 0 (16–17)	0.85	0.25	0.68	0.60

In a further experiment, we divide the *Photi-LakeIce* dataset into six folds, see Table 3.20. This makes it possible to perform experiments with a larger amount of training data, given that in previous experiments the loss had not fully saturated. As expected

from a high-capacity statistical model, more training data improves the results i.e., it seems feasible to build a practical system if one is willing to undertake a bigger (but still reasonable and realistic) annotation effort. An exception in this experiment is lake Sihl (camera 2), where the performance drops. This confirms the observation above that this camera is the most difficult one to segment in our dataset, and the domain gap from St. Moritz to Sihl is too large to bridge without appropriate adaptation measures. One solution might be fine-tuning with at least a small set of cleverly picked samples from the target lake, but this is beyond the scope of the present paper.

Table 3.20.: Results (IoU) of *leave one dataset out* experiments. Cameras 0 and 1 monitor lake St. Moritz while camera 2 captures lake Sihl.

Training Set	Test Set	Water	Ice	Snow	Clutter	mIoU
Camera 0 (17–18), Cameras 1 and 2 (2 winters)	Camera 0 (16–17)	0.98	0.90	0.96	0.62	0.86
Camera 0 (16–17), Cameras 1 and 2 (2 winters)	Camera 0 (17–18)	0.83	0.78	0.95	0.59	0.78
Camera 1 (17–18), Cameras 0 and 2 (2 winters)	Camera 1 (16–17)	0.99	0.92	0.91	0.69	0.87
Camera 1 (16–17), Cameras 0 and 2 (2 winters)	Camera 1 (17–18)	0.92	0.81	0.96	0.55	0.81
Camera 2 (17–18), Cameras 0 and 1 (2 winters)	Camera 2 (16–17)	0.35	0.25	0.46	—	0.35
Camera 2 (16–17), Cameras 0 and 1 (2 winters)	Camera 2 (17–18)	0.66	0.30	0.36	—	0.44

To assess the lake ice segmentation visually, we depict qualitative webcam results for cameras 0 and 1 in Fig. 3.18. *Deep-U-Lab* successfully segments correctly in challenging scenarios. For instance, our network performed well even when shadows appeared on the lake either from clouds or nearby mountains (Fig. 3.18 row 1). To determine how well the *Deep-U-Lab* predictions follow the ground truth, especially during freezing and thawing periods, we plot time series results that include all the transition as well as non-transition days from a full winter (17–18, see Fig. 3.19). Per image, we sum the numbers of ice and snow pixels and divide by the sum of all lake pixels. The resulting fractions of frozen pixels are aggregated into a daily value by taking the median. Optionally, the daily values are further processed with another 3-day median to improve temporal coherence. The daily fractions of frozen pixels (y-axis) are displayed in chronological order (x-axis), for the ground truth, daily prediction, and smoothed prediction. Smoothing across time improves the final results by ≈3%.

3.3.3. Ice-on/off results

We go on to estimate the ice-on and ice-off dates using our satellite- and webcam-based approaches, results are shown in Table 3.21. A comparison with the ground truth dates (estimated by visual interpretation of webcams by a human operator) is also provided. Additionally, we compare our results with the in-situ temperature analysis results reported in Tom et al. (2019). We can only show the results for one winter (16–17), since ground truth is not available for 17–18.

Prior to the estimation of the two dates, we combine the time series results of both MODIS and VIIRS (Fig. 3.13) in order to minimise data gaps due to clouds, by simply filling in missing days in the VIIRS time series with MODIS results, whenever the latter are available. Even after merging the two time series, some gaps still exist during the critical transition periods. Note that the presence of gaps near the ice-on/off dates could

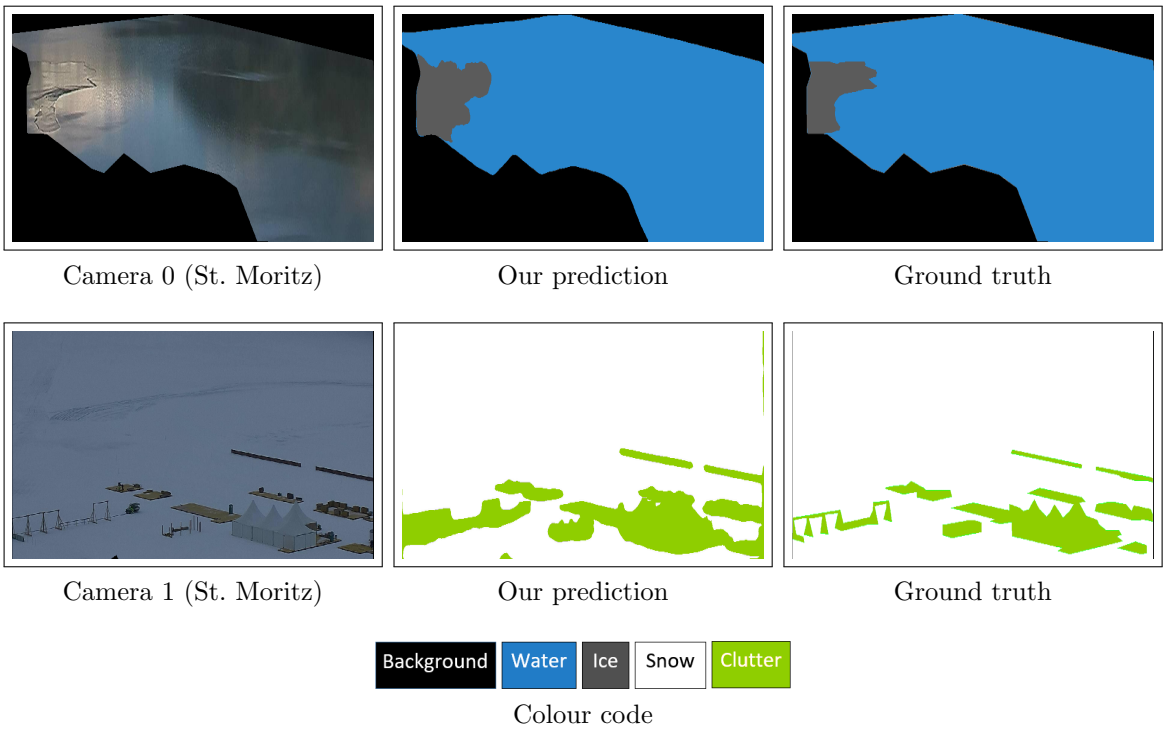


Figure 3.18.: Qualitative lake ice segmentation results on webcam images. The colour code is also shown.

affect the accuracy and confidence in the estimated dates. This is one of the risks when using optical satellite image analysis and where webcams could constitute a valuable alternative, if sufficient coverage can be ensured for a lake of interest.

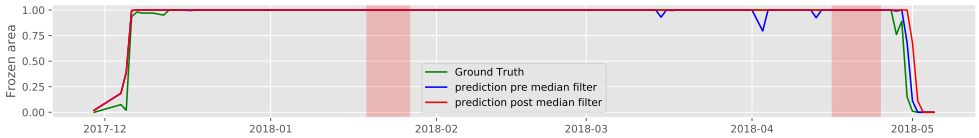


Figure 3.19.: *Cross-camera* time series results (winter 17–18) of lake St. Moritz using *Deep-U-Lab*. Results for camera 1 (when trained on camera 0 data) is displayed. All dates are shown in chronological order on the x-axis and the respective results (percentage of frozen pixels) are plotted on the y-axis. Data lost due to technical failures are shown as red bars.

Table 3.21.: Ice-on/off dates (winter 16–17). Ground truth dates are shown in the order of confidence in case of more than one candidate.

Dates	Ground Truth	Satellite Approach	Webcam Approach	In-Situ [T] (Tom et al., 2019)
ice-on (Sihl)	1 January 2017	3 January 2017	4 January 2017	28–29 December 2016
ice-off (Sihl)	14 March 2017, 15 March 2017	10 March 2017	14 February 2017	16 March 2017
ice-on (Sils)	2 January 2017, 5 January 2017	6 January 2017	-	31 December 2016
ice-off (Sils)	8 April 2017, 11 April 2017	31 March 2017	-	10 April 2017
ice-on (Silvaplana)	12 January 2017	15 January 2017	-	14 January 2017
ice-off (Silvaplana)	11 April 2017	30 March 2017	-	14 April 2017
ice-on (St. Moritz)	15–17 December 2016	1 January 2017	14 December 2016	17 December 2016
ice-off (St. Moritz)	30 March–6 April 2017	7 April 2017	18 March–26 April 2017	5–8 April 2017

We estimate the ice-on/off dates from the combined time series and show them as "satellite approach" in Table 3.21. The best results are obtained with an RBF kernel for St. Moritz and with a linear kernel for rest of the lakes. In most cases the ice-on/off dates have an offset of 1–4 days from the ground truth. Exceptions are the ice-off dates of Sils and Silvaplana. Note that data from only one winter (of the lake being tested) is present in the corresponding training set. It appears that training data from more winters is critical to estimate accurate ice-on/off dates. We note that there could also be noise in the ground truth ice-on/off dates due to human interpretation errors. Using the satellite approach, significant errors in ice-on/off estimation exist for St. Moritz. Recall that the daily decision for lake St. Moritz is taken based on just four pixels and based only on cloud-free days in MODIS. This clearly points to the fact that MODIS (and even more, VIIRS) imagery is not the best choice for very small lakes. The results obtained with webcams show a higher accuracy for lake St. Moritz (see Table 3.21 and Fig. 3.19). Here, we use camera 0 to estimate the two phenological dates, since it has a better coverage of the lake than camera 1. Note that no data is available between 18.03.2017 and 26.04.2017 due to a technical problem with the camera, and ice-off unfortunately occurred during that period. While we obtain excellent results for St. Moritz with webcams, the accuracy of ice-on/off for lake Sihl is not good, primarily because of the limited image quality with low spatial resolution (see Table 3.6), compression artefacts, and acute view angles (see Fig. 3.5).

3.4. Discussion

3.4.1. VIIRS and MODIS analysis

The optical satellite sensors such as MODIS and VIIRS can clearly serve as a basis for routine monitoring of lake ice (especially for global coverage) and the results achieved show a high level of accuracy. One weakness is their inability to penetrate clouds, especially during lake freeze-up and break-up. The main advantage of MODIS is the availability of longer time series data. In addition, MODIS has useful bands in various areas of the electromagnetic spectrum. However, there are several disadvantages too. The radiometric quality is not that good, moreover, the sensor is very old and its absolute geolocation is less accurate than that of VIIRS (more important for small lakes). Furthermore, MODIS data is expected to eventually be discontinued, whereas VIIRS operation is guaranteed over a longer future period (JPSS-1/NOAA-20 until 2024; JPSS-2 with same suite of sensors will be launched in 2021 with designed life time of 7 years; JPSS-3 and -4 are in the planning phase).

Poor spatial resolution (particularly for VIIRS), makes it impossible to operate our satellite methodology on small-size lakes up to at least $2\,\mathrm{km}^2$. Another issue is the significant confusion between (thin) ice and water since the similar reflectance of these two classes can confuse the classifier. Unfortunately there exist very few non-transition dates with no clouds, snow-free ice, mixture of thin ice, and water in both the MODIS and VIIRS datasets, such that training a reliable model for these situations still remains a challenge. Moreover, the presence of label noise in the ground truth impacts the training. Such noise occurs mainly because most of the webcams are not optimally configured and

it is very difficult to capture the whole lake in a single webcam frame. This problem is even bigger for larger lakes. Integrating the visual interpretations from multiple cameras observing a lake is cumbersome as well as challenging. One possible solution for the noise problem could be to not train on dates near the transition period, for which label noise in the ground truth is more likely. It is, however, equally possible that this would even aggravate the problem, as the conditions seen in the training set would become even less representative of the transition periods. Large-scale in-situ measurements are an alternative to webcam-based ground truth, but are not realistic for wide-area coverage. Furthermore, imperfections of the cloud-masks bring in more errors in the high-accuracy range where our method operates.

Regarding our SVM-based methodology, the RBF kernel tends to not generalise as well as its linear counterpart. However, this behaviour may depend on the available training data. Under our current experimental conditions, the linear kernel overall has the upper hand, but this assessment could still change when using data from multiple winters.

3.4.2. Webcam analysis

For webcam data featuring sufficient spatial resolution, we see a great potential for lake ice monitoring. We do note that webcam placement is restricted by practical considerations. Selecting and/or mounting webcams for lake ice monitoring will normally be a compromise between the ideal geometric configuration and finding a place where the device can be installed with reasonable effort. For the ideal placement, the usual perspective imaging rules apply, most importantly viewing directions from above are preferable over grazing angles, and viewing directions directly towards the sun should be avoided as much as possible.

One question that still remains unanswered is: What is the reason that results in *Deep-U-Lab* outperforming the *Tiramisu* lake ice network (Jégou et al., 2016, Xiao et al., 2018, Tom et al., 2019)? One possible explanation is that our model profits from the smarter dilated convolutions and multi-scale pyramid pooling at the feature extraction stage, effectively letting the network grasp a relatively broader context as opposed to the *Tiramisu* network. Additionally, our model heavily benefits from the pre-trained weights to learn with still limited training data for the lake ice task. Our *Deep-U-Lab* model did not converge when we tried to train it without transfer learning whereas pre-trained weights for FC-DenseNet are not available, so that we can not at the moment quantify the influence of pre-training.

Regarding the computational efficiency of the CNN approach, (off-line) training for 100k steps on camera 0 (820 images) takes ≈24 h on a PC equipped with a NVIDIA GeForce GTX 1080 Ti graphics card (for cross-camera experiment, lake St. Moritz, winter 16–17). Testing takes ≈10 minutes for the 1180 images of camera 1.

3.5. Conclusion

We investigated the potential of machine-learning based image analysis, in combination with various image sensors to retrieve lake ice. So far such an approach has rarely been explored, especially with regard to the many small lakes on Earth (particularly in mountainous regions), but it can be a valuable source of information that is largely independent of in-situ observations as well as models of the freezing/thawing process. We put forward an easy-to-use, SVM-based approach to detect lake ice in MODIS and VIIRS satellite images and show that it delivers conclusive results. Additionally, we set a new state-of-the-art from webcam-based lake ice monitoring, using the *Deep-U-Lab* network, and have in that context also automated the detection of lake outlines as a further step towards operational monitoring with webcams. Finally, we introduced a new, public webcam dataset with pixel-accurate annotations.

To detect lake ice from MODIS and VIIRS optical satellite imagery, we proposed a simple, generic machine learning-based approach that achieves high accuracy for all tested lakes. Though we focused on Swiss Alpine lakes, the proposed approach is very straight-forward and hence the results could hopefully be directly applied to other lakes with similar conditions, in Switzerland and abroad, and possibly to other sensors with similar characteristics. We demonstrated that our approach generalises well across winters and lakes (with similar geographical and meteorological conditions). In addition to the lake ice detection from space, we have proposed the use of free data from terrestrial webcams for lake ice monitoring. Webcams are especially suitable for small lakes (ca. up to $2\,\mathrm{km}^2$), which cannot be monitored by VIIRS-type sensors. Despite the limited image quality, we obtained promising results using deep learning. Webcams have good ice detection capability with a much higher spatial resolution compared to satellites. However, one has no control over the location, orientation, lake area coverage, and image quality (often poor) of public webcams. In addition, there are no, or too few, webcams for some lakes. On the positive side, webcams are largely unaffected by cloud cover. For continental/global coverage, satellite-based monitoring is clearly the method of choice, again confirming the advantage of satellite observations for large-scale Earth observation. For focused, local campaigns and as a source of reference data at selected sites, the webcam-based monitoring method may be an interesting alternative. Note also, it may in certain cases be warranted to install dedicated webcams (respectively, surveillance cameras) with pan, tilt and zoom functionalities optimised for lake ice monitoring.

One way to circumvent the problem of clouds with optical satellite sensors is to use microwave data. In particular, Sentinel-1 SAR data (GSD 10 m, freely available) looks very promising (Tom et al., 2020a). Optical sensors such as Sentinel-2 and Landsat-8 are visually easier to interpret w.r.t. lake ice than Radar, and have a better spatial resolution than MODIS and VIIRS. Although they are not suitable as a stand-alone source for lake ice monitoring due to their low temporal resolution (under ideal conditions five days), they may still in certain cases be useful to fill gaps in VIIRS/MODIS results.

We consider our satellite-based approach as a first step and ultimately hope to produce a 20-year time series, using MODIS data since 2000. It will be interesting to correlate the longer-term lake freezing trends with other climate time series such as surface temperature or CO_2 levels. We expect that such a time series will be helpful to draw conclusions

about the local and global climate change.

One technical finding of our study is that the prior learning-based approaches (Xiao et al., 2018, Tom et al., 2019) did not fully leverage the power of deep CNNs to observe lake ice. At the methodological level, we clearly demonstrated the potential of machine (deep) learning systems for lake ice monitoring, and hope that this research direction will be pursued further. Given the good cross-winter and cross-camera generalisation of the models and computational efficiency at inference time (on GPU for the CNN part), an operational deployment is within reach. Our results show that employing the state-of-the-art CNN frameworks was highly effective for ice analysis, especially during the transition periods. What still needs improvement is cross-lake generalisation. We do expect that a *Deep-U-Lab* model trained using data from a couple of winters could consistently reach >80% IoU on the four major classes. From the segmentation results we were in many cases able to determine the ice-on and ice-off dates to within 1–2 days and for that task the relatively better quality webcams were particularly helpful, as satellite-based segmentation was less reliable during the transition periods. An interesting direction may be to reduce the one-time effort for ground truth labelling with techniques such as domain adaptation or active learning.

For monitoring small lakes, integrating the webcam results with in-situ temperature measurements seems to be a possible future direction. Additionally, for such lakes, usage of UAVs equipped with both thermal and RGB cameras could be a plausible option, but may be difficult to operationalise due to the need for accurate geo-referencing, lack of robustness in cold weather, as well as legal flight restrictions. An intriguing extension of the webcam-based approach could be to use crowd-sourced imagery for lake ice detection. We published some preliminary results in one of our recent works (Prabha et al., 2020). A large, and exponentially growing, number of images are available on the internet and social media. With the advance of smartphones equipped with cameras and the habit of selfies, many personal images show a lake in background. Still regular coverage of a given site is hard to ensure, and accurate geo-referencing of such images is challenging.

3.6. Acknowledgements

This research was funded mainly by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss (in the framework of GCOS Switzerland) and partially by the Sofja Kovalevskaja Award of the Humboldt Foundation. The APC was funded by ETH Zurich. We express our gratitude to all the partners in our MeteoSwiss projects for their support. Additionally, we acknowledge Mathias Rothermel, Muyan Xiao, and Konstantinos Fokeas for their contributions in collecting and annotating the images of Photi-LakeIce dataset. We also thank Hotel Schweizerhof for providing the webcam data of lake St. Moritz.

4 Recent Ice Trends in Swiss Mountain Lakes: 20-year Analysis of MODIS Imagery

Manu Tom, Tianyu Wu, Emmanuel Baltsavias, Konrad Schindler
Under review

Abstract

Depleting lake ice can serve as an indicator for climate change, just like sea level rise or glacial retreat. Several Lake Ice Phenological (LIP) events serve as sentinels to understand the regional and global climate change. Hence, it is useful to monitor long-term freezing and thawing patterns of lakes. In this paper we report a case study for the Oberengadin region of Switzerland, where there are several small- and medium-sized mountain lakes. We observe the LIP events, such as freeze-up, break-up and ice cover duration, across two decades (2000-2020) from optical satellite images. We analyse time series of MODIS imagery by estimating spatially resolved maps of lake ice for these Alpine lakes with supervised machine learning (and additionally cross-check with VIIRS data when available). To train the classifier we rely on reference data annotated manually based on publicly available webcam images. From the ice maps we derive long-term LIP trends. Since the webcam data is only available for two winters, we also validate our results against the operational MODIS and VIIRS snow products. We find a change in Complete Freeze Duration (CFD) of -0.76 and -0.89 days per annum (d/a) for lakes Sils and Silvaplana, respectively. Furthermore, we observe plausible correlations of the lake freezing and thawing trends with climate data such as air temperature, sunshine and precipitation measured at nearby meteorological stations. We notice that mean winter air temperature has negative correlation with the freeze duration events and break-up events, and positive correlation with the freeze-up events. Additionally, we observe strong negative correlation of sunshine during the winter months with the freeze duration and break-up events. Furthermore, we note that increased precipitation in the months January to May (likely to be predominantly snow), favours later break-up.

4.1. Introduction

Scientists around the globe strive to understand the changing climate, to find ways to mitigate the impact of associated extreme weather conditions, and to protect the environment for future generations (Rolnick et al., 2019). The repercussions of climate change are foreseen to amplify in the next few decades. Furthermore, the latest climate models underline the need for urgent mitigation (Forster et al., 2020). *"Human activities are estimated to have caused approximately* $1.0°C$ *of global warming above pre-industrial levels, with a likely range of* $0.8°C$ *to* $1.2°C$*. Global warming is likely to reach* $1.5°C$ *between 2030 and 2052 if it continues to increase at the current rate"*, said the IPCC special report on impacts of global warming (Masson-Delmotte et al., 2018).

Many studies have reported on the response of LIP trends to climate variations (Duguay et al., 2006, Howell et al., 2009, Brown and Duguay, 2010, Kang et al., 2012, Surdu et al., 2014, Sharma et al., 2019). Local weather patterns and lake ice formation processes are inter-connected. Hence, monitoring the long-term LIP trends can provide integral cues on the regional and global climate. Increasing temperatures cause decreasing trends in the lake ice formation process. Air temperature in the vicinity of a lake affects the ice formation process within the lake and vice versa. Moreover, there are potential positive feedbacks, as frozen lakes have higher albedo (especially when covered with snow), and thus lower absorption and evaporation (Wang et al., 2018c, Slater et al., 2021). In addition to its usefulness for climate studies, lake ice monitoring is also crucial to organise safe transportation especially in lakes that freeze only partially, to conserve freshwater ecology, to trigger warnings against ice shoves caused by wind during the break-up period, and for winter tourism.

In the present case study, we aim to monitor lakes of the Oberengadin region in the Swiss Alps (which reliably freeze every winter) on a daily basis during the winter months, to derive the spatio-temporal extent of lake ice.[1] Specifically, we estimate the four important LIP events: *Freeze-Up Start (FUS), Freeze-Up End (FUE), Break-Up Start (BUS) and Break-Up End (BUE).* Using these four dates, we also estimate the *Complete Freeze Duration (CFD)* and *Ice Coverage Duration (ICD)*, refer to Table 4.1 for definitions. Some publications have termed FUE and BUS as ice-on and ice-off dates, respectively (Hendricks Franssen and Scherrer, 2008, Tom et al., 2020c). However, other researchers (and the NSIDC database, `https://nsidc.org/`) consider BUE as ice-off (Duguay et al., 2015). Regarding the ice-on/off dates, the Global Climate Observing System (GCOS) requirements are daily observations at an accuracy of ± 2 days (`https://gcos.wmo.int/en/essential-climate-variables/lakes/ecv-requirements`).

In this work, we focus on estimating the spatio-temporal extent of the ice cover from optical satellite data.[2] Compared to other sensors, MODIS and VIIRS satellite data have several advantages such as wide area coverage, good spectral and fine temporal resolution (daily), free availability etc. Additionally, compared to other optical satellites such as Landsat-8, Sentinel-2 and the like, MODIS and VIIRS offer the best spatio-temporal resolution trade-off for the application of single-sensor lake ice monitoring, even

[1] Note that we do not include the ice thickness.

[2] we have previously also used webcams (Xiao et al., 2018, Tom et al., 2020c) and Sentinel-1 Synthetic Aperture Radar (Tom et al., 2020a, SAR) for lake ice monitoring.

though the spatial resolution is moderate (250-1000m Ground Sampling Distance, GSD). In addition, the global coverage is beneficial to eventually scale up to country- or world-wide monitoring. On the other hand, cloud cover is a bottleneck for optical satellite data analysis. An important asset is the availability of large time series, e.g., MODIS data is available for the entire period since 2000, contrary to other sensor data like airborne or terrestrial photography, webcams etc. This makes it possible to implement a 20-year analysis and to derive LIP trends.

The last decades have seen the rise of Machine Learning (ML) as a tool for data analysis in remote sensing and the Earth sciences. That is, large-scale statistical data analysis is used to capture the complex input-output relationships in a data-driven manner. ML is a powerful tool to recognise the underlying patterns in data where mechanistic models are lacking or too complicated. We leverage it to create a 20 year time series of ice cover in Swiss mountain lakes primarily using the Terra MODIS (`https://terra.nasa.gov/about/terra-instruments/modis`) data, and show empirically that the ice formation indeed follows a decreasing trend. We cast lake ice detection as a 2-class (*frozen, non-frozen*) per-pixel supervised classification problem. Class *frozen* represents both snow-on-ice and snow-free-ice pixels, while *non-frozen* denotes open water. As part of our study, we compare the performance of three popular ML methods: Support Vector Machine (Cortes and Vapnik, 1995, SVM), Random Forest (Breiman, 2001, RF), and XGBoost (Chen and Guestrin, 2016, XGB). Additionally, we assess the sensitivity of these classifiers to the respective hyper-parameters. We find that a linear SVM offers the best generalisation across winters and lakes for our data, and derive LIP from the resulting time series by fitting a piece-wise linear model per winter.

4.1.1. Operational lake ice / snow products

To our knowledge, the only operational lake ice product at present is the Climate Change Initiative Lake Ice Cover (Crétaux et al., 2020, CCI LIC). A comparison of our results with CCI LIC is however not possible, since none of our target lakes are included in the list of 250 lakes covered by the product. A second product, Copernicus Lake Ice Extent (LIE, `https://land.copernicus.eu/global/products/lie`), is still in pre-operational stage due to accuracy issues, and coverage only starts in 2017. Though not designed for lake ice, the MODIS Snow Product (Hall and Riggs, 2016, MSP) and VIIRS Snow Product (VSP, `https://nsidc.org/sites/nsidc.org/files/technical-references/VIIRS-snow-products-user-guide-final.pdf`) are also possible options for comparison, since lakes in the Alps are typically snow-covered for most of the frozen period. We cross check our results with these two snow products, see Section 4.5.2. More details on all the mentioned products can be found in Table B.1 (Appendix B).

4.1.2. Definitions used

We exclude mixed pixels and work only with pixels that lie completely inside a lake, termed *clean pixels*. *Non-transition dates* are the days when a lake is either completely frozen or completely non-frozen, the remaining days in a winter season are termed *transition dates*. By winter season, we denote all dates from September till May for our

purposes.

4.2. Related work

4.2.1. LIP trend analysis studies

The LIP trends of several lakes with different geographical conditions have been studied and reported in the literature. Though most of them use information from various ice databases (e.g., NSIDC), some studies directly derived the trends from radar and optical satellite data. In the following we focus on studies that, like ours, analyse MODIS and/or VIIRS optical satellite imagery to examine trends over multiple winters.

Šmejkalová et al. (2016) extracted the LIP trends (2000–2013) for 13,300 Arctic lakes (area >1 km^2) using MODIS imagery, and earlier break-up trends were noticed. They reported a mean shift in BUS in the range: -0.10 d/a (Northern Europe) to -1.05 d/a (central Siberia), and BUE in the range: -0.14 d/a to -0.72 d/a. Kropáček et al. (2013) studied the LIP trends of 59 lakes on the Tibetan Plateau from 2001 to 2010 using MODIS data. However, the estimated LIP trends varied across the target lakes and it was concluded that 10 year time span is too short to draw firm conclusion about LIP trends. Gou et al. (2015) analysed the ice formation trends (2000–2013) in lake Nam Co (Tibet, area 1920 km^2) using MODIS and in-situ data and found strong correlations with air temperature and wind speed patterns. This study found that high wind speeds during winter time could potentially expedite the freeze-up process. Additionally, this work reported a significant reduction of the total freeze duration. Gou et al. (2017) later analysed Nam Co for the period 2000 till 2015 using multiple MODIS products (MOD11A1, MYD11A1, MOD09GQ, MYD09GQ, and MOD10A1) and reported delayed FUS (0.58 d/a) as well as BUS (0.09 d/a), and reduced ice duration (-0.49 d/a) trends. Another study (Yao et al., 2016) also noted increasingly shorter freeze duration during the period 2000–2011 when investigating the lakes in Hoh Xil region (Tibet, 22 lakes with area > 100 km^2), using MODIS, Landsat TM/ETM+, and meteorological data. In addition, that work estimated late freeze-up and early break-up trends. They reported that the FUS, FUE, BUS, BUE, CFD and ICD shifted on average by 0.73, 0.34, −1.66, −0.81, −1.91, −2.21 d/a respectively. Cai et al. (2019) also analysed 58 lakes located on the Tibetan Plateau during the period from 2001 till 2017 using both Terra and Aqua MODIS imagery. For 47 lakes, a later FUS was noticed (0.55 d/a) while for the remaining 11 lakes an earlier FUS was observed (-0.44 d/a). For 50% of the target lakes, an earlier BUE (-0.69 d/a) was observed, however, for the other half a later BUE (0.39 d/a) was noted. Additionally, they reported a reduced ice cover duration for 40 lakes (-0.8 d/a), while for 18 lakes an increase was observed (1.11 d/a).

Yang et al. (2019) used MODIS to estimate the LIP trends for 8 large lakes (106 - 3461 km^2) in Northeastern China from 2003 to 2016. Later FUS (0.65 d/a), earlier BUE (-0.19 d/a) and shorter freeze duration (-0.84 d/a) trends were noticed. Qi et al. (2020) used AVHRR, MODIS, and Landsat data to extract the LIP of Qinghai lake (China, area of 4294 km^2) for the period 1980–2018. They estimated a shift of 0.16, 0.19, −0.36, and −0.42 d/a for FUS, FUE, BUS and BUE respectively, also pointing towards progressively

later freeze-up and earlier break-up. Additionally, they computed the decreasing patterns in ICD (−0.58 d/a) and CFD (−0.52 d/a). The study also identified correlations between the LIP and climate indicators like the Accumulated Freezing Degree-Days (AFDD), wind speed, precipitation, etc. during the winter season. Cai et al. (2020) used a threshold-based method to extract LIP trends from MODIS snow product for 23 lakes (2001-2018, Xinjiang Uygur Region, area: 11 to 1004 km^2) in China. They found that the ICD decreased (-1.08 d/a) in 16 out of the 23 lakes and increased (1.18 d/a) for the rest. They also reported later freeze-up (0.52 d/a) and earlier break-up (-0.51 d/a) in 17 and 18 lakes respectively. Additionally, they found that the freeze-up events are more affected by lake-specific factors such as area, mineralisation, etc.; while climatic factors like water surface temperature have more influence on the break-up events. That work also emphasised that lake surface water temperature has a stronger influence on the LIP events than air temperature.

Latifovic and Pouliot (2007) used the historical data record of AVHRR satellite in addition to in-situ measurements to perform long-term (1950-2004) trend analysis in Canadian lakes, via an automated profile feature extraction procedure, confirming later freeze-up (0.12 d/a) and earlier break-up (—0.18 d/a) for the majority of lakes that were analysed. They suggested that their procedure to extract the LIP events is not sensor-specific and could be applied to other satellite data, too. Murfitt and Brown (2017) also used MODIS data, to extract lake ice trends (2001-2014) for the regions Ontario and Manitoba in Canada. However, the discovered trends varied across regions. Zhang et al. (2021) put forward a new LIP database (4241 lakes with a minimum area of 1 km^2) for Alaskan lakes covering the period 2000-2019. Ice-on/off dates and freeze duration values included in this dataset were extracted from MODIS data using a threshold-based method. Additionally, they estimated the following LIP trends: later freeze-up (0.29 d/a) and earlier break-up (-0.55 d/a) were recorded for 289 and 440 lakes, respectively, while earlier freeze-up (-0.33 d/a) and later break-up (0.75 d/a) were noticed only for 11 and 4 lakes, respectively.

Compared to MODIS, the literature on lake ice monitoring with VIIRS data is limited. Sütterlin et al. (2017) estimated the LIP dates for winter 2016–17 in selected Swiss lakes using the lake surface water temperature (LSWT) derived from visible and near-infrared reflectances, and thermal infrared band (I_5) of VIIRS data. Later, for winter 2016–17, Tom et al. (2020c) estimated the LIP dates of lakes Sihl, Sils, Silvaplana, and St. Moritz from VIIRS and MODIS data. To our knowledge, no multi-winter LIP trend analysis based on VIIRS data has been reported yet.

To summarise, most related works reviewed so far have found trends towards late freeze-up, early break-up and declining freeze duration. The prevalent methods are physics-inspired models based on empirical indices and thresholds. Most studies focus on large lakes, often in the Arctic and sub-Arctic regions. Beyond analysing the lakes of the Oberengadin region, in the present work we also show that supervised ML models are able to detect lake ice with high accuracy, hoping that the outcome may be useful for future research. To our knowledge, none of the earlier trend studies applied data-driven ML methods to identify lake ice.

For Swiss lakes, a previous study (Hendricks Franssen and Scherrer, 2008) verified

that the lake ice formation and surrounding air temperature are heavily correlated. They deducted an empirical relationship between sum of negative degree days (same as AFDD) and the lake ice formation process, and modelled the probability of ice cover via binomial logistic regression. Though this approach gathered and used the temperature data from 1901–2006 to study eleven lakes in the lower-lying Swiss plateau, none of the target mountain lakes were included. Our case study focuses on the Oberengadin region, with three main lakes: *Sils*, *Silvaplana* and *St. Moritz* (with the latter very small for the GSD of MODIS). Moreover, we include another Alpine lake Sihl to check generality.

Our goal in this work is lake ice monitoring using only image data from optical satellites, which provides a direct, data-driven observation not influenced by model assumptions about the ice formation process. We see satellite imagery as an independent information source and consider image analysis complementary to other methods of lake ice modelling. Furthermore, compared to our satellite-based approach, which can easily analyse the whole lake area, it is difficult to effectively derive the spatial extent of lake ice from the temperature (point) observations recorded at one or a few nearby weather stations, even if they are situated in the immediate vicinity of the lake (which is not always the case). This is even more true in the Alpine terrain that we target, due to strong micro-climatic effects.

4.2.2. Lake ice observation with machine and deep learning

ML algorithms have become a standard tool for several environmental remote sensing research problems, including our earlier works on monitoring lake ice cover. In Tom et al. (2018), we already investigated pixel-wise classification of the spatio-temporal extent of lake ice from MODIS and VIIRS imagery with SVM. Each pixel was classified as either *frozen* or *non-frozen* in a supervised manner. Though this approach achieved strong results (including generalisation across winters and across lakes with similar geographic conditions), the test set at the time did not have a complete winter of reference data (including the critical freeze-up and break-up periods), due to technical problems. Later, Tom et al. (2020c) presented extensive experiments on data from two full winters and confirmed the efficacy of SVM for lake ice monitoring with MODIS and VIIRS. Both these works dealt with small- and mid-sized Swiss mountain lakes. Xiao et al. (2018) and Prabha et al. (2020) explored the potential of convolutional neural networks (CNNs) for lake ice detection in terrestrial webcam images (RGB). They performed a supervised classification of the lake pixels using the Tiramisu (Jégou et al., 2016), respectively Deeplab v3+ (Chen et al., 2018b) networks, into the four classes: *water, ice, snow* and *clutter*. An integrated approach using both the satellite and webcam observations was discussed in Tom et al. (2019) to estimate the ice-on and ice-off dates. Recently, Hoekstra et al. (2020) proposed an automated approach for ice vs. water classification in RADARSAT-2 data, combining unsupervised Iterative Region Growing using Semantics (IRGS) and supervised RF labelling. A deep learning approach to lake ice detection in Sentinel-1 SAR imagery has been described in Tom et al. (2020a), and achieved promising results, including generalisation across lakes and winters. Very recently, Wu et al. (2021b) compared the capabilities of four different ML methodologies: Multinomial Logistic Regression (MLR), SVM, RF, and Gradient Boosting Trees (GBT) for lake ice observation using the MODIS

Top of Atmosphere (TOA) product. They modelled lake ice monitoring as a 3-class (*ice, water, cloud*) supervised classification problem. The four classifiers were tested on 17 large lakes from North America and Europe with areas >1040 km^2, and achieved >94% accuracy. RF and GBT showed better generalisation performance on this dataset of large lakes.

4.3. Study area and data

4.3.1. Study area

We process four small-to-medium-sized Swiss Alpine lakes: *Sihl, Sils, Silvaplana* and *St. Moritz*, see Fig. 4.1 and Table B.2 (Appendix B). While most of the earlier works (Gou et al., 2015, Yao et al., 2016, Qi et al., 2019, 2020) on long time series monitoring of lake ice with MODIS concentrated on larger lakes, many lakes that freeze are actually small- or medium-sized mountain lakes, especially outside the (sub-)Arctic regions. The lakes we analyse are relatively small in area (0.78 - 11.3 km^2), representative for this category.

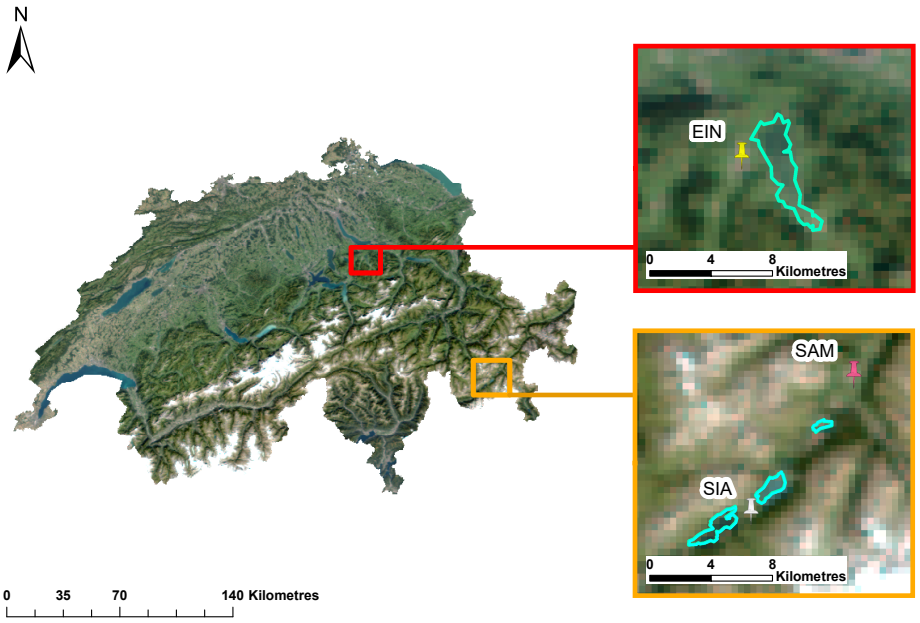


Figure 4.1.: MODIS orthophoto map (RGB composite, red: band 1, green: band 4, blue: band 3) of Switzerland (left) captured on 7 September 2016. Red and amber rectangles show the regions Einsiedeln (around lake Sihl) and Oberengadin (with lakes Sils, Silvaplana and St. Moritz, from left to right) respectively. Inside each zoomed rectangle on the right, the respective lake outlines are shown in light green and the nearest meteorological stations (EIN, SIA and SAM) are marked using pins.

The lakes were chosen according to the needs of two projects of the Swiss GCOS office (Tom et al., 2019, 2020b). For the three small lakes in the region Oberengadin

(Sils, Silvaplana, St. Moritz), located at an altitude > 1750 m, there are long in-situ observation series (important for climate studies), and they are also included in the NSIDC lake ice database (`https://nsidc.org`), although not updated recently. The fourth lake (Sihl) from the region Einsiedeln has been chosen mainly to test generality, as it is relatively larger, lies at a lower altitude on the North slope of the Alps, and has different environmental conditions, see Table B.2 (Appendix B). The three lakes in Oberengadin fully freeze every year, whereas lake Sihl does not (but still freezes in most winters).

For these four lakes there is no reference freeze/thaw data available from the past two decades. Hence, we study the weather patterns in the regions near the lakes. For each lake, the temperature and precipitation data recorded at the nearest meteorological stations are shown in Fig. 4.2. On the top and bottom rows, we plot the mean temperature and total precipitation (solid curves) during the winter months on the y-axis, against the winters on the x-axis in a chronological order. Additionally, in both rows, we plot the linearly fitted trend curve (dotted line) for each station.

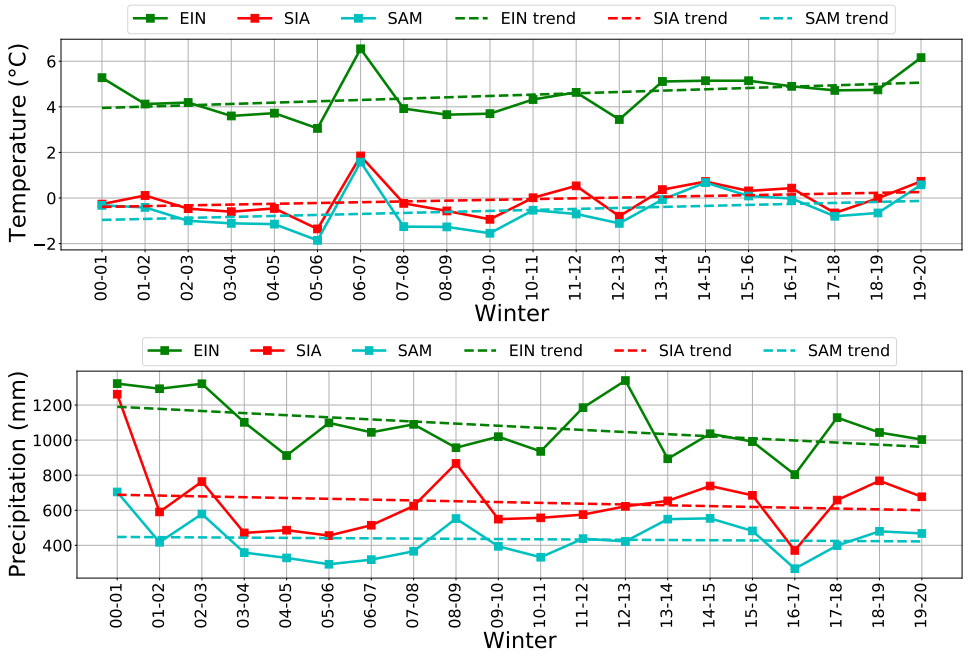


Figure 4.2.: For the region near the lakes, the mean winter air temperature (row 1) and total winter precipitation (row 2) are plotted (solid curve) on the y-axis against the winters shown in a chronological order on the x-axis. The data from nearest meteorological stations: EIN (Sihl), SIA (Sils and Silvaplana) and SAM (St. Moritz) are used. The corresponding trends (linear fit, dotted curve) are also shown with the same colour. Data courtesy of MeteoSwiss.

For lakes Sihl and St. Moritz, the nearest meteorological stations are Einsiedeln (EIN) and Samedan (SAM) respectively as shown in Fig. 4.1 (see also Table B.2 in Appendix B). Lakes Sils and Silvaplana are located next to each other and hence share the same meteorological station: Segl Maria (SIA). The station Einsiedeln is located at a relatively lower altitude and closer to the Swiss plateau, while the other two stations are in the Engadin valley at higher altitudes. This explains why the absolute temperature

is relatively higher for Einsiedeln. Additionally, the stations SIA and SAM are located within 20 km from one another, see Fig 4.1, and hence have similar temperature and precipitation values. Exceptionally high winter temperature was recorded at all three stations in winter 2006–07.

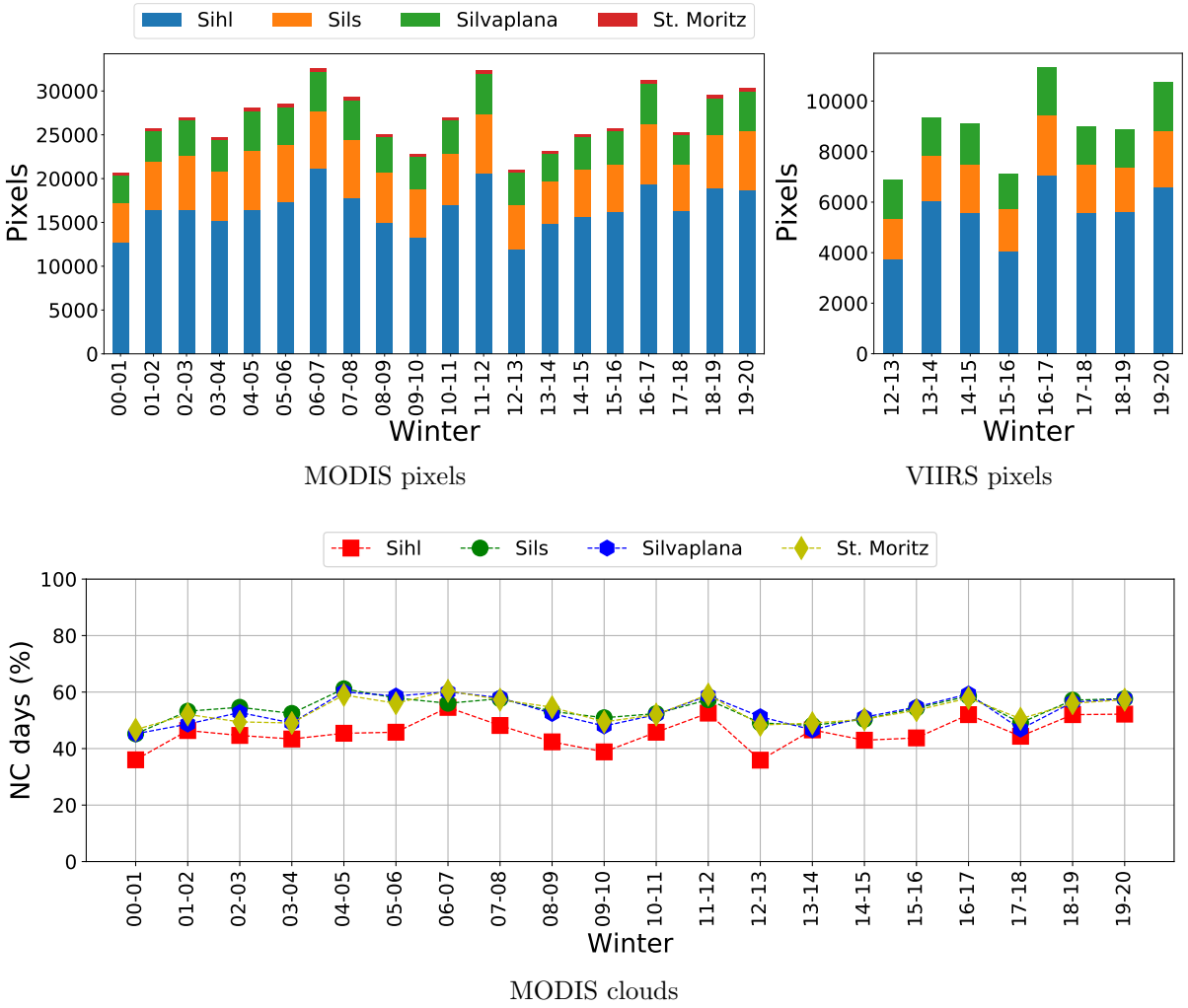
It can be seen from Fig. 4.2 that during the past 20 winters, at all the three stations, the mean temperature follows an increasing trend. On the other hand, precipitation (snow+rain) has a decreasing pattern. While Meteoswiss (MCH) has reported a significant trend of temperature increase in the Swiss Alps since 1864, they have so far not confirmed a significant precipitation trend (`https://www.meteoswiss.admin.ch/home/climate/climate-change-in-switzerland/temperature-and-precipitation-trends.html`). Over the shorter period of the past 20 winters, precipitation has been slightly declining. Warmer winters at higher altitudes in Switzerland could be linked to a decrease in precipitation, see Rebetez (1996). The pattern of precipitation over the 20-year period differs somewhat between Einsiedeln and the two other (similar) stations, e.g., see the winters 08–09, 12–13.

4.3.2. Data

In our analysis, we use the data from Terra MODIS and Suomi NPP VIIRS (`https://ncc.nesdis.noaa.gov/VIIRS/`) satellites downloaded from the LAADS (`https://ladsweb.modaps.eosdis.nasa.gov`) and NOAA (`https://www.avl.class.noaa.gov/`) databases, respectively. For MODIS processing, we downloaded the MOD02 (geolocated and calibrated radiance, level 1b, TOA), MOD03 (geolocation) and MOD35_L2 (cloud mask) products and pre-processed using *MRTSWATH* (`https://lpdaac.usgs.gov/tools/modis_reprojection_tool_swath/`, re-projection and re-sampling) and *LDOPE* (Roy et al., 2002, cloud mask) software. For VIIRS, we downloaded the Scientific Data Record (SDR) data for the imagery bands, IICMO and VICMO products for the cloud masks, and GITCO (for image bands) and GMTCO (for cloud masks) for terrain corrected geolocation. VIIRS pre-processing is done using the following software packages: *SatPy* (`https://satpy.readthedocs.io/`) for assembling the data granules, mapping and re-sampling, *H5py* (`https://www.h5py.org`) for cloud mask extraction, *PyResample* (`https://resample.readthedocs.io`) and *GDAL* (`https://gdal.org`) for re-sampling of cloud masks. As in Tom et al. (2020c), we use only twelve (five) selected MODIS (VIIRS) bands to form the feature vector. The bands that offer maximum separabilty for the task of lake ice monitoring were automatically chosen by the supervised XGB feature selection algorithm (Chen and Guestrin, 2016). The spectral coverage of these bands is shown in Fig. B.1 (Appendix B) where MODIS (M) and VIIRS (V) bands are displayed as blue and red bars, respectively. For each band, the bar width is proportional to the corresponding bandwidth.

We analyse MODIS data from all winters since 2000–01 (20 winters), and VIIRS data from all winters since 2012–13 (8 winters). In each winter, we process all the dates from the beginning of September until the end of May on which at least 30% of a lake is free of clouds, according to the mask. Fig. 4.3 displays more details of the MODIS and VIIRS data that we use as a stacked bar chart (one colour per lake). For all target lakes,

the total number of cloud-free, clean pixels in each winter is shown on the y-axis, against the winters in chronological order on the x-axis.



MODIS pixels

VIIRS pixels

MODIS clouds

Figure 4.3.: First row displays the *clean, cloud-free* pixels (from *transition* and *non-transition* dates) from the four target lakes (Sihl, Sils, Silvaplana, St. Moritz) used in our experiments. Data from both MODIS (20 winters, 4 lakes) and VIIRS (8 winters, 3 lakes) is displayed. Second row shows the percentage of at least 30% Non-Cloudy (NC) days during each winter (derived from the MODIS cloud mask). Winter 00–01 represents the dates from September 2000 till May 2001 (and similarly other winters).

In Fig. 4.3, note that some winters are relatively less cloudy. This explains why the number of pixels vary across winters, even for the same lake and sensor. On average there are 16558, 5899, 3972 and 329 MODIS pixels per winter for the lakes Sihl, Sils, Silvaplana and St. Moritz respectively. Similarly, there are 5538, 1901 and 1606 VIIRS pixels for the lakes Sihl, Sils and Silvaplana respectively. Due to its small size there exist no clean pixel for St. Moritz in VIIRS imagery bands (Tom et al., 2020c), hence we exclude it from the VIIRS analysis. The number of pixels per acquisition is proportional to the lake area and hence varies across lakes even for a given sensor. Additionally, for a given lake, the number of VIIRS pixels per acquisition is lower compared to MODIS, due to

the higher GSD of VIIRS imagery bands ($\approx 375m$) compared to MODIS (250m). We super-resolved all low resolution MODIS bands (500m, 1000m) to 250m using bilinear interpolation prior to the analysis. This step is not required for VIIRS as all used bands have the same GSD.

It can be inferred from Fig. 4.3 that in both VIIRS and MODIS plots, the cloud patterns of lakes Sils, Silvaplana and St. Moritz are quite similar, due to geographical proximity (see also Fig. 4.1). Minor differences exist (in few winters) between the two very nearby lakes Sils and Silvaplana due to cloud mask errors (see also Section 4.4.1). Lake Sihl has a different cloud coverage pattern compared to other three lakes, due to its lower altitude and different surrounding topography. Both MODIS and VIIRS have daily temporal resolution, but the data capture can happen at different times within a day. Consequently, the cloud pattern (and hence the cloud masks) can differ even between MODIS and VIIRS acquisitions from the same day. For all four lakes, Fig. 4.3 shows the percentage of non-cloudy days (at least 30% cloud-free pixels) in each winter season. It can be seen that ≈40 to ≈60% of all days are not usable due to clouds, significantly reducing the effective temporal resolution. We notice that the data loss is worse for lake Sihl located near the plateau.

Ground truth

We use the same ground truth as in Tom et al. (2020c), which is based on visual interpretation of freely available high-resolution webcams monitoring the target lakes. One label (*fully frozen, fully non-frozen, partially frozen*) per day is assigned. Two different operators looked at each image, i.e., a second expert verified the judgement of the first operator to minimise interpretation errors. When deciphering a webcam image was difficult, additional images were used from other webcams viewing the same lake (if available), images from the same webcam but at other acquisition times on the same day, and images of the same webcam for the days before and after the given observation day. We also improved the webcam-based ground truth using sporadic information available from media reports, and by visually interpreting Sentinel-2 images, whenever available and cloud-free. No webcam data is available from the winters before 2016-17. Moreover, the manual interpretation process is very labour intensive. Thus, ground truth is available only for two winters (2016–17, 2017–18).

Even though visual interpretation is the standard practice, a certain level of label noise inevitably remains in the ground truth, due to factors such as interpretation errors, image compression artefacts, large distance and flat viewing angle on the lake, etc. Furthermore, the webcams used are not optimally mounted for lake ice monitoring application and hence do not always cover the full lake area (or even a major portion of it), even for the smallest lake St. Moritz. Still, the ground truth serves the purpose, in the sense that it has significantly fewer wrongly labelled pixels than the automatic prediction results. We see no possibility to obtain more accurate, spatially explicit ground truth for our task.

4.4. Methodology

4.4.1. Pre-processing

We perform the same pre-processing steps as in Tom et al. (2020c). First, the absolute geolocation error for both sensors (0.75, respectively 0.85 pixels x- and y-shifts for MODIS; 0.0, respectively 0.3 pixels x- and y-shifts for VIIRS) are corrected. The generalised (Douglas and Peucker, 1973) lake outlines are then backprojected onto the images to extract the clean pixels. Binary cloud masks are derived from the respective cloud mask products to limit the analysis only to cloud-free pixels. We noticed some errors in both MODIS and VIIRS cloud masks. The most critical ones are false negatives, where an actually cloudy pixel goes undetected. Such cases can corrupt model learning and inference and introduce errors in the predicted ice maps.

4.4.2. Machine learning for lake ice extraction

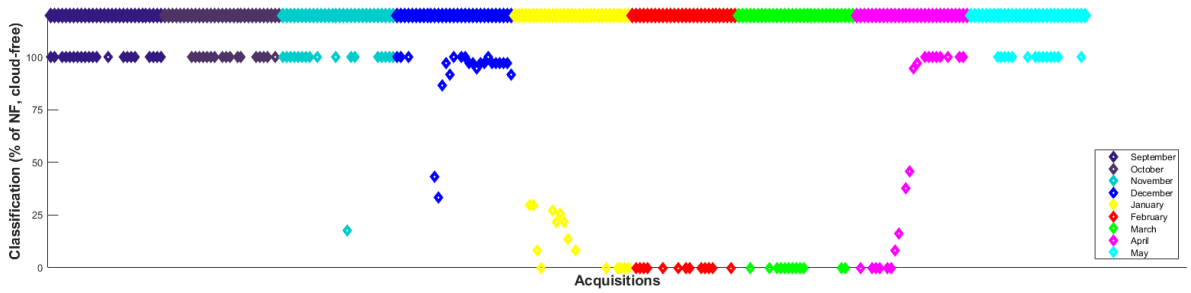
We model lake ice detection in optical satellite images as a per pixel 2-class (*frozen, non-frozen*) supervised classification problem. For each pixel, the feature vector is formed by directly stacking the 12 (5) bands of MODIS (VIIRS), see Fig. B.1 (Appendix B). We treat *snow-on-ice* and *snow-free-ice* as a single class: *frozen*. Class *non-frozen* denotes the water pixels. The class distributions in winters 2016–17 and 2017–18 are shown in Fig. B.2 (Appendix B). There is a significant class imbalance in our dataset, since we include all cloud-free dates from September till May, of which only a minority is frozen. We have tested four off-the-shelf ML classifiers: linear SVM [SL], SVM with Radial Basis Function (RBF) kernel [SR], Random Forest [RF] and eXtreme Gradient Boosting [XGB] to perform pixel-wise supervised classification in order to predict the state of a lake.

We recall that SVM is a linear large-margin classifier, which can be extended to non-linear class boundaries with the kernel trick. The choice of kernel is critical and depends on the data distribution. In our case we tested both linear and RBF kernels. RF is an ensemble learning approach which relies on bagging (bootstrap aggregation) of multiple decision trees constructed from the data, using randomisation to decorrelate the individual trees. XGB is also an ensemble method based on (shallower) decision trees, but iteratively learns the trees with gradient descent, such that each tree corrects error of earlier ones. XGB is highly scalable and exploits sparsity. We note that several comparison studies exist in the literature for other applications than ours, e.g., (Ogutu et al., 2011, Wainer, 2016, Pham et al., 2020).

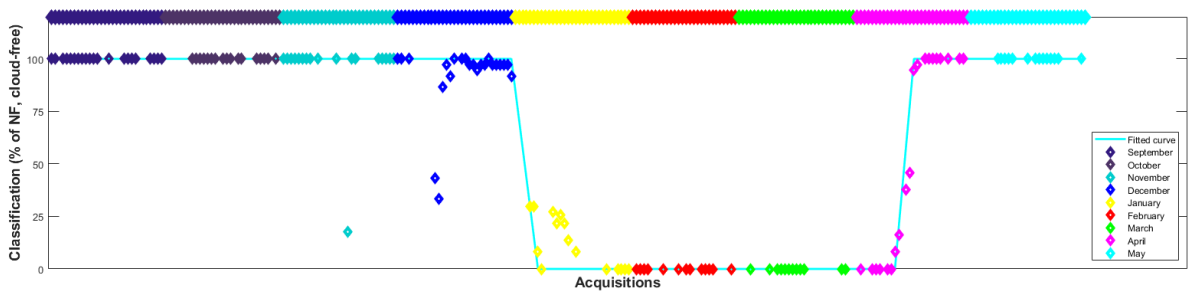
Note also, while there recently has been a strong interest in deep learning for remote sensing tasks, it is not suitable for our particular application, due to the scarcity of pixels with reliable ground truth. The lakes that we monitor are small and ground truth is available for two winters (see Section 4.3.2, and Fig. B.2 in Appendix B), which is too little to train data-greedy deep neural networks. Also, given the large GSD and limited need for spatial context, we do not expect deep models to greatly outperform our shallower ones, see Section 4.5.

4.4.3. LIP estimation

In each winter, using the trained ML model, we process all available acquisitions that are at least 30% cloud-free and generate pixel-wise classification maps (one per acquisition). To recover the temporal evolution (per winter), the percentage of non-frozen pixels is computed from each classification map and is plotted on the y-axis against the acquisition time on the x-axis. Then, as in Tom et al. (2020c), multi-temporal smoothing is performed as a post-processing step using a Gaussian kernel with standard deviation 0.6 days and window width 3 days. An example MODIS results timeline for lake Sils from winter 2006–07 is shown in Fig. 4.4 (top). Results from different months are displayed with different colours, see the legend. In addition, only the acquisitions which are at least 30% non-cloudy are displayed.



Post-processed results timeline before curve fitting



After curve fitting

Figure 4.4.: Piece-wise linear ("U with wings") curve fitting example. NF indicates Non-Frozen.

In the post-processed timeline, we find all the potential candidates for the following four critical dates: FUS, FUE, BUS and BUE, see Table 4.1 for the corresponding definitions. Within a winter, it is possible that >1 candidates exist per critical date which all satisfy the respective definition. In order to weed out some obviously spurious candidates, we enforce the constraint that the four dates must occur in the following chronological order: FUS→FUE→BUS→BUE. Then we exhaustively search for the optimal set of four dates among the remaining candidates. To that end, we fit a continuous, piece-wise linear *"U with wings"* shape to the per-day values of percentage of non-frozen pixels, such that the fitting residuals z are minimised (see example fit in Fig. 4.4 [bottom], shown in cyan

Table 4.1.: Key LIP events.

Event	Definition
FUS	30% or more of the non-cloudy portion of the lake is frozen and the just previous non-cloudy day should be < 30% frozen
FUE	70% or more of the non-cloudy portion of the lake is frozen and the just previous non-cloudy day should be < 70% frozen
BUS	30% or more of the non-cloudy portion of the lake is non-frozen and the just previous non-cloudy day should be < 30% non-frozen
BUE	70% or more of the non-cloudy portion of the lake is non-frozen and the just previous non-cloudy day should be < 70% non-frozen
ICD	BUE - FUS
CFD	BUS - FUE

colour). In detail, the loss function for the fit is defined as:

$$L_{LIP} = \frac{1}{P} \cdot \sum_{i=1}^{N} H_\phi(z) \tag{4.1}$$

where N is the total number of available acquisitions that are at least 30% cloud-free.

$$H_\phi(z) = \begin{cases} z^2 & |z| \leq \phi \\ 2\phi|z| - \phi^2 & |z| > \phi \end{cases} \tag{4.2}$$

is the Huber norm of the residual. For the shape parameter ϕ, we use a constant value of 1.35 which offers a good trade-off between the robust l_1-norm for large residuals and the statistically efficient l_2-norm for small residuals (Owen, 2006).

Per lake, we assume that each critical date occurs only once per winter, which is always true in Oberengadin. Lake Sihl does not always fully freeze. As it lies outside of the target region and is included mostly to ensure generality of the ice classifier, we do not extract the LIP events for Sihl. Moreover, we decide to exclude lake St. Moritz since it is too small for the GSD of MODIS (only 4 clean pixels), making the fraction of frozen pixels overly susceptible to noise. We thus prefer to study only the two main lakes in Oberengadin, Sils and Silvaplana, in terms of long-term lake ice trends. These two lakes fully freeze every year and typically have a single freeze-up and break-up period. To further stabilise the LIP estimates we include a weak prior for each phenological date, in the form of a diffuse Gaussian distribution.

The prior probability (P) is given by:

$$P = P_{fus} \cdot P_{fue} \cdot P_{bus} \cdot P_{bue} \tag{4.3}$$

where P_{fus}, P_{fue}, P_{bus}, and P_{bue} are Gaussian normal distributions for the events FUS, FUE, BUS and BUE, respectively. The prior formalises the knowledge that freeze-up normally occurs around the end of December and takes around three days, and break-up occurs around the end of April over a similar period, for both target lakes. In order not to

bias the estimation, but only to minimise the risk of implausible results, we choose very wide Gaussians ($\sigma = 1$ month). Furthermore, we impose a constraint that the duration of freeze-up (FUE-FUS) and break-up (BUE-BUS) is not more than two weeks.

Inspired by Qi et al. (2020), we additionally compute further indicators that can be derived from the four critical dates, namely ICD and CFD, see Table 4.1 for details. We use 30% as the threshold to estimate the four dates. For example, a date is considered as FUS candidate if 30% or more of the non-cloudy portion of the lake is frozen. Some studies based on MODIS (Reed et al., 2009, Yao et al., 2016, Qi et al., 2020) have used 10% as threshold, while another approach (Kropáček et al., 2013) even employed 5%. All of them monitored larger lakes (45 to 4294 km^2 in area). We empirically found that for our comparatively tiny lakes the above thresholds are too strict and a threshold of 30% is needed to ensure reliable decisions. To see why, consider that in the best case (Sils, cloud-free) a lake has 33 clean pixels, but that number can go down to as few as 7 (Silvaplana at 70% cloud cover). Note also that on such small lakes a large portion of all pixels is very close to the lake's shoreline, where the residual absolute geolocation error (in the worst case 0.5 pixel) may have a significant impact.

4.5. Results

In our experiments, we perform a comparison of the performance of various ML classifiers, derive a 20-year time series of the key phenological dates, and perform correlation of our results with the regional weather trends. In addition to the overall classification accuracy (Acc), we report a stricter measure: mean intersection-over-union (mIoU) for fair reporting of the results on a dataset with imbalanced class distribution. More details can be found in the following sub-sections.

4.5.1. Choice of machine learning method

We first conduct experiments on the two winters 2016–17 and 2017–18, for which ground truth is available. These experiments serve to compare the performance of the SVM, RF, and XGB classifiers and determine which one is most suitable for our task and dataset. We train the models and report quantitative results only on non-transition dates, since per-pixel ground truth is not available for the transition dates.

Hyper-parameter tuning

In SVM, the two main hyper-parameters are *cost* and *gamma* (for the linear variant just the *cost*) which indicate misclassification cost and kernel width respectively. Higher values of *cost* means the model chooses more support vectors which effectively increases the variance and decreases the bias and results in overfitting and vice versa. When *gamma* is set to a high value, more weight is assigned to the points close to the hyperplane compared to the far away ones. Higher *gamma* values also result in low bias and high variance thus causing overfitting. For RF, *num_trees* represents the number of trees in the forest while *num_variables* denotes the number of predictors (features) to select at random for each

split. Note that the *minimum_leaf_size* is set to 1 in all our RF experiments. In XGB, the three most important hyper-parameters are number of trees (*num_trees*), *learning_rate* and the tree specific parameter: *colsample_bytree*. *Learning rate* shrinks the contribution of each new tree to make the boosting procedure more conservative and thus the resulting model more robust. *Colsample_bytree* controls the fraction of features (spectral bands) to be used in each boosting iteration.

For each ML approach, the optimal hyper-parameters are first independently determined with a grid search. The best-performing hyper-parameters thus chosen are shown in Table 4.2. We use these parameters in all further experiments. However, note that the parameters are dataset-dependent, too. Overall, RF, XGB and SL exhibited fairly stable results across a range of hyper-parameters, whereas SR was very sensitive.

Table 4.2.: Optimum hyper-parameters of each classifier and sensor estimated using grid search.

Method	**Sensor**	**Hyper-parameters**
SL	MODIS	*cost* 0.1
SR	MODIS	*cost* 10, *gamma* 1
RF	MODIS	*num trees* 500, *num variables* 10
XGB	MODIS	*num trees* 1000, *colsample_bytree* 1, *learning rate* 0.2
SL	VIIRS	*cost* 0.1
SR	VIIRS	*cost* 10, *gamma* 1
RF	VIIRS	*num trees* 500, *num variables* 3
XGB	VIIRS	*num trees* 500, *colsample_bytree* 1, *learning rate* 0.3

Four-fold cross validation

As the first experiment, we combine the data (independently for MODIS and VIIRS) of all the available lakes from winters 2016–17 and 2017–18 and perform 4-fold cross validation and report the overall accuracy and $mIoU$, see Table 4.3. For both MODIS and VIIRS, the performance of all classifiers except SL is more or less the same, with accuracy differences below 1%. While SR performs marginally best on MODIS data, XGB scores well on VIIRS. Though lower than other three classifiers, the performance of SL is also very good on both sensors.

Generalisation experiments

In order to study how well the classifiers generalise across space and time, we train a model on all except one lake (respectively, winter) and test on the held-out lake (winter).

Fig. 4.5 displays the results (bar graphs showing overall accuracy and mIoU) of the four classifiers for *leave-one-lake-out* setting on MODIS (top row) and VIIRS (bottom row) data. It can be seen that the performance varies across lakes and sensors. For MODIS, XGB has, on average, a narrow advantage over SR and RF, with SL a bit

Table 4.3.: Four-fold cross-validation results (in %) on MODIS and VIIRS data. The data from both winters 2016–17 and 2017–18 are used in this analysis. Overall classification accuracy and mean intersection-over-union (mIoU) scores are shown. The best results are shown in bold.

Sensor	Feature vector	Method	Accuracy	mIoU
MODIS	All 12 bands	SL	93.4	83.9
MODIS	All 12 bands	SR	**99.4**	**98.5**
MODIS	10 bands (random)	RF	98.9	97.2
MODIS	All 12 bands	XGB	99.3	98.3
VIIRS	All 5 bands	SL	95.1	88.4
VIIRS	All 5 bands	SR	97.1	93.1
VIIRS	3 bands (random)	RF	97.6	**94.5**
VIIRS	All 5 bands	XGB	**97.7**	**94.5**

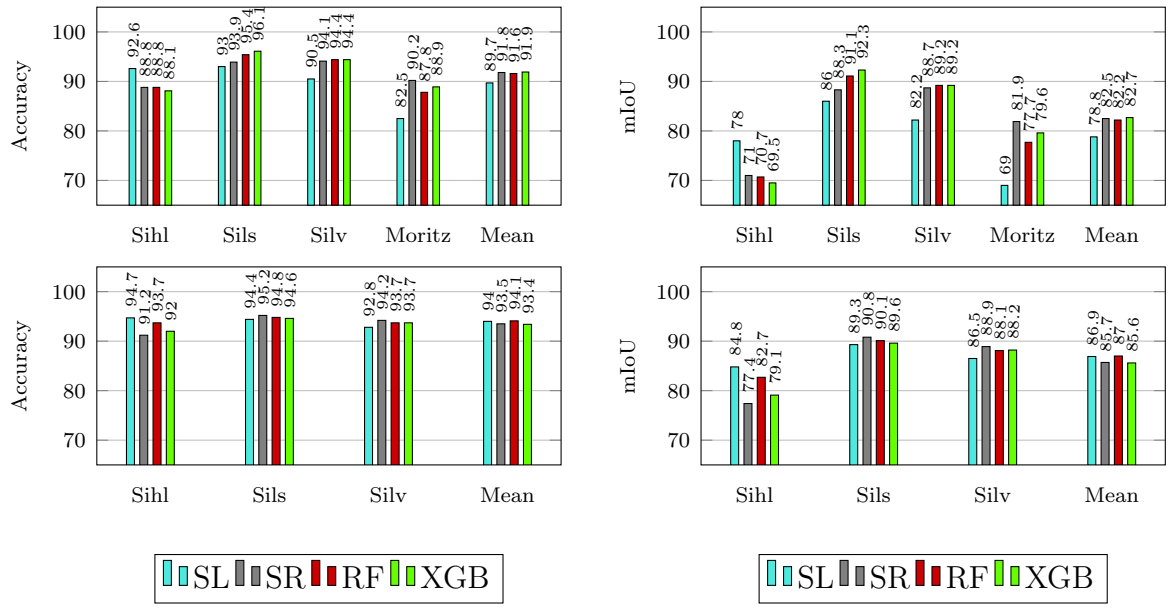


Figure 4.5.: Generalisation across lakes results on MODIS (top row) and VIIRS (bottom row) data for the classifiers SL, SR, RF and XGB on lakes Sihl, Sils, Silvaplana (Silv) and St. Moritz (Moritz). Both overall accuracy (left column) and mIoU (right column) are shown.

behind. For VIIRS, on the other hand, RF marked the best performance closely followed by SL, SR and XGB.

On both sensors, the best performance (especially in terms of mIoU) is achieved for the lakes Sils and Silvaplana. This is likely due to them having the most similar characteristics and imaging conditions, see Table B.2 (Appendix B). I.e., pixels from one of them are representative also of the other one, such that the classifier trained in one of the two generalises well to the other. Lake St. Moritz (only for MODIS) has too few clean pixels per acquisition to draw any conclusions about generalisation. However, we still include it in our processing to study how far lake ice monitoring with MODIS can be pushed (in terms of lake area) – indeed, the classification is $> 82.5\%$ correct. Lake

Sihl from the region Einsiedeln is different compared to other three lakes from the region Engadin in terms of area, weather, surrounding topography etc., c.f. Section 4.3.1. Hence, the performance on lake Sihl is interesting to assess geographical generalisation over longer distances. It can be seen in Fig. 4.5 that for lake Sihl the SL classifier performs best on both MODIS and VIIRS data, suggesting that the other (non-linear) models already overfit to the specific local conditions of Oberengadin.

As a second generalisation experiment, more important for our time series analysis, we check how well the trained classifiers can be transferred across different winters. We train on one winter and test the resulting model on the held-out winter (*leave-one-winter-out*), see Fig. 4.6. We only have data from two consecutive winters (2016–17, 2017–18) to perform this analysis. Still, we believe that the experiment is representative for generalisation to unseen years, since the weather conditions in different years are largely uncorrelated (c.f. Fig. 4.2). In particular for the two available winters, 2017–18 was markedly colder than the previous year, see Fig. 4.2.

Fig. 4.6 (top row) clearly shows for MODIS that SL adapts best to a new winter, with significantly higher generalisation losses for XGB, RF, and SR. For VIIRS (Fig. 4.6, bottom row), SL also copes best with generalisation across different years. Here RF can keep up, whereas SR and XGB again suffer higher generalisation losses.

Figure 4.6.: Generalisation across winters results on MODIS (top row) and VIIRS (bottom row) data for the classifiers SL, SR, RF and XGB. Both overall accuracy (left column) and mIoU (right column) are shown.

Overall, all classifiers exhibit a certain performance drop when having to generalise beyond the exact training conditions. Table 4.4 shows the detailed performance drops w.r.t. Table 4.3. Since mIoU is a stricter measure than accuracy, the drops are more pronounced. It is interesting to note that in both generalisation experiments the most robust (least affected) classifier is the (linear) SL. With all other classifiers the drop is much higher for unseen winters than for unseen lakes. We conclude that SL is the safest option four our task, where data from multiple different winters must be processed,

and use it for all further experiments. Recall, though, that we only have a relatively small dataset at our disposal from few small mountain lakes over two winters. It is quite possible that the small volume and specific geographical conditions aggravate the tendency to overfit, and that a higher-capacity, non-linear classifier will work best if a larger and more diverse dataset were available, or if the conditions were less variable (large lakes in smooth terrain).

Table 4.4.: Generalisation loss (across lakes / winters) encountered by each classifier: SL, SR, RF and XGB. Drop (in % points) for overall accuracy and mIoU are shown in black and grey respectively.

Sensor	Loss type	SL	SR	RF	XGB
MODIS	across lakes	3.7/5.1	7.6/16.0	7.3/15.0	7.4/15.8
MODIS	across winters	1.3/2.8	12.0/26.6	10.7/23.6	9.9/22.4
VIIRS	across lakes	1.1/1.5	3.6/7.4	3.5/7.5	4.3/8.9
VIIRS	across winters	2.6/5.7	5.5/12.2	5.1/11.9	6.7/14.5

4.5.2. Experiments on MODIS data from 20 winters

Test-train split

So far, we have used only on parts of the available ground truth for training, so as to evaluate the method. We now move on to the actual longer-term analysis, where we process MODIS data from all the 20 winters since 2000–01 (inclusive). Details of the training set for each tested winter is shown in Table 4.5. To avoid systematic biases in the estimated ice maps due to overfitting to a particular year, we proceed as follows: we train the SL model on all non-transition dates of 2016–2017 and use it to estimate lake ice coverage for all days in 2017–2018 (including transition dates). We repeat that procedure in the opposite direction, i.e., we train on all non-transition days of 2017–2018 and perform inference for all dates of 2016–2017. Then, we merge all non-transition dates from both winters into a new, larger *two-winter* training set, which we further augment with an *auxiliary dataset*. The latter contains all acquisitions of lakes Sils and Silvaplana captured during the remaining 18 years in September (when the lakes are never frozen) and in February (when the lakes are always frozen). The purpose of the auxiliary dataset is to cover a wider range of weather and lighting conditions that might not have been encountered in the two winters with annotated ground truth, for better generalisation. Data of lake Sihl is not included in the auxiliary set, as it does not freeze reliably, St. Moritz is ignored due to its negligible number of pixels. The two-winter and auxiliary datasets are merged and used to train a SL model, which is then used to predict ice cover maps for the 18 remaining winters.

Qualitative results

Exemplary qualitative results on some selected days (one per winter) of lake Sihl are shown in Fig. 4.7. The respective dates are displayed below each sub-figure. The lake

Table 4.5.: Test-train split for the MODIS data from 20 winters.

Test set	**Training set**
winter 2016–17	winter 2017–18
winter 2017–18	winter 2016–17
winters till 2015–16, from 2018–19	winters 2016–17 and 2017–18, auxiliary set

outline overlaid on the MODIS band B_1 is shown in green. Pixels detected as frozen and non-frozen are shown as blue and red squares, respectively. The results include fully-frozen, fully non-frozen and partially frozen days.

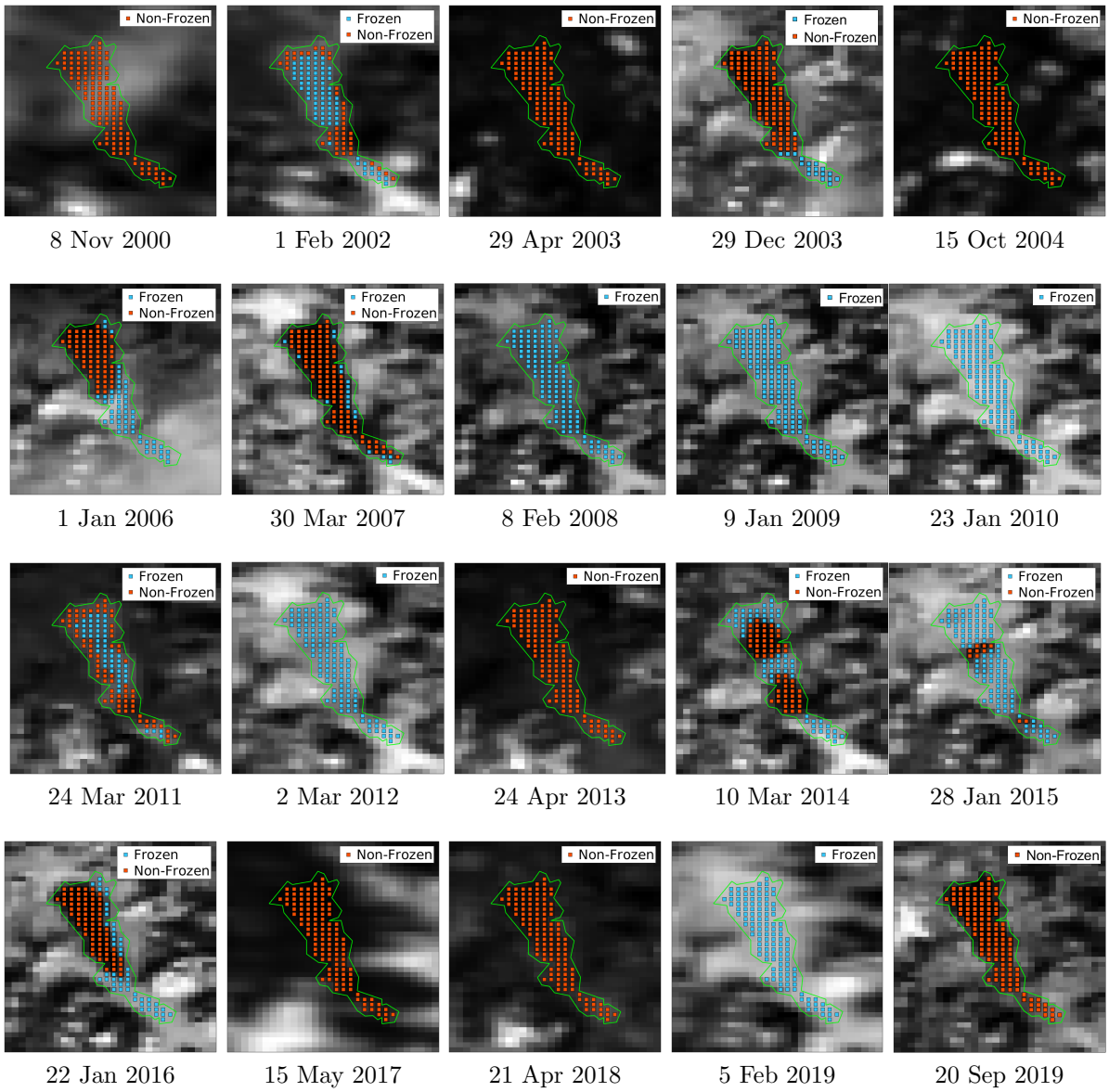


Figure 4.7.: MODIS qualitative classification results (overlaid on band B_1 from the respective day) for lake Sihl on selected dates from the past 20 winters using the SL classifier. Blue and red squares are overlaid on the pixels detected as frozen and non-frozen respectively.

Additional check using VIIRS data from 8 winters

A direct quantitative analysis is not possible, since no ground truth is available for 18 out of the 20 winters. In order to validate our MODIS results, we additionally process the VIIRS data from 8 winters (since winter 2012–13, inclusive) and compare the results to MODIS. Since a pixel-to-pixel comparison is not straight-forward due to different GSDs, we fit the timelines per winter for each lake as described before (Fig. 4.4, top row) and compute absolute differences (AD) between the daily estimates for the percentage of frozen pixels. The AD is computed only on dates when both MODIS and VIIRS acquisitions are present, and when the lake is at least 30% cloud-free. The ADs are then further aggregated to obtain a Mean Absolute Difference (MAD) for each winter. Fig. 4.8 (top row, left) shows, for each lake, the mean and standard deviation of the MAD across the 8 common winters. The low mean values (3.5, 5.8 and 4.3 percent respectively for Sihl, Sils and Silvaplana) show that our MODIS and VIIRS results are in good agreement, especially considering that a part of the MAD is due to the difference in GSD between MODIS (250m GSD) and VIIRS (≈375m GSD). Note also that, the acquisition times during the day (and hence the cloud masks) can differ; and that, although the absolute geolocation has been corrected for both sensors, errors up to 0.5 pixel can still remain (Aksakal, 2013) and affect the selection of clean pixels near the lake shore.

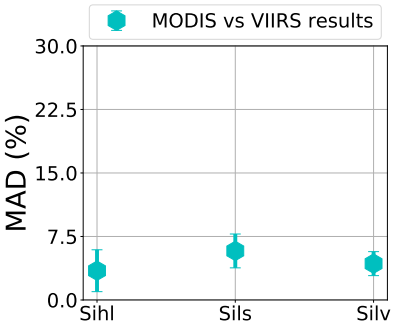
Comparison with MODIS and VIIRS snow products

We compare our MODIS (20 winters) and VIIRS (8 winters) results to the respective operational snow products: MSP (collection 6, MOD10A1), VSP (collection 1, VNP10A1F). For the regions of interest, VSP has some data gaps, hence the comparison is done whenever VSP data is available. For actual snow cover mapping, errors of 7-13% have been reported for MSP (Hall and Riggs, 2016). Our findings are in line with this: for the two 2016–17 and 2017–18 (non-transition days only) we observe an error of 14% w.r.t. our ground truth, see Fig. 4.8 (middle row).

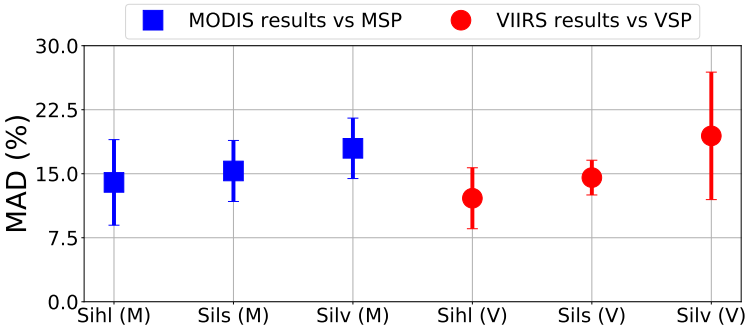
Similar to the evaluation methodology explained in Section 4.5.2, for each lake, we first estimate the percentage of frozen pixels per day using our MODIS and VIIRS results. Since a pixel-to-pixel registration is difficult in the presence of absolute geolocation shifts and/or GSD differences, the daily percentage of frozen pixels is also computed from the snow products and the MAD is estimated for each winter. See Fig. 4.8 (bottom row) for our MODIS results vs MSP comparison. For the three lakes, the per-winter MAD is shown on y-axis against the winters on x-axis. We again exclude lake St. Moritz because of its minuscule area.

Overall, the 20-year time series inter-comparison (per-lake mean and standard deviation of MAD, Fig. 4.8, top row, right) does not suggest large, systematic inconsistencies. On average, our MODIS and VIIRS results deviate by mean MAD values of 14-18% and 12-19% respectively. These deviations are only a little higher than the estimated error of the snow products, and relatively stable across different years.

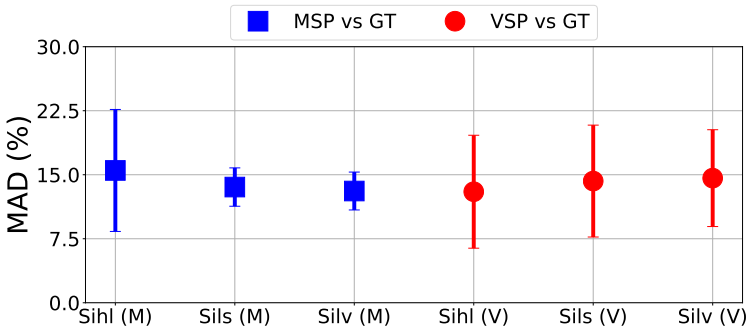
It is important to point out that the snow products are an imperfect proxy for lake ice, because a lake can be frozen but not snow-covered, especially near freeze-up when it has not yet snowed onto the ice. Also, mixed ice and water cases go undetected in



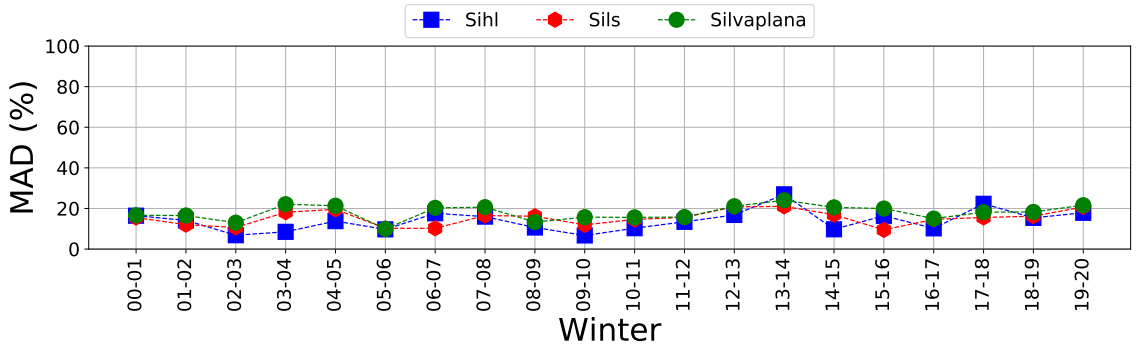
Our MODIS vs. VIIRS results



Our results vs respective snow products



Snow products vs. our ground truth (GT) for 2 winters.



Our MODIS results vs. MSP for 20 winters

Figure 4.8.: Row 1 shows the comparison of our MODIS and VIIRS results for the 8 common winters (left sub-figure) and comparison of our MODIS (20 winters) and VIIRS (8 winters) results with the respective snow products (right sub-figure). Row 2 shows the deviations between the two snow products and our webcam-based ground truth. Row 3 displays per-winter MAD (MODIS) for each lake.

MSP (Hall and Riggs, 2016). Fig. 4.8 (middle row) shows that the snow products are less consistent with the manually annotated ground truth than our ice maps. In fact, most deviations between our estimates and the snow products occur around the transition dates, mostly freeze-up. Additionally, MSP and VSP use a less conservative cloud mask than we do (accepting not only *confident clear* and *probably clear*, but also *uncertain clear* as cloud-free). Despite these issues, the inter-comparison provides a second check for our results. For completeness, we note that our algorithm has similar issue and thin

ice is sometimes confused with water: firstly, snow-free ice is rare and underrepresented in the training set. Secondly, it appears predominantly near the transition dates (especially freeze-up) when we do not have pixel-accurate ground truth. Thirdly, thin ice and water are difficult to distinguish, we observed that even human interpreters at times confused them when interpreting webcam images.

It is interesting to note that, for both sensors, the mean MAD is inversely proportional to the lake area (see Fig. 4.8, row 1, right). This hints at residual errors in the products' geolocation, which would affect smaller lakes more due to the larger fraction of pixels near the lake outline. Besides the < 0.5 pixel inaccuracy of our maps, inaccurate geolocation of the snow products has been reported (more for MODIS, less for VIIRS) especially for freshwater bodies, due to uncertainties in gridding, reprojection etc. (Hall and Riggs, 2016).

LIP trends using MODIS data

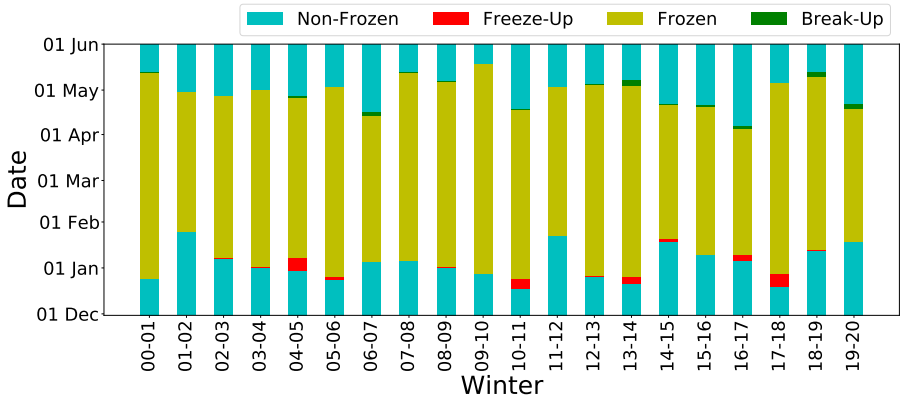
As discussed in Section 4.4.3, we fit "U with wings" polygon to each winter to estimate the four critical dates: FUS, FUE, BUS and BUE. Sometimes, these phenological dates are defined such that a second, consecutive day with similar ice conditions is required to confirm the event. We do not enforce this constraint, because, quite often, the days after a potential freeze-up or break-up date are cloudy, and looking further ahead runs the risk of pruning the correct candidates.

Using the estimated LIP dates from 20 winters, we plot their temporal evolution for lakes Sils (top) and Silvaplana (bottom) in Fig. 4.9. On the y-axis, all the dates from 1 December to 1 June (we skip September till November since no LIP events were detected during these months), while on the x-axis we show the winters in a chronological order. In each winter, the non-frozen, freeze-up, frozen and break-up periods are displayed in cyan, red, yellow and dark green colours, respectively.

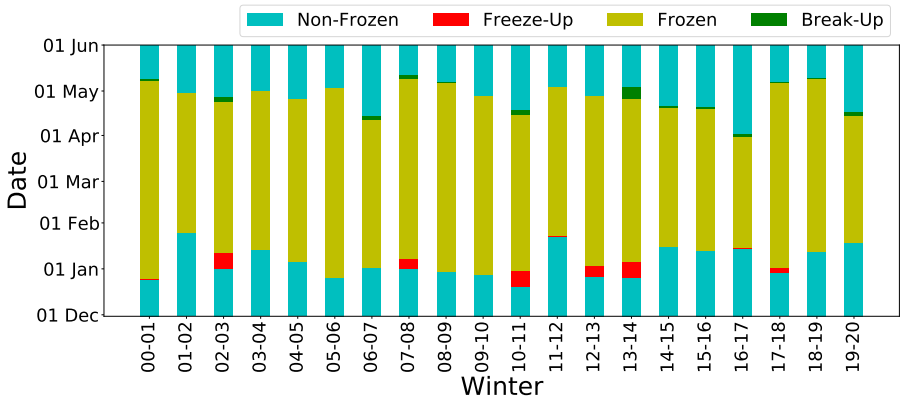
It can be seen from Fig. 4.9 that the freeze-thaw patterns of both lakes vary considerably across winters. For lake Sils (Silvaplana), on average, the FUS occurred on 3 January (5 January) followed by a freeze-up period of 3 (3) days until FUE on 6 January (8 January). Additionally, on average, the lake remained fully frozen (CFD) for 113 (108) days until BUS on 29 April (26 April) and the break-up period lasted 1 (1) day until BUE on 30 April (27 April). The average number of days from FUS to BUE (ICD) is 117 (112).

The Oberengadin region with lakes Sils and Silvaplana is a single valley (Fig. 4.1, Table B.2 in Appendix B) and hence the two have similar weather conditions. Silvaplana is relatively deeper but has smaller area than Sils, making them comparable in terms of volume, too. So similar LIP patterns can be expected. However, the clouds above the lakes (especially on partly cloudy days), and the associated cloud mask errors, can cause small differences. In winter 2016-17, the ice-on date of the two lakes, confirmed by visual interpretation of webcams, differ by 7 (low confidence) to 10 (medium confidence) days, see also Tom et al. (2020c).

In most of the winters, the LIP characteristics of these lakes derived using our ap-



Lake Sils



Lake Silvaplana

Figure 4.9.: 20 winter temporal LIP characteristics of lakes Sils and Silvaplana estimated from MODIS using SL classifier.

proach are in agreement, see Fig. 4.9. However, there are some outliers too (>10 days deviation). A notable outlier is the break-up period in winter 2009-10. For Sils (Silvaplana), BUS and BUE were both estimated as 19 May (28 April). This drift primarily happened because of a huge data gap due to clouds and cloud mask errors. During the period from 28 April till 20 May, Silvaplana had $>30\%$ cloud-free MODIS acquisitions only on 28 April, 29 April, 8 May and 20 May, and the lake was detected as non-frozen on all these dates. However, Sils had MODIS acquisitions on 28 April, 29 April, 5 May and 19 May. On 5 May the lake was detected as 100% frozen due to a false negative cloud mask, although break-up had started on the two earlier dates (75%, respectively 60% frozen) and the lake was ice-free on May 19. We also checked the Landsat-7 acquisitions on 20 April 2010 and 22 May 2010 and found that both lakes were fully covered by snow on the former date and fully non-frozen on the latter date. No cloud-free Landsat-7 data is available in between these two dates. For Sils, the actual BUS probably happened on 29 April ($>30\%$ non-frozen) and BUE soon after (likely on 30 April, since the BUE of Silvaplana was detected as 28 April and Sils was detected $<70\%$ non-frozen on 29 April). However, both dates went undetected until 19 May, because of the clouds in combination with the maximum allowed duration of 2 weeks for the break-up.

In winter 03-04, the freeze-up periods of Sils (FUS on 1 January, FUE on 2 January) and Silvaplana (FUS and FUE on 14 January) were also detected far apart, again due to

a data gap because of clouds. Sils was estimated 68% and 90% frozen on 1 and 2 January, respectively, so they were chosen as FUS and FUE. On lake Silvaplana, the sequence for 1-5 January was 4%→13%→0%→21%→0% frozen. Then 14 January and 21 January were both found 100% frozen, so the fitting chose 14 January as both FUS and FUE. No cloud-free MODIS data exist on the intermediate dates 6-13 January and 15-20 January, and we could also not find any cloud-free Landsat-7 images between 21 December 2003 and 29 January 2004 (both inclusive) to check, but could confirm 0% ice cover on 20 December and 100% cover on 30 January. Connecting all the dots, we speculate that FUS and FUE of Silvaplana occurred soon after 5 January.

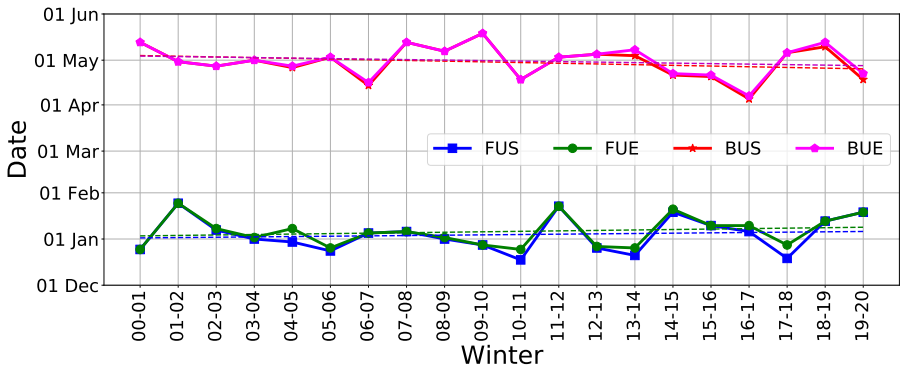
In winter 2013-14, our method asserts FUE of Sils on 26 December and of Silvaplana on 6 January. Between those dates there were a number of partially frozen dates, but with more ice cover for Sils than Silvaplana. Additionally, 2-5 January were cloudy, leading the fitting to chose the earlier date for the former, but the later one for the latter. We again checked with Landsat-7 that on 15 December both lakes were fully non-frozen, whereas on and 25 January both lakes were fully snow-covered. There exist no cloud-free Landsat-7 image in between these two dates to pin down the dates more accurately.

In some winters, there is almost no freeze-up and/or break-up period detected by our algorithm. This is partly a byproduct of the relatively loose threshold needed to estimate the initial candidates for our small lakes (see Section 4.4.3), bringing the start and end dates of the transition closer together; and also influenced by frequent cloud cover during the critical transition dates (often more than half of all days c.f. Section 4.3c). For instance, if a couple of adjacent dates are cloudy during break-up (and the real BUS occurred during one of these dates) and on the next non-cloudy day, the lake is estimated 70.1% non-frozen, then our fitting will choose this date as both BUS and BUE.

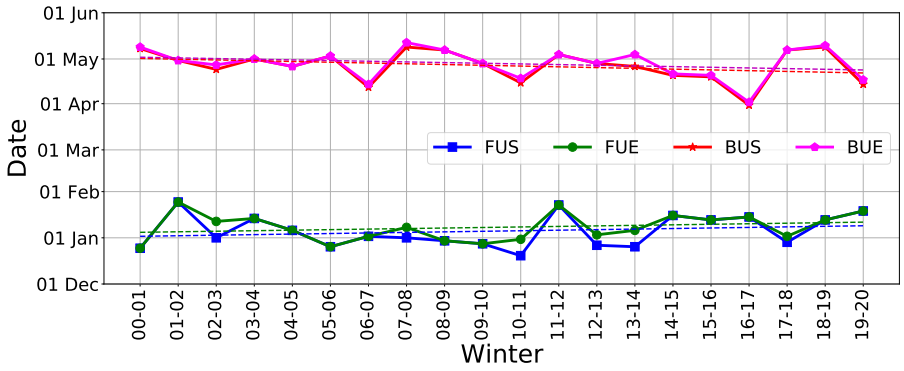
We go on to analyse the freeze-up (FUS, FUE) and break-up (BUS, BUE) patterns, by plotting time series of the four critical dates over the past 20 winters for lakes Sils and Silvaplana, see Fig. 4.10. Here also, we plot all the dates from 1 December till 1 June in each winter on the y-axis and the winters in chronological order on the x-axis. Additionally, per phenological date, we fit a linear trend. Progressively later freeze-up and earlier break-up is apparent for both lakes.

In each winter, we also derive the remaining LIP events (ICD, CFD) listed in Table 4.1. Their trends are shown in Fig. 4.11, with the duration in days on the y-axis and the winters in chronological order on the x-axis. Obviously, ICD and CFD are decreasing for both lakes.

Quantitative trend values that we estimated for all the LIP events of Sils and Silvaplana are shown in Table 4.6. As explained above, we correct obvious failures of the automatic analysis, and set the following corrections for lake Sils: BUS and BUE occurred on 29 April and 30 April respectively in winter 2009–10. Similarly, for Silvaplana: FUS and FUE occurred on 6 January in winter 2003–04 and FUE occurred on 26 December in winter 2013–14. For completeness we also fit trends without the correction – these differ only slightly and confirm that the corrections hardly impact the overall picture. The trend towards earlier break-up (BUS, BUE) is more pronounced than the one towards later freeze-up (FUS, FUE), for both Sils and Silvaplana. It is interesting to note that



Lake Sils



Lake Silvaplana

Figure 4.10.: 20 winter ice freeze-up (FUS, FUE) and break-up (BUS, BUE) trends for lakes Sils and Silvaplana from MODIS estimated using SL classifier.

the decrease in freeze duration (ICD and CFD) is stronger for the slightly smaller lake Silvaplana.

Table 4.6.: Estimated LIP trends (black). Results before manual correction of the automatic results are shown in grey.

Lake	FUS	FUE	BUS	BUE	ICD	CFD
Sils	0.23	0.31	-0.46/-0.47	-0.32/-0.34	-0.55/-0.57	-0.76/-0.78
Silvaplana	0.45/0.37	0.38/0.36	-0.51	-0.45	-0.9/-0.82	-0.89/-0.87

Correlation of LIP events with meteorological data

We have also studied the (centred and normalised) cross correlation $\in [-1, 1]$ between the LIP events (corrected version) and climate variables such as temperature, sunshine, precipitation and wind during the 20 winters. The results are shown in Fig. 4.12 for the lakes Sils (top) and Silvaplana (bottom). Air temperature (measured at 2m above ground) and precipitation data were collected from the nearest meteorological station SIA. However, we used the sunshine and wind measurements at station SAM, since these were not available for the complete 20 winters time span at SIA. We did not use the cloud information (number of non-cloudy pixels) from MODIS data as a measure of sunshine

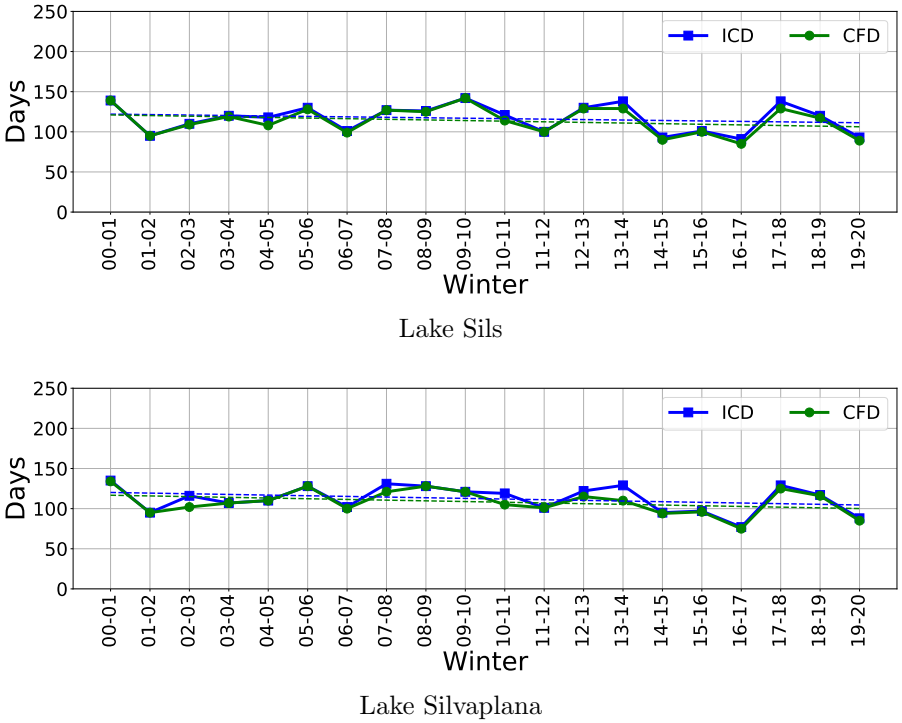


Figure 4.11.: 20 winter ice duration (ICD, CFD) trends for lake Sils and Silvaplana from MODIS estimated using SL classifier.

duration, since that would ignore the evolution throughout the day, and suffers from a non-negligible amount of cloud mask errors.

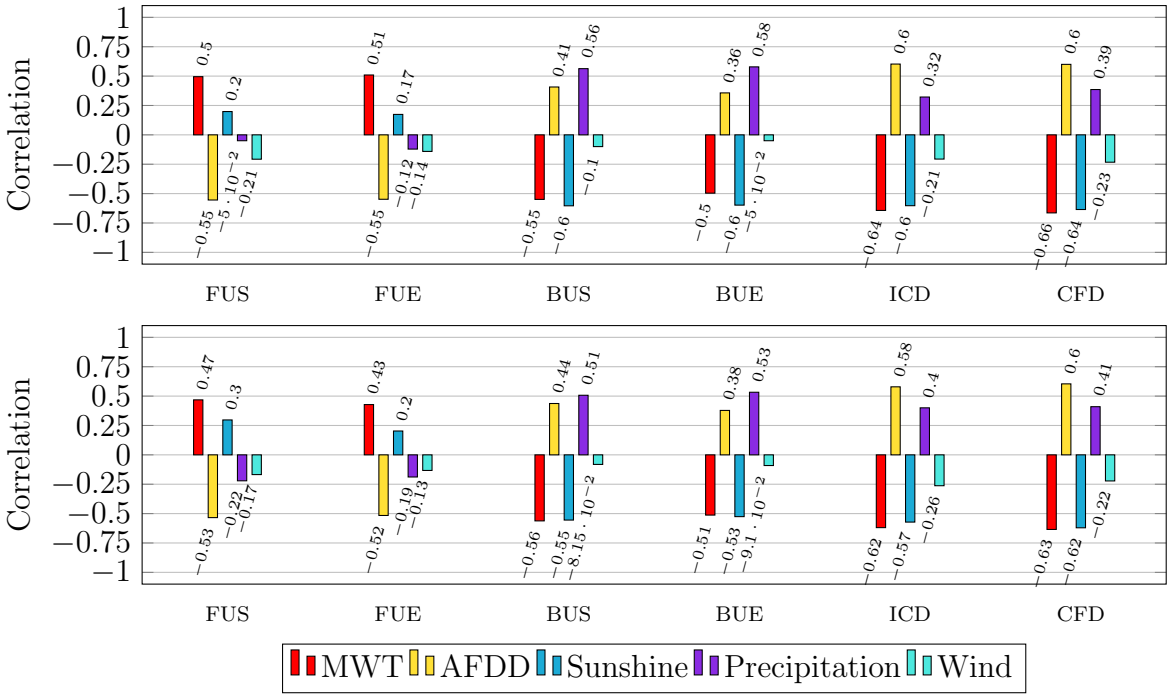


Figure 4.12.: Bar graphs showing the 20 winter correlation of the LIP events (Sils on top row, Silvaplana on bottom) with climate variables.

Mean Winter Temperature (MWT) corresponds to the air temperature (in °C) averaged over the whole winter season (September till May). As expected, Fig. 4.12 shows that MWT has strong negative correlation with the freeze durations CFD and ICD, negative correlation with the break-up events BUS and BUE, and positive correlation with the freeze-up events FUS and FUE. We conclude that, indeed, as winters got warmer over the past 20 years the lakes froze later and broke up earlier. For both lakes, the relationship of MWT with CFD is shown in Fig. 4.13 (row 1). Further significant correlations are displayed in Section B.5 (Appendix B).

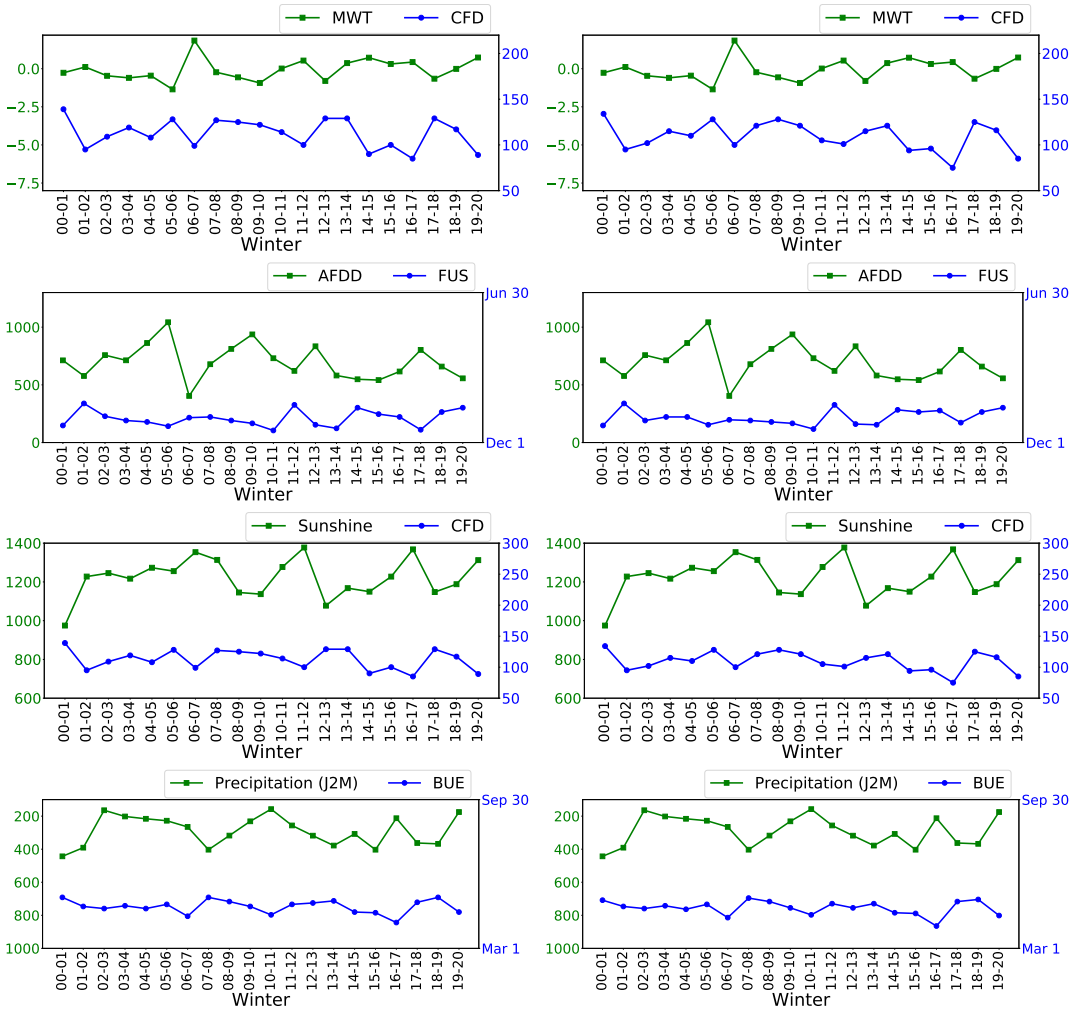


Figure 4.13.: Correlation of LIP events and weather variables: MWT (°C) and CFD (days) are shown in first row, AFDD (°C) and FUS in second row, total winter sunshine (hours) and CFD in third row, total precipitation (mm) in the months January to May (J2M) and BUE in last row. Results for lakes Sils and Silvaplana are displayed in left and right columns respectively.

AFDD represents the cumulative sum (of daily mean temperature) on the days with average air temperature below the freezing point (0°C) in a winter season. AFDD is a popular proxy for the thickness of ice cover (Beyene and Jain, 2018, Qi et al., 2020). For both Sils and Silvaplana, AFDD has strong positive correlations with ICD and CFD, strong negative correlation with the freeze-up events FUS and FUE, and moderate positive correlation with ice break-up events BUS and BUE, see Fig. 4.12, again indicating

that in colder winters (higher AFDD) the freeze-up occurs earlier and the break-up later, leading to longer freeze duration. The relatively weaker correlation for the break-up indicates that freeze-up play a larger role for that event. As an example, the correlation with FUS is shown in Fig. 4.13, for further plots see Section B.5 (Appendix B).

To study the effect of sunshine on LIP events, we correlate the total winter sunshine (hours) with the freeze length events ICD and CFD, total sunshine in the months September to December (S2D) with the freeze-up events FUS and FUE, and the total sunshine from January to May (J2M) with the break-up events BUS and BUE. Here, we assume that the sunshine in the months after freeze-up has no connection with freeze-up events. Similarly, we assume that the sunshine in the early winter months (September till December) does not affect the break-up events. We notice strong negative correlation of the total winter sunshine with ICD, CFD and break-up events. The more sunshine in the months near break-up, the earlier the ice/snow melts, which also reduces the total freeze duration. An example correlation with CFD is visualised in Fig. 4.13, further significant correlations are displayed in Section B.5 (Appendix B).

We also check the relationship between the LIP events and the total precipitation (snow and rain) during the winter months. Similar to sunshine analysis, we correlate the total precipitation during the months from September till December, January till May and September till May to the freeze-up (FUS, FUE), break-up (BUS, BUE) and freeze duration events (CFD, ICD) respectively, see Fig. 4.12. Notable are the break-up events with good positive correlation. More precipitation in the months January to May (likely to be predominantly snow), favours later break-up, and vice-versa. Correlation with BUE is shown in Fig. 4.13 and with BUS in Section B.5 (Appendix B).

Finally, inspired by Gou et al. (2015), we also looked at the effect of wind on the LIP events, which may also influence lake freezing. We correlated the mean winter wind speed (km/h) with CFD and ICD, mean wind speed from September to December with FUE and FUS, and mean wind speed from January to May with BUS and BUE. However, we did not find any significant correlations, see Fig. 4.12.

4.5.3. Discussion

In any ML-based system, the variety in the training dataset has a critical influence on the model being learnt. Our dataset consists of small lakes and has significant class imbalance. This is a biased, but realistic scenario, representative of mountain lakes in sub-Arctic and temperate climate zones. For supervised classification, proper ground truth information is available only from the winters 2016–17 and 2017–18. Using the data from these two winters, we first performed thorough experimentation with different classifiers. Additionally, we did an inter-comparison of the individual performance of these ML classifiers. For the four-fold cross validation experiment, the highest performance on MODIS and VIIRS data were reported for SR and XGB classifiers respectively, see Table 4.3. However, SR, RF and XGB suffered a significant generalisation loss compared to the SL counterpart. I.e., non-linear classifiers tended to overfit on our relatively small dataset. We emphasise that it does not contradict the findings of Wu et al. (2021b) for a dataset of larger lakes with many more training pixels. There, RF and GBT performed

very well. We empirically selected SL as the most suitable classifier for our lakes of interest, and we recommend to repeat the empirical exercise when moving to different geographical conditions. Our MODIS and VIIRS results validate each other in a relative sense (< 5.6 % MAD in the worst case, lake Sils), but could be subject to a common bias. In the absence of ground truth there is no way to assess our absolute accuracy, but as an external check against a methodologically different mapping scheme we inter-compared our results against the respective operational snow products. The deviations were in the expected range (MAD < 20 %).

Besides its advantages, like dense coverage, regular time series and homogeneous observation conditions, satellite data also has disadvantages. The trade-off between spatial and temporal resolution makes it difficult to monitor smaller lakes – with 21 MODIS (9 VIIRS pixels) for lake Silvaplana and only 4 MODIS pixels for lake St. Moritz, our study goes to the limit in that respect. A further, often-named obstacle for optical satellite observation are occlusions due to clouds, which significantly reduce the effective temporal resolution and also cause irregular gaps in the time series. These unpredictable data gaps are particularly troublesome for ice phenology, because the critical events occur over a short time and at times of the year where cloud cover is frequent in sub-Arctic and mid latitudes.

Data gaps due to clouds are the main source of error in our LIP estimation, besides cloud mask errors, confusion between water and thin/floating snow-free ice, and quantisation effects around hard thresholds. This makes phenological observations challenging – in particular the uncertainties of our predictions are largest during freeze-up, because of the frequent, but short-lived presence of snow-free ice. Still, it appears that our classifier copes better with the reflectance of ice than simple index-based snow products, and it is likely that more training data and, if available, additional spectral bands could further improve its detection.

As said before, a limiting factor for small lakes is the GSD, as decision based on very few pixels become overly unreliable and prone to statistical fluctuations, and even small geolocalisation errors have a large effect. Our work is also on the challenging ends of the spectrum in terms of local weather conditions: in a drier climate the observations would be less affected by cloud cover (we process lakes with as little as 30% cloud-free area to obtain sufficient temporal coverage), and fewer clouds also means fewer cloud-mask errors.

4.6. Conclusion

In this paper, we have reported results from a case study in Southeastern Switzerland, where we have retrieved lake ice phenology based on MODIS optical image time series. On the one hand, we have tried to push the limits of MODIS data for the analysis of small-to-medium-sized lakes, and have shown that even for such high-Alpine lakes it is possible to derive meaningful correlations between the 20-winter lake ice phenological trends and climate data. On the other hand, for the tested lakes, we have confirmed that a dedicated machine learning scheme maps lake ice more accurately than the classical

index- and threshold-based approaches.

As expected, our results point towards later freeze-up (FUS at a rate of 0.23 d/a for lake Sils, respectively 0.45 d/a for Silvaplana and FUE at a rate of 0.31 d/a for lake Sils, respectively 0.38 d/a for Silvaplana), earlier break-up (BUS: -0.46 d/a for lake Sils, respectively -0.51 d/a for Silvaplana and BUE: -0.32 d/a for lake Sils, respectively -0.45 d/a for Silvaplana) and decreasing freeze duration (ICD: -0.55 d/a for lake Sils, respectively -0.9 d/a for Silvaplana and CFD: -0.76 d/a for lake Sils, respectively -0.89 d/a for Silvaplana). We also observed significant (but not surprising) correlations with climate indicators such as temperature, sunshine and precipitation.

Our machine learning approach is generic and easy to apply to other sensors beyond MODIS and VIIRS (given training data). Importantly, the VIIRS sensor is projected to ensure continuity well into the future, opening up the possibility to establish an even longer time series. One solution for the cloud issues of optical satellites is to complement/replace them with radar observations, e.g., Sentinel-1 SAR. We have done preliminary research in this direction (Tom et al., 2020a). SAR-optical data fusion holds great promise, particularly in view of the GCOS requirement to monitor lake ice at daily temporal resolution (with an accuracy of ± 2 days).

In the future, it will be interesting to cross-check and validate existing temperature-based models like Hendricks Franssen and Scherrer (2008) against our data-driven results. Beyond a simple inter-comparison, temperature measurements could be used to eliminate gross prediction errors, and to bridge temporal gaps in optical satellite image-based predictions that occur due to clouds or fog.

We expect that machine learning-based ice detection itself could still be further improved with additional training data. I.e., pixel-accurate annotations during transition dates, as well as for more winters and a wider variety of lakes. Unfortunately, gathering such data is not only a considerable, tedious effort, but also poses its own challenges. In most locations and for older data, no corresponding webcam data (or similar regular photography) is available; even when available, its coverage is almost invariably incomplete; and even with usable webcam and satellite imagery, manual annotation is not trivial and prone to mistakes exactly in the situations that are most critical also for computational analysis (such as thin, black ice). We speculate that, given the enormous archive of unlabelled satellite data, approaches such as unsupervised, semi-supervised or active learning may be applicable and could improve the detector.

Yet another interesting research direction is to close the gap between knowledge-driven, model-based top-down models of lake ice formation and data-driven, bottom-up machine learning. Introducing expert knowledge about ice formation and associated physical constraints into machine learning models could also reduce the need for training data, and get the best of both worlds in terms of accuracy as well as interpretability of the model (Camps-Valls et al., 2018, 2020). How to best bridge the gap between statistical machine learning models and physical process models is an open question and an active research direction in the Earth sciences and beyond.

4.7. Funding

This work is part of the project *Integrated lake ice monitoring and generation of sustainable, reliable, long time series* funded by the Swiss Federal Office of Meteorology and Climatology MeteoSwiss in the framework of GCOS Switzerland.

4.8. Acknowledgements

We are grateful to Damien Bouffard (EAWAG, Switzerland) for providing advice regarding the correlation of meteorological variables and LIP events.

5 Lake Ice Detection from Sentinel-1 SAR with Deep Learning

Manu Tom, Roberto Aguilar, Pascal Imhof, Silvan Leinss, Emmanuel Baltsavias, Konrad Schindler
ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3-2020, 409–416, 2020
(Author version; for typeset version please refer to the original conference paper)

Abstract

Lake ice, as part of the Essential Climate Variable (ECV) *lakes*, is an important indicator to monitor climate change and global warming. The spatio-temporal extent of lake ice cover, along with the timings of key phenological events such as *freeze-up* and *break-up*, provide important cues about the local and global climate. We present a lake ice monitoring system based on the automatic analysis of Sentinel-1 Synthetic Aperture Radar (SAR) data with a deep neural network. In previous studies that used optical satellite imagery for lake ice monitoring, frequent cloud cover was a main limiting factor, which we overcome thanks to the ability of microwave sensors to penetrate clouds and observe the lakes regardless of the weather and illumination conditions. We cast ice detection as a two class (*frozen, non-frozen*) semantic segmentation problem and solve it using a state-of-the-art deep convolutional network (CNN). We report results on two winters (2016-17 and 2017-18) and three alpine lakes in Switzerland. The proposed model reaches mean Intersection-over-Union (mIoU) scores >90% on average, and >84% even for the most difficult lake. Additionally, we perform cross-validation tests and show that our algorithm generalises well across unseen lakes and winters.

5.1. Introduction

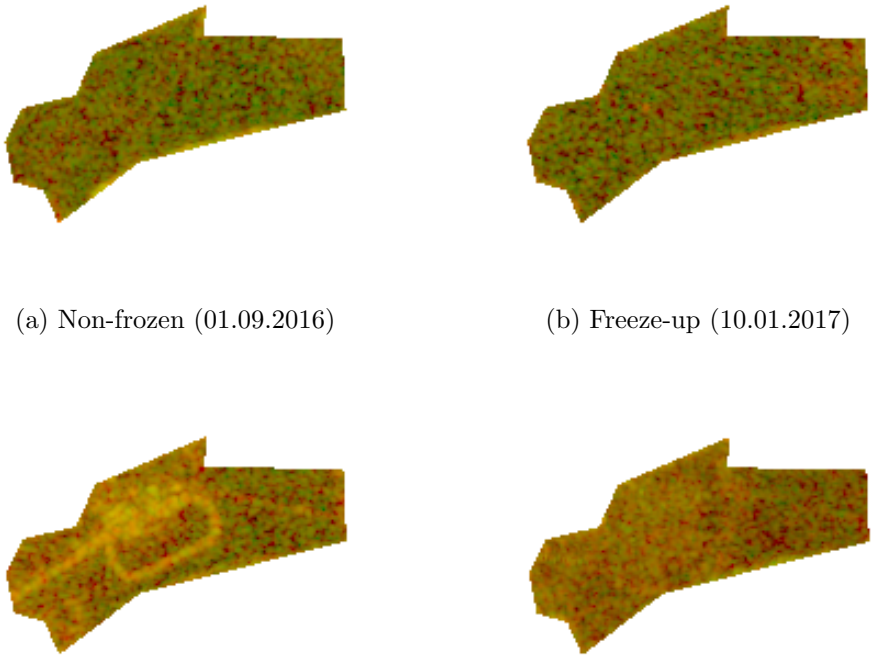
Climate change is one of the main challenges humanity is facing today, calling for new methods to quantify and monitor the rapid change in global and local climatic conditions. Various lake observables are related to those conditions and provide an opportunity for long-term monitoring, among them the duration and extent of lake ice. Remote sensing of lake ice also fits well with the Climate Change Initiative (CCI+, 2017) by the European

Space Agency (ESA), where *lakes* and *lake ice* were newly included. Additionally, CCI+ promotes long-term trend studies and climate studies, as recognised by the Global Climate Observing System (GCOS). Furthermore, lake ice influences various economic and social activities, such as winter sports and tourism, hydroelectric power, fishing, transportation, and public safety (e.g., winter and spring flooding due to ice jams). In addition, its impact on the regional environment and ecological systems is significant, which further underlines the need for detailed monitoring.

Satellites are a secure source for remote sensing of the cryosphere and for sustainable, reliable, and long term trend analysis. Additionally, satellite images are currently the only means to monitor large regions systematically and with short update cycles. This increasing importance of satellite observations has also been recognised by the GCOS. Recently, Tom et al. (2018) proposed a machine learning-based semantic segmentation approach for lake ice detection using low spatial-resolution (250m-1000m) optical satellite data (MODIS and VIIRS). Although the nominal temporal resolution of those sensors is very good (daily coverage), the main drawback of this methodology is frequent data loss due to clouds, which reduces the effective temporal resolution. This is critical, since important phenological variables depend on frequent and reliable observation. In particular, the *ice-on* date is defined as the first day when the lake surface is (almost) completely frozen and remains frozen on the next day, and *ice-off* is defined symmetrically as the first day where a significant amount of the surface is liquid water, and remains in that state for another day (Hendricks Franssen and Scherrer, 2008). The GCOS accuracy requirement for these two dates is ± 2 days. Systems based on optical satellite data will fail to determine these key events if they coincide with a cloudy period. Moreover, low spatial resolution of MODIS and VIIRS is also a bottleneck for spatially explicit ice mapping. Higher resolution optical sensors like Landsat-8 or Sentinel-2 do not provide a solution, due to their low temporal resolution and susceptibility to clouds. On the contrary, Sentinel-1 represents a favourable trade-off between spatial and temporal resolution. Additionally, Radar is unaffected by clouds, which in many regions is a considerable advantage.

Here we propose to use Sentinel-1 SAR data, which mostly meets the requirements of lake ice monitoring, and additionally comes for free and with a commitment to ensure continuity of the observations. Its spatial and temporal resolution (GSD ca. 10 m / revisit period if 1.5 days) make it possible to derive high-resolution ice maps almost on a daily basis. For completeness, we mention that, taking into account estimation uncertainty, the temporal resolution of Sentinel-1 falls just short of the 1-day temporal resolution requirement of lake ice monitoring, it can still provide an excellent "observation backbone" for an operational system that could fill the gaps with optical satellite data (Tom et al., 2018) or webcams (Xiao et al., 2018).

Converting a Sentinel-1 image to a lake ice map boils down to 2-class semantic segmentation, i.e., assigning each lake pixel to one of two classes, *frozen* or *non-frozen*. We do this with the *Deeplab v3+* semantic segmentation network (Chen et al., 2018b). Examples of Sentinel-1 SAR composites over the lake St. Moritz is visualised in Fig. 5.1, showing the *VV* amplitude in the red channel, and the *VH* amplitude in the green channel. The examples include the states *non-frozen* (01.09.2016, water), *freeze-up* (10.01.2017), *frozen* (08.02.2017, snow on top of ice) and a *break-up* date (23.03.2017).



(a) Non-frozen (01.09.2016)

(b) Freeze-up (10.01.2017)

(c) Snow with skate tracks (08.02.2017)

(d) Break-up (23.03.2017)

Figure 5.1.: Examples of RGB composites of Sentinel-1 SAR data (RGB = [*VV, VH, 0*]) of lake St. Moritz showing the lake in the four different states specified in the sub-captions.

Target lakes and winters. We analyse three selected lakes in Switzerland (*Sils, Silvaplana, St. Moritz*, see Table 5.1) over the period of two winters (2016-17 and 2017-18). These three lakes are located close to each other in the same geographic region, referred to as *Region Sils*. The lakes are comparatively small and situated in an Alpine environment, and they are known to reliably freeze over completely every year during the winter months. For the two winters 2016-17 and 2017-18, all available images were collected for the nine months between September 1 and May 31. After back-projecting the digitised lake outline polygons from Open Street Map (OSM) on to the SAR images, for each lake, we extract the *lake pixels* which lie inside the lake outline. In low-spatial resolution satellite images such as *MODIS* and *VIIRS*, only few such *lake pixels* are available (Tom et al., 2018), which makes the analysis of very small lakes such as St. Moritz difficult or even impossible. Thanks to the higher spatial resolution, the Sentinel-1 time series provides us with millions of lake pixels, which makes it possible to train powerful deep learning models for segmentation, which are extremely data-hungry.

Contributions. We address the problem of lake ice detection from Sentinel-1 SAR data, as an alternative to optical satellite data which is impaired by clouds. In the process, we show that a deep learning model pre-trained on an optical RGB dataset can nevertheless be re-used successfully as initialisation for fine-tuning to Radar data. To our knowledge, our work is the first one that utilises Radar data and deep learning for lake

ice detection, though it has been used for sea ice analysis.

Table 5.1.: Characteristics of the target lakes. Altitude (L) and altitude (S) denote the altitudes of the lake and nearest meteo station respectively. The distance to station is also shown.

	Sils	**Silvaplana**	**St. Moritz**
Area (km^2)	4.1	2.7	0.78
Altitude (L) (m)	1797	1791	1768
Max. depth (m)	71	77	42
Meteo station	Segl Maria	Segl Maria	Samedan
Dist. to lake (km)	0.5	1	5
Altitude (S) (m)	1804	1804	1709

5.2. Related work

Many studies discussed the trends in lake ice formation in different parts of the globe. Duguay et al. (2006) presented the trends in lake freeze-up and break-up across Canada for a long period from 1951 until 2000. Later, Hendricks Franssen and Scherrer (2008) reported the decreasing tendency in lake freezing in several Swiss lakes. Then, Brown and Duguay (2010) reviewed the response and role of ice cover in lake-climate interactions. This paper observed that the ability to accurately monitor lake ice will be an important step in the improvement of global circulation models, regional and global climate models and numerical weather forecasting. Brown and Duguay (2011) used the Canadian Lake Ice Model (CLIMo) to simulate lake ice phenology across the North American Arctic from 1961–2100, using two climate scenarios produced by the Canadian Regional Climate Model (CRCM). They projected changes to the ice cover using 30-year mean data between 1961–1990 and 2041–2070, which suggested a probable drift in freeze-up (0–15 days later) and break-up (10–25 days earlier). Duguay et al. (2015) presented an overview of the progress of remote sensing for lake and river ice. For lakes, that work reviewed a number of topics, including ice cover concentration, ice extent and phenology, and ice types, as well as ice thickness, snow on ice, snow/ice surface temperature, and grounded and floating ice cover on shallow Arctic and sub-Arctic lakes.

Lake ice monitoring using Radar data. Duguay and Lafleur (2003) proposed to determine the depth and thickness of ice in shallow lakes and ponds using the Landsat Thematic Mapper and European Remote Sensing (ERS)-1 SAR data. Almost a decade later, Surdu et al. (2014) conducted a study of the shallow lakes on the north slope of Alaska to find the response of ice cover on the climate conditions using Radar remote sensing and numerical analysis. A machine learning-based automated ice-vs-water classification was proposed by Leigh et al. (2014) using dual polarisation SAR imagery. Later, Surdu et al. (2015) performed a study on the ice freezing and thawing detection in shallow lakes from Northern Alaska with spaceborne SAR imagery. Antonova et al. (2016) monitored ice phenology in lakes of the Lena river delta using TerraSAR-X backscatter. Du et al. (2017) performed an assessment of lake ice phenology in the Northern

Hemisphere from 2002 to 2015. Pointner et al. (2018) studied the effect of the lake size on the accuracy of a threshold-based classification of ground-fast and floating lake ice from Sentinel-1 SAR data. Duguay and Wang (2019) presented various algorithms such as thresholding, Iterative Region Growing with Semantics (IRGS) and k-means for the generation of a floating lake ice product from Sentinel-1 SAR data for various permafrost regions (Alaska, Canada and Russia). Geldsetzer et al. (2010) used RADARSAT-2 SAR data to monitor ice cover in lakes during the spring melt period in the Yukon area of the Canadian Arctic. They put forward a threshold-based classification methodology and observed that the *HH* and *HV* backscatter from the lake ice have significant temporal variability and inter-lake diversity. Murfitt et al. (2018) used the RADARSAT-2 imagery to develop a threshold-based method to determine lake phenology events for the mid-latitude lakes in Central Ontario from 2008 to 2017. Wang et al. (2018a) also used RADARSAT-2 imagery (dual polarised) for developing a lake ice classification system acquired over lake Erie, with the IRGS method. Additionally, Gunn et al. (2018) used the polarimetric RADARSAT-2 (C-Band) to observe the scattering mechanisms of bubbled freshwater lake ice. SAR data analysis is challenging, and deep learning could play a significant role because of its ability to learn task-specific, hierarchical image features.

Lake ice monitoring with webcams. Xiao et al. (2018) described a system that detects lake ice in webcam data with the help of a deep neural network. Public webcams have two main advantages compared to optical satellite images. Firstly, they are usually not affected by clouds, except for the comparatively rare case of dense fog. Secondly, they have a very high temporal resolution (up to one image per 10 min). Although the approach generated excellent results, it also has disadvantages. Webcams are usually placed arbitrarily (e.g., too far away or covering a small lake area), and often only low above the lake, leading to great scale differences between front and back of the lake surface. Moreover, they are prone to hardware failure, and, being very cheap cameras, they have poor spectral and radiometric quality with significant compression artifacts. Another practical problem with webcams is that it is difficult to operationalise them at country- or even world-scale.

Optical data for lake ice monitoring. Tom et al. (2018) proposed a machine learning-based methodology for lake ice detection using low resolution optical satellite images. The main problem with optical satellite images is the data loss due to clouds. However, the authors showed that the algorithm produces consistent results when tested on data from multiple winters. In addition, Barbieux et al. (2018) used Landsat-8 multispectral data for extraction of frozen lakes and water-vs-ice classification. Recently, Tom et al. (2019) put forward a feasibility study, which targeted for a unified lake ice monitoring system that combines images from optical satellites, in-situ temperature data and webcam images.

5.3. Data

Sentinel-1 SAR consists of two identical satellites (*S1A* and *S1B*) operational in space with 180° phase shift, following a sun-synchronous, near-polar orbit. The two satellites orbit the Earth at an altitude of 693 km and have a repeat cycle of 12 days at the equator

(effectively 6 days with *S1A* and *S1B*). The same point on Earth is mapped several times within one repeat cycle. Due to the large across-track area coverage of the satellites and the latitude of our target area in Switzerland (and most other areas where lakes freeze), the revisit time is further reduced. For Region Sils, it can bee seen from Table 5.2 that the temporal resolution in winter 2017-18 is better than that of 2016-17. This is because of missing data from *S1B*. Though *S1B* was launched in April 2016, the corresponding data is fully available in the Google Earth Engine (GEE) platform (see Section "Data Collection" below) only from March 2017. In addition, Sentinel-1 covers the Region Sils in four orbits. See Table 5.3 for the details. Footprints of the four orbits are shown in Fig. 5.2.

Table 5.2.: Dataset statistics. Non-transition days, on which a lake is almost fully frozen / non-frozen, and transition days (partially frozen) are shown. Lake pixels are those which lie completely inside the lake polygon. # acq. denotes the number of acquisitions.

Lake	Winter	Non-transition days		Transition days	Total # acq.	Temporal resolution (days)	# lake pixels
		Non-frozen	Frozen				
Sils	2016-17	40	42	37	119	2.3	40908
	2017-18	76	65	40	181	1.5	
Silvaplana	2016-17	36	44	39	119	2.3	26563
	2017-18	85	66	30	181	1.5	
St. Moritz	2016-17	66	42	11	119	2.3	7521
	2017-18	84	77	20	181	1.5	

Table 5.3.: Details of four orbits scanning *Region Sils* such as orbit number, flight direction, acquisition time in Universal Coordinate Time (UTC), and approximate incidence angle.

Orbit	Flight dir.	Acquisition time	Incidence angle
15	ascending	17:15	41.0°
66	descending	05:35	32.3°
117	ascending	17:06	30.8°
168	descending	05:26	41.7°

The *S1A* and *S1B* satellites have the same SAR system on board which sends out frequency-modulated electromagnetic waves (C-frequency band) and detects the backscattered echoes of the surface. From the reflected signal, the SAR sensor measures the amplitude and phase. In our research, we use only the amplitude information. We work with the Level-1 Ground Range Detected (GRD) product in Interferometric Wide (IW) swath mode. That product has no phase information anymore, and has a nearly square footprint (10m×10m per pixel). It also has good temporal resolution (see Table 5.2) and four polarisations (*VV, VH, HH, HV*). However, for the Region Sils, data is acquired only in *VV* and *VH* modes. The distribution of backscatter values of *frozen* and *non-frozen* pixels in these bands are shown in Fig. 5.3. Note that the separability in *VV* appears better than in *VH*.

The radar backscatter is influenced in a complex manner by a variety of factors, which can be grouped into two main categories. Firstly, the *sensor parameters* such as

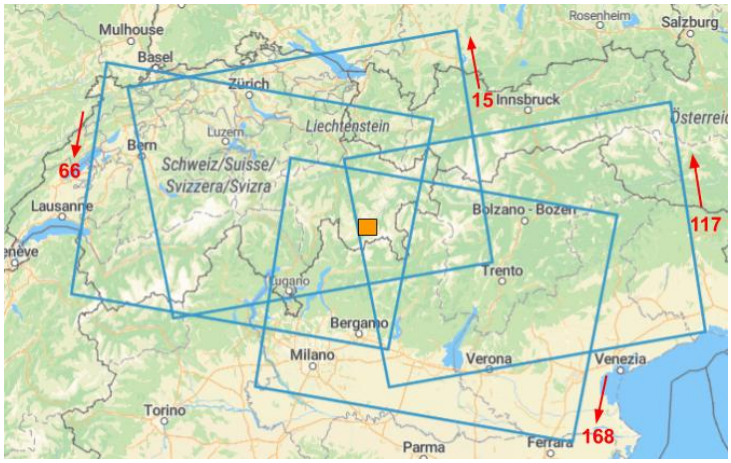
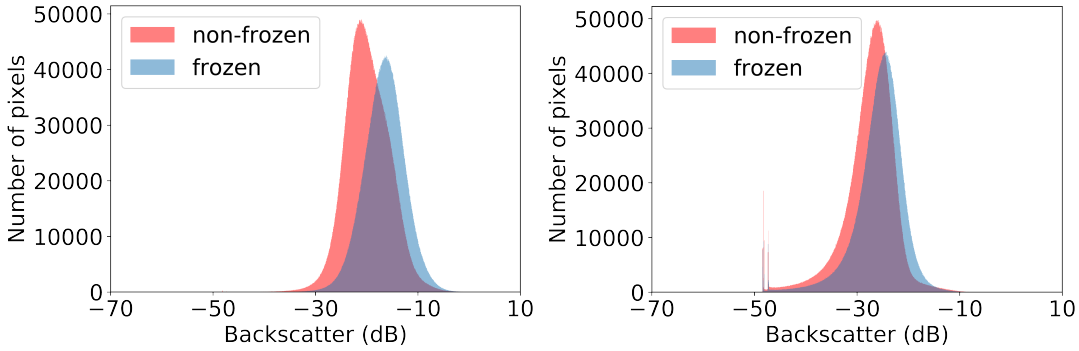


Figure 5.2.: The four Sentinel-1 orbits (15, 66, 17, 168) with the corresponding directions (*ascending* or *descending*) which covers the *Region Sils* (shown as orange filled rectangle).



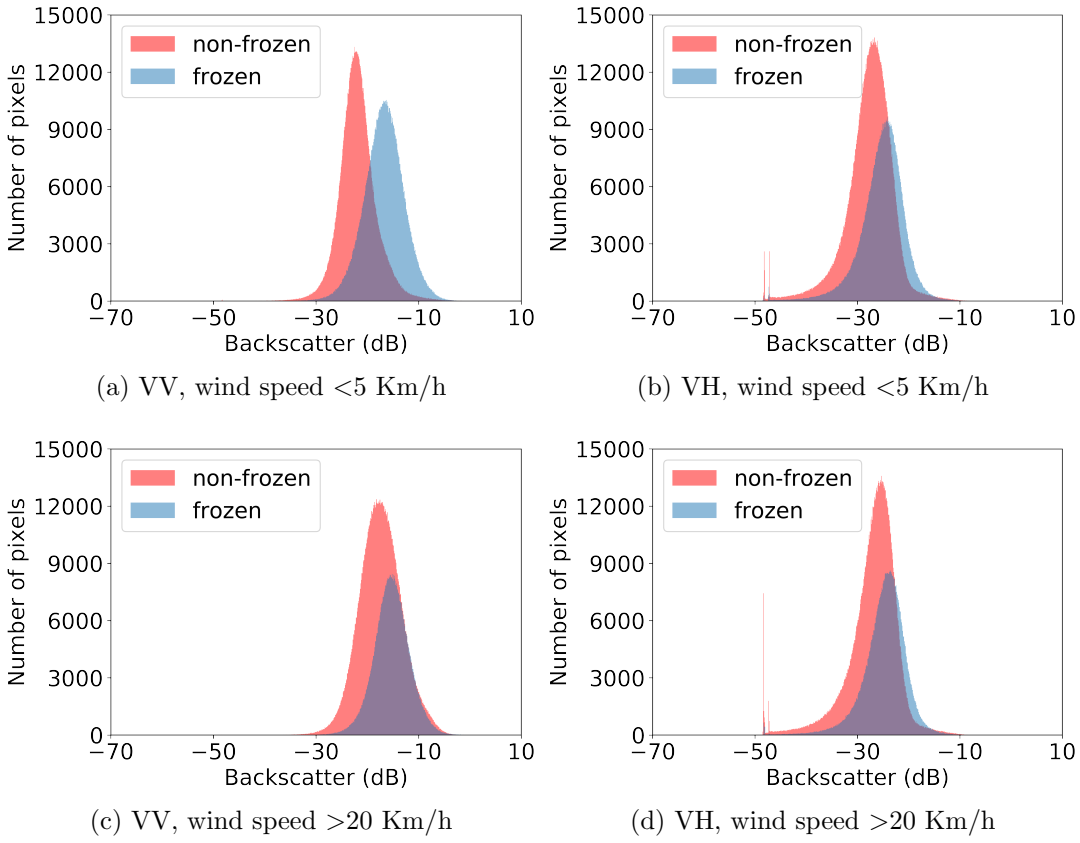
(a) Distribution of VV

(b) Distribution of VH

Figure 5.3.: Distribution of frozen and non-frozen pixels for VV and VH polarisations (combined data from 3 lakes, 2 winters). Best if viewed on screen.

wavelength (5.54 cm), incident angle (20° to 46°) and polarisation. Secondly, it depends on *surface parameters* which can be either geometrical factors such as roughness, landscape topography, etc. or physical factors such as the permittivity of the surface material. Significant factors for lakes are also wind speed and direction, and the water content in snow. For smooth and plain water, almost all radiation is scattered away from the sensor making it appear very dark. As the wind speed picks up, waves occur on the water surface and significant scattering can occur. When perfectly plain water is covered by perfectly plain ice, microwaves penetrate the ice without absorption and are reflected at the ice-water interface away from the sensor, and the ice covered lake would in theory appear completely black. In reality, cracks in the ice scatter some microwaves back to the sensor. Therefore, visible and well located cracks are clear indicators for ice cover. Furthermore, in reality, the ice-water interface is never completely smooth, therefore some scattering can occur at these boundaries which, however, can be weak. The older the ice, the more air bubbles are enclosed in it, which increase the backscatter within the ice volume by direct backscattering and also by double reflection of microwaves at the air-bubbles and the ice-water interface. With snow cover, the air-ice interface becomes increasingly rough

which further increases the backscatter signal. Finally, with snow melt, the liquid water content of the overlying snow pack increases and significantly reduces the backscatter signal, as the snow-water mixture of wet snow absorbs a significant fraction of the microwave energy.



(a) VV, wind speed <5 Km/h

(b) VH, wind speed <5 Km/h

(c) VV, wind speed >20 Km/h

(d) VH, wind speed >20 Km/h

Figure 5.4.: Distribution of frozen and non-frozen pixels for VV and VH polarisations in different wind speed conditions (combined data from 3 lakes, 2 winters). Best if viewed on screen.

Data collection and pre-processing. GEE is a cloud-based platform for large-scale geo-spatial data analysis (Gorelick et al., 2017). It stores and provides data of various satellite missions, performs data pre-processing and makes them freely available for education and research purposes. The Sentinel-1 backscatter coefficients (in decibels) were downloaded from the GEE platform after several inbuilt pre-processing steps such as GRD *border noise removal* which corrects the noise at the border of the images, *thermal noise removal* for correcting the thermal noise between the sub-swaths, *radiometric calibration* which calculates the backscatter intensity using the GRD metadata, *terrain correction* to correct the side looking effects using the digital elevation model (SRTM, 30m), and *log-scaling* to transform the approximate distribution of the SAR responses from *Chi-squared* to *Gaussian* (see Fig. 5.3). Note that we did not perform any absolute geolocation correction, since the back-projected lake outlines suggested a sufficient accuracy.

Transition and non-transition days. All the data from two winters was divided into two categories: *non-transition* dates where the lake is almost fully frozen or fully

non-frozen, and *transition* dates with partially frozen lake surface. Both freeze-up and break-up dates belong to the transition category. The dataset statistics are shown in Table 5.2. Note that, since lake St. Moritz is relatively small in area and volume, it freezes and melts faster and has fewer transition days.

Ground truth. For each lake, one label (*fully frozen / non-frozen*, *partially frozen / non-frozen*) per day was assigned by a human operator after visual interpretation of the visible part of the lake from freely available webcam data. The ground truth thus generated was further enriched by visual interpretation of the Sentinel-2 images whenever available. However, some remaining noise in the ground truth is likely due to interpretation errors, as a result of overly oblique viewing angles of webcams and compression artefacts in the images. During the transition days, ground truth estimation was very difficult since we had partially frozen and non-frozen states and there was a difficulty to discriminate transparent ice and water. Thus, transition days were not used for quantitative analysis.

5.4. Methodology

Semantic segmentation. We define lake ice detection as a two class (*frozen, non-frozen*) pixel-wise classification problem and tackle it with the state-of-the-art semantic segmentation network *Deeplab v3+* (Chen et al., 2018b). The *non-frozen* class comprises of only *water* pixels. Whereas a pixel is considered to be part of the *frozen* class if it is either *ice* or *snow*, since in the target region, the frozen lakes are covered by snow for much of the winter. The standard procedures in machine learning-based data analysis are followed. The dataset is first divided into mutually exclusive *training* and *test* sets. The *Deeplab v3+* model is then fitted on the training set. Lastly, the trained model is tested on the previously unseen test dataset.

Deeplab v3+ (Chen et al., 2018b) is a deep neural network for semantic segmentation, which has set the state-of-the-art in multiple benchmarks, including among others the PASCAL VOC 2012 dataset (Everingham et al., 2015). It combines the advantages of both Atrous Spatial Pyramid Pooling (ASPP) and encoder-decoder structure. Atrous convolution allows to explicitly control the resolution of the features computed by the convolutional feature extractor. Moreover, it adjusts the field-of-view of the filters in order to capture multi-scale information. *Deeplab v3+* also incorporates depthwise separable convolution (per-channel 2D convolution followed by pointwise 1×1 convolution) which significantly reduces the model size. The network architecture is shown in Fig. 5.5.
Network parameters. We used the *mobilenetv2* implementation of *Deeplab v3+*, as available in *TensorFlow*. The *train crop size* was set to 129×129 (effective patch size is 128×128) and the *eval crop size* to the full image resolution. All models were trained for $40'000$ iterations with a batch size of 8. Atrous rates were set to $[1, 2, 3]$ for all experiments. The cross-entropy loss function was minimised with standard stochastic gradient descent, with a base learning rate of 10^{-3}.

Transfer learning. Deep supervised classification approaches need lots of labelled data and a large amount of resources to train a model from scratch. Such data volumes

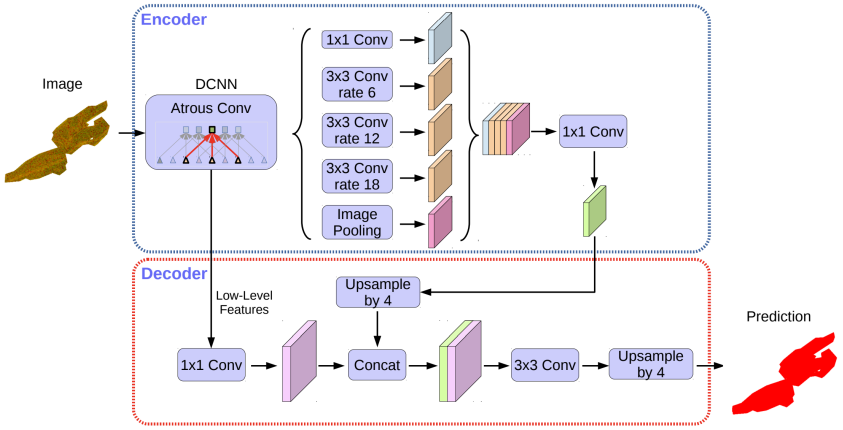


Figure 5.5.: *Deeplab v3+* architecture. Best if viewed on screen.

are often not available. Even if they are, labelling them is costly and increases the computational cost of model training. *Transfer learning* mitigates this bottleneck by using an already trained model from some related task as a starting point. Given the fact that the initial layers of a neural network learn rather generic local image properties, a model trained on a huge image dataset can be re-utilised on a different dataset with a much smaller amount of fine-tuning (re-training) to the specific characteristics of the new data. We use a *Deeplab v3+* model pre-trained on the PASCAL VOC 2012 close-range dataset as the starting point and fine-tune it on the relatively small Sentinel-1 SAR dataset (see Table 5.2). Surprisingly, we find that pre-training on RGB amateur images of indoor scenes, vehicles, animals, humans etc. greatly enhances the performance even on a data source as different as interferometric Radar, compared to training from scratch only on the SAR data. Note, all weights were fine-tuned, we did not freeze any network layers.

5.5. Experiments, results, and discussion

We use various measures to quantify performance, including recall, precision, overall accuracy, and the IoU score (Jaccard index). In all experiments described in the paper, we used only *lake pixels* from the non-transition dates to train the network and to compute performance metrics. This is due to the lack of reliable ground truth during the transition dates. Additionally, whenever the ground truth cannot be established for a non-transition date due to foggy webcam images and/or clouds in Sentinel-2, it is exempted from the training set. However, qualitative analysis is done on all the dates.

Quantitative results: Semantic segmentation. For developing an ideal operational system for lake ice monitoring, the data from a couple of lakes from a few winters would have to be used to train the model, which can then be deployed on unseen lakes and winters. However, generating ground truth for each lake is a tedious task. Nevertheless, we make sure that the data from at least one lake from one full winter is in the training set for the classifier to learn the proper class decision boundaries.

We employ Cross-Validation (CV), i.e., the data is partitioned into k folds, usually

of approximately the same size. Then, the evaluation is done k times, each time using one fold as test set and the union of all remaining folds as training set. *Leave-one-out cross-validation* is the setting where the number of folds equals the number of instances (in our case the number of winters/lakes) in the dataset.

The goal of *leave-one-winter-out CV* is to investigate the generalisation capability of a model trained on one winter when tested on a different winter. The results are shown on Table 5.4. It can be seen that we achieve excellent results for both winters with average accuracies of 95.5% and 94.8% for 2016-17 and 2017-18 respectively. The results show that the model generalises well across the potential domain shift caused by the specific conditions of different winters, without having seen data from any day within the test period.

Table 5.4.: Results for winter 2016-17 (top) and 2017-18 (bottom). Data from all the three lakes from winter 2016-17 was used to train the model that was tested on winter 2017-18, and vice versa. Confusion matrices are shown. Units are in millions of pixels, except for precision, recall, and **accuracy** (bottom right cell in each table).

True \ Prediction	Non-frozen	Frozen	**Recall**
Non-frozen	3.06	0.01	99.7%
Frozen	0.29	2.89	90.8%
Precision	91.3%	99.6%	**95.5%**

True \ Prediction	Non-frozen	Frozen	**Recall**
Non-frozen	5.77	0.19	96.7%
Frozen	0.44	4.59	91.1%
Precision	92.9%	95.8%	**94.8%**

We also report results of a *leave-one-lake-out* CV experiment to check the generalisation capacity of the model across lakes. The results are shown on Table 5.5. While testing for all data of a lake (e.g., *Sils*) from two winters, the data from the other two lakes (e.g., *Silvaplana, St. Moritz*) from the same two winters is used for training. The prediction achieves >91.5% overall accuracy for all three lakes. See Table 5.6 for the per-class and mean IoU values for each lake. The worst performance is noted for St. Moritz where the lake itself is smaller than the size of a single patch (128×128). Note also, St. Moritz has the least mIoU, especially for class *frozen*. This is because of the presence of tents and other infrastructure on the frozen lake St. Moritz, which is not present in the other two lakes used in training. To assess the per-class performance in detail, we also report the precision-recall curves in Fig. 5.6. For both the frozen and non-frozen classes, the area under the curve is nearly optimal for lake Sils and very good performance is achieved on lakes Silvaplana and St. Moritz.

Quantitative results: Time series. The *ice-on* and *ice-off* dates are of particular interest for climate monitoring. From the per-day semantic segmentation results, we

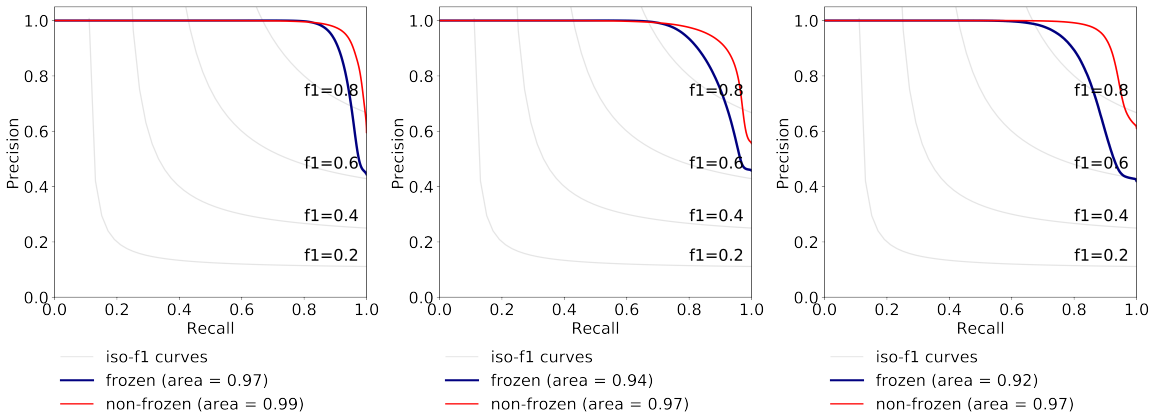


Figure 5.6.: Precision-recall (PR) curves for lakes Sils (left), Silvaplana (middle), and St. Moritz (right) for winter 2017-18. The *iso-f*1 curve connects all points in the PR space with same $F1$ score. Combined data of all 3 lakes from winter 2016-17 was used to train the model. Best if viewed on screen.

Table 5.5.: Results for lake Sils (top), Silvaplana (middle), and St. Moritz (bottom). Confusion matrices are shown for the leave-one-lake-out cross-validation experiment. Units are in millions of pixels, except for precision, recall, and **accuracy** (bottom right cell in each table).

True \ Prediction	Non-frozen	Frozen	**Recall**
Non-frozen	4.69	0.02	99.4%
Frozen	0.13	4.22	96.9%
Precision	97.3%	99.4%	**98.3%**

True \ Prediction	Non-frozen	Frozen	**Recall**
Non-frozen	3.08	0.10	96.4%
Frozen	0.11	2.78	96.1%
Precision	96.4%	96.4%	**96.4%**

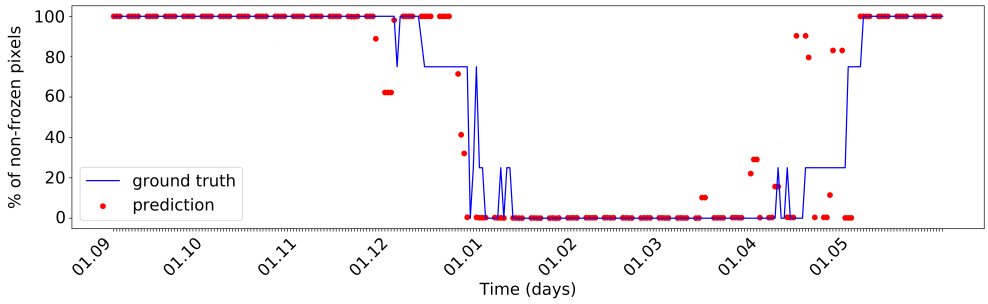
True \ Prediction	Non-frozen	Frozen	**Recall**
Non-frozen	1.00	0.10	94.1%
Frozen	0.62	0.82	88.5%
Precision	94.1%	88.5%	**91.5%**

estimate the daily percentage of frozen surface for each observed lake. Thus, for each available SAR image, we compute the percentage of frozen pixels, throughout the entire winter. An example time series for lake Silvaplana in winter 2017-18 is shown in Fig. 5.7a. Although we do not have per-pixel ground truth on the partially frozen transition days, we know whether the lake has more water (shown with a value of 75% in the ground truth) or more ice/snow (shown with a value of 25% in the ground truth). Even though some

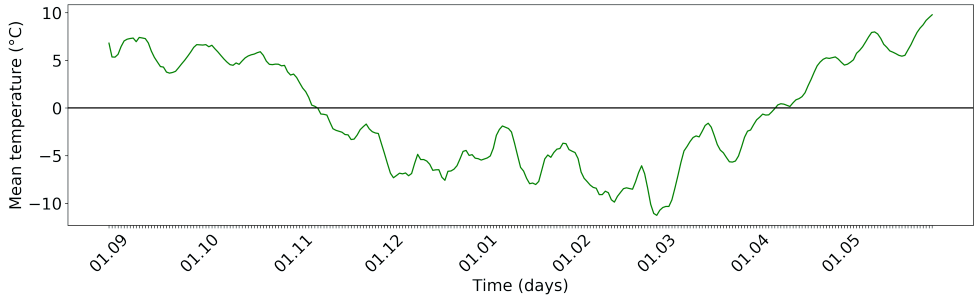
Table 5.6.: Per-class- and mean IoU values of frozen and non-frozen classes for each lake. The data of a lake from two winters ($2016-17$ and $2017-18$) is tested using a model trained on the data from the other two lakes from both winters.

IoU \ Lake	Sils	Silvaplana	St. Moritz
Non-frozen	96.7%	93.3%	85.6%
Frozen	96.4%	92.7%	82.9%
Mean	96.5%	93.1%	84.3%

miss-classifications exist during the transition days, the non-transition days are almost always predicted correctly, likely because the network was trained solely on non-transition days. For a comparison, we also plot the time series of temperature values (sliding window mean of the daily median, window size 7 days) obtained from the nearest meteo station in Fig. 5.7b. Sub-zero values in this graph correlate (with some time lag) with the period in which the lakes are fully or partially frozen.



(a) Time series of percentage of non-frozen pixels for lake Silvaplana from winter 2017-18.



(b) Temperature (temporal moving average of daily median with window size of 7 days) from the nearest meteo station.

Figure 5.7.: Correlation of our results (winter 2017-18) on lake Silvaplana with the ground truth and the auxiliary temperature data. Best if viewed on screen.

Qualitative analysis. Exemplary qualitative results are depicted in Fig. 5.8. We show the classification results on frozen, non-frozen, and transition dates along with the probability map (blue means higher probability of frozen, red means higher probability of non-frozen). For better interpretation of the result, especially for the transition date, we show the corresponding image from Sentinel-2.

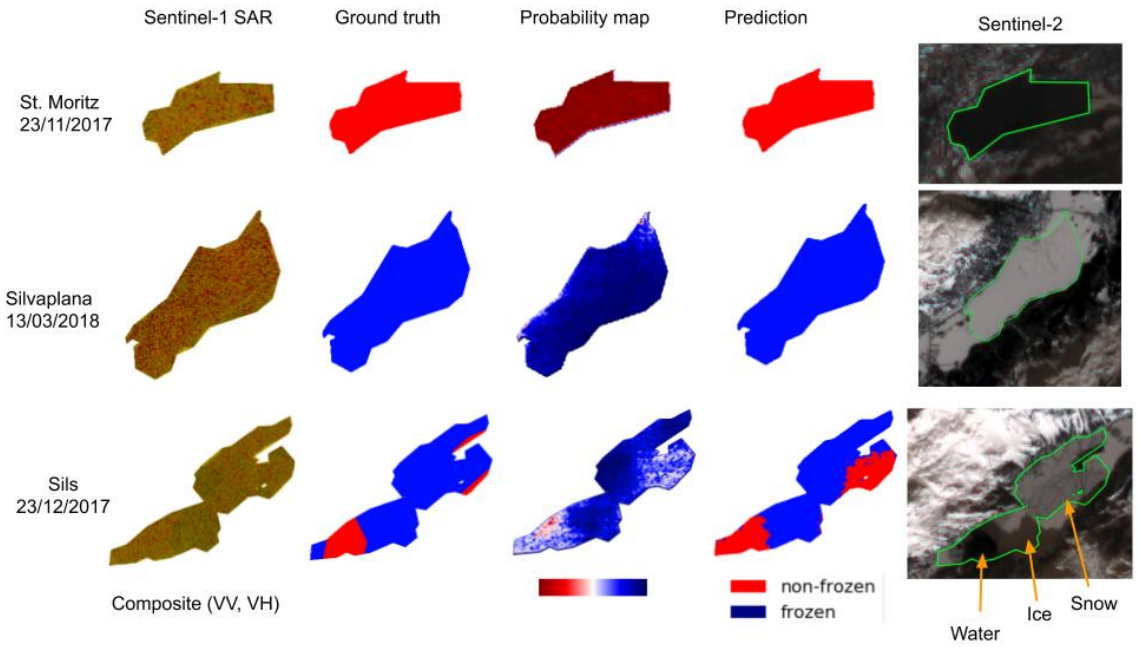


Figure 5.8.: Qualitative results for lake St. Moritz on a non-frozen day (row 1), lake Silvaplana on a frozen day (row 2), and lake Sils on a transition day (row 3). For each lake we show the Sentinel-1 composite image (column 1), the ground truth (column 2), the predicted frozen probability map derived from the network logits (column 3), and the corresponding binary classification map (column 4). Additionally, column 5 shows a corresponding Sentinel-2 image for better visual interpretation.

Miscellaneous experiments. In all the experiments reported so far, we used the data from all four orbits (both ascending and descending) and both polarisations (VV and VH). To study the individual effect of polarisations VV and VH, we drop either of them and report the corresponding results on Table 5.7 (top). Note that mIoU drops by almost 25% when VV is left out, while it drops by only 3.7% without VH, confirming the significance of polarisation VV for lake ice detection. This finding also aligns with the visual differences in Fig. 5.3. However, while VH appears to be less discriminative overall, it is much less affected by wind speed – see Fig. 5.4. We believe that using also VH may improve robustness in windy conditions, where discriminating water from ice/snow should be particularly challenging because of increased surface roughness due to waves. However, we do not have enough days with strong wind to quantitatively corroborate this hypothesis. From our current data it appears that the system can handle calm and moderately windy days practically equally well. In another experiment, we drop VH but use VV from two temporally adjacent acquisitions (VV and VV_prev), thus simultaneously feeding the network with data from two different days. The mIoU rises by 2%, see Table 5.7 (top). However, we noted some stability issues especially during fully frozen days. We believe this is due to the fact we removed VH. We did another experiment to check the effect of acquisition time. Here, we used the data from both VV and VH. However, we drop the data from some orbits (see Table 5.3). Table 5.7 (bottom) shows that the mIoU drops by 7.1% and 5.7% respectively when the data from only descending orbits (66 and 168) or only ascending orbits (15 and 117) were used.

A final experiment was done to assess the influence of the (rectangular) training patch size, see Fig. 5.9. Somewhat surprisingly, even the move from an already large context of 64×64 pixels to 128×128 pixels (1.3×1.3 km) still brought a marked improvement, hence we always use that patch size in our system.

Table 5.7.: Per-class- and mean IoU values of frozen and non-frozen classes with different polarisations (top table) and orbits (bottom table). Data from all three lakes from winter 2016-17 was tested using a model trained on the data from all three lakes from winter 2017-18. *Asc* and *Dsc* denotes ascending and descending orbits respectively.

IoU \ Polarisation	VV, VH	VH	VV	VV, VV_prev
Non-frozen	91.0%	71.0%	87.7%	93.3%
Frozen	90.6%	60.1%	86.6%	91.9%
Mean	90.8%	65.9%	87.1%	92.6%

IoU \ Direction	Asc, Dsc	Dsc	Asc
Non-frozen	91.0%	84.7%	85.9%
Frozen	90.6%	82.7%	84.3%
Mean	90.8%	83.7%	85.1%

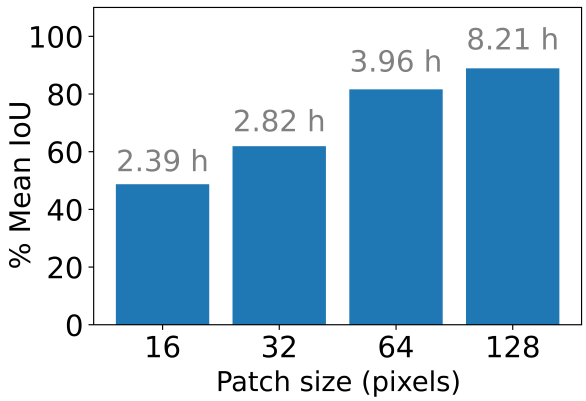


Figure 5.9.: Mean IoU values obtained with different input patch sizes. The total time taken to complete training and testing is also indicated, in hours (h). The results are for all three lakes combined, training on 2016-17 and testing on 2017-18.

5.6. Conclusion and outlook

We have described a system for reliable monitoring of lake ice based on Sentinel-1 SAR imagery, with the potential to retrieve long, consistent time series over many years (assuming continuity of the satellite mission). The proposed method has been demonstrated for three different Swiss lakes over two complete winters, and obtains good results (mIoU

90% on average, and >84% even for the most difficult lake), even when generalising to an unseen winter or lake. Given the main advantage of SAR data for our purposes – its ability to observe with very good spatial and temporal resolution independent of clouds – we see the possibility to extend our method into an operational monitoring system. A logical next step would be to process longer time series, which unfortunately is not yet possible with Sentinel-1. It is quite possible that even a moderate time span, say 20 years, would suffice to reveal trends in lake freezing patterns and perhaps also correlations with climate change. Another future direction is an integrated monitoring concept, using SAR together with optical satellite imagery and optionally images from webcams, to ensure reliable identification of *ice-on* and *ice-off* dates within the GCOS specification of ±2 days.

5.7. Acknowledgements

This work is part of the project *Integrated lake ice monitoring and generation of sustainable, reliable, long time series* funded by Swiss Federal Office of Meteorology and Climatology MeteoSwiss in the framework of GCOS Switzerland. Also, we thank Anton B. Ivanov from Skoltech for his support.

6 Learning a Sensor-invariant Embedding of Satellite Data: A Case Study for Lake Ice Monitoring

Manu Tom, Yuchang Jiang, Emmanuel Baltsavias, Konrad Schindler
Under review

Abstract

Fusing satellite imagery acquired with different sensors has been a long-standing challenge of Earth observation, particularly across different modalities such as optical and Synthetic Aperture Radar (SAR) images. Here, we explore the joint analysis of imagery from different sensors in the light of representation learning: we propose to learn a satellite embedding (feature representation) within a deep neural network. Our application problem is the monitoring of lake ice on Alpine lakes. To reach the temporal resolution requirement of the Swiss Global Climate Observing System (GCOS) office, we combine three image sources: Sentinel-1 SAR (S1-SAR), Terra MODIS and Suomi-NPP VIIRS. The large gaps between the optical and SAR domains and between the sensor resolutions make this a challenging instance of the sensor fusion problem. Our approach can be classified as a feature level fusion that is learnt in a data-driven manner. The proposed network architecture has separate encoding branches for each image sensor, which feed into a single latent embedding. I.e., a common feature representation shared by all inputs, such that subsequent processing steps deliver comparable output irrespective of which sort of input image was used. By fusing satellite data, we map lake ice at a temporal resolution of <1.5 days. The network produces spatially explicit lake ice maps with pixel-wise accuracies >91.3% (respectively, mIoU scores >60.7%) and generalises well across different lakes and winters. Moreover, it sets a new state-of-the-art for determining the important ice-on and ice-off dates for the target lakes, in many cases meeting the GCOS requirement.

6.1. Introduction

Multi-modal satellite data analysis is an important capability and an active area of research in remote sensing and Earth observation (Schmitt and Zhu, 2016). Its aim is to combine the data acquired with different sensors. With the ever-increasing number of operational satellites and the greater variety of imaging sensors, such a combined analysis becomes even more important. One important advantage of combining (or "fusing") data from multiple satellites is a higher temporal sampling frequency, so as to obtain denser time series of dynamic processes.

A whole body of literature exists on combining satellite data from different sensors, at various levels of processing (pixel level, feature level, decision level). In most cases the fused data have comparable spatial resolution, e.g., S1-SAR and Sentinel-2 (S2), TerraSAR-X and ALOS PRISM, etc. Here we aim to merge several sensors with high temporal resolution, leading to a large gap in spatial resolution: while S1-SAR acquires an image of our mid-latitude target area every 1.5-4.5 days at 10m Ground Sampling Distance (GSD), optical sensors with similar revisit times have moderate spatial resolutions (GSD of 250m for MODIS, respectively ≈375m for VIIRS). See Table 6.1. Due to the large resolution gap we combine them at the feature level. This makes it possible to use texture information from the high-resolution S1-SAR and circumvents pixel-accurate matching and co-registration, which is challenging across different modalities with very different radiometries and large changes in GSD, with associated uncertainties in absolute geo-location and relative co-registration.

Our goal here is to analyse a time series composed of images from different sensors. Intuitively, this problem is solved if we can transform all input images to the same, sensor-invariant representation, while at the same time making sure that representation preserves the information necessary for the intended downstream task, in our case the detection of lake ice. We leverage the ability of deep neural networks to learn at the same time a complex, non-linear mapping of the inputs into a latent feature representation, via sequences of convolutions and element-wise, non-linear activation functions; and the subsequent mapping from that representation to the desired output variables. Our proposed network includes a separate encoder branch with individual layout and weights for each input sensor, to map its raw image values to predictive features. However, the different encoders all share the same feature embedding, which forms the input for two branches: an auxiliary branch that outputs a per-pixel classification into frozen (including snow-covered ice and snow-free ice) or non-frozen (water) and background classes, and serves mainly to train the representation; and a branch that regresses the fraction of the lake covered by water (non-frozen pixels). The joint embedding means that the encoder branches must learn to produce a single representation that is shared across all sensors (respectively, branches), while at the same time being a suitable basis for the prediction tasks.

Fusion of multi-source satellite data has been employed for a wide variety of Earth observation tasks, including land cover classification, object detection, change detection, etc. For instance, SAR-optical fusion has been applied to sharpen (spatially) low-resolution optical images by fusing them with high-resolution SAR images (Sanli et al., 2013), to improve the information extraction by exploiting the complementary nature of radar and

optical sensing (Schmitt and Zhu, 2016), and for cloud removal from optical images (Gao et al., 2020, Meraner et al., 2020). In this paper, we use fusion to construct denser time series from heterogeneous data, for the specific application of lake ice monitoring. To that end we infer two outputs: a spatially explicit map of frozen and non-frozen lake pixels at every time step; and a time series of the fraction of water, which serves as a basis for estimating the critical events of lake ice phenology (LIP), such as *ice-on* and *ice-off* dates.

Lake freezing and thawing are strongly correlated to local warming of the atmosphere (Tom et al., 2021b, Qi et al., 2020), and *Lake Ice Cover* has been included in the list of essential climate variables (`https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables`). Hence, lake ice monitoring is important to support climate science and cryosphere research, and ultimately to support climate change mitigation (Rolnick et al., 2019).

For lake ice monitoring, the requirements of the Swiss GCOS office are daily observations and an accuracy of ±2 days for the critical ice-on/off dates (Tom et al., 2019). Several papers have discussed lake ice observation with the help of machine learning, using a single satellite sensor (Tom et al., 2018, 2020a, Wu et al., 2021b), but unfortunately none of those sensors has so far met the criteria of ±2 days. In earlier work (Tom et al., 2020c) we have integrated MODIS and VIIRS data at the decision level, but still failed to meet the GCOS target in most cases. Terrestrial webcams deliver image sequences that would satisfy the GCOS criteria (Xiao et al., 2018, Prabha et al., 2020), but unlike satellite-based methods, they are not suitable for systematic and large-scale monitoring, since most lakes do not have webcams observing them, and even if there is a camera its field of view rarely covers the entire lake surface.

In this work, we combine data from multiple satellite sources to improve the temporal resolution, and thus the accuracy of the critical LIP events around the freeze-up and break-up periods. MODIS and VIIRS satellite images are available daily, at the cost of reduced spatial resolution. Still, the GSD is adequate for all but the smallest lake. The main issue with optical satellite images is data loss due to clouds, which can significantly decrease the effective temporal resolution. A case in point is our target application of monitoring Alpine lakes, as they are frequently obscured by clouds, especially around the important ice-on/-off dates. This means that adding further optical sensors, such as S2 or Landsat-8, will not solve the problem; as they are equally affected by clouds. On the contrary, SAR will not be influenced by clouds. S1-SAR provides regular coverage (for the target regions in Switzerland at least every 4.5 days, but in the best case even every 1.5 days, depending on the number of usable orbits as well as data availability in early stages of the mission). See Table 6.1 for details. Contrary to the moderate-resolution optical sensors, S1-SAR has a much higher spatial resolution. Therefore, texture analysis becomes important, while at the same time the correct texture description for the task is not obvious. A natural solution is to use a deep convolutional network. Note also, the much higher resolution means many more pixels are available within the area of a given lake, playing to the strength of data-hungry deep networks. Lake ice mapping with S1-SAR has been shown to work rather well and has the added advantage that even tiny lakes can be monitored, but by itself the revisit time is not sufficient to meet the

GCOS specification (Tom et al., 2020a). Below we will show that, at least for our target lakes, this target can in many cases be met with the combination of MODIS, VIIRS and S1-SAR.

Table 6.1.: Details about the satellite data that we use.

	MODIS	**VIIRS**	**Sentinel-1 SAR**
Satellite type	optical	optical	radar
Spatial resolution	250–1000m	375–750m	10m
Temporal resolution	1d	1d	1.5–4.5d (for Switzerland)
Spectral resolution	36 bands	22 bands	C-band (3.8–7.5cm),
	(0.4–14.2μm)	(0.4–12.0μm)	4 polarisations (mainly VV, VH)
Cloud problems	severe	severe	nil
Cloud mask issues	slight	slight	NA
Costs	free	free	free
Availability	very good	very good	very good
	(via VIIRS continuity)		(HV / HH only partially available)

6.1.1. Related work

Satellite data fusion has been a long-standing topic in remote sensing and beyond. Arguably one of the most basic forms of data fusion is pan-sharpening, i.e., fusing an image with high spectral, but low spatial resolution with one that has high spatial resolution but only a single (normally broader) spectral band, eg., Ehlers et al. (2010). Related are more complicated spectral-spatial fusion scenarios (Ranchin and Wald, 2000, Melgani and Serpico, 2002, Huang and Song, 2012, Huang et al., 2013a,b, Lanaras et al., 2017a,b, 2018) as well as spatio-temporal fusion techniques (Song et al., 2018, Zhu et al., 2018). For readers looking for a general overview of remote sensing data fusion, there are several well-crafted review papers (Zhang, 2010, Joshi et al., 2016, Schmitt and Zhu, 2016, Schmitt et al., 2017, Kulkarni and Rege, 2020) that discuss technical challenges, solutions, applications, and trends.

Here, we will limit our review to directly related work, by which we mean recent (deep) representation learning methods in the context of SAR-optical satellite data fusion. Furthermore, on the application side we will review learning-based approaches to lake ice monitoring.

Fusing optical and SAR satellite images with deep learning

Several authors have used Convolutional Neural Networks (CNNs) to integrate the complimentary information in optical and SAR data, typically using the concept of two-stream networks with two input branches (one for each modality). Mou et al. (2017) proposed to use a two-stream pseudo-Siamese CNN ("SARptical Convolutional Network") to match patches of urban scenes in very high spatial resolution TerraSAR-X (1.25m pixel spacing) and airborne UltraCAM optical imagery (20 cm GSD). They achieved an overall accuracy of 97.48% at a 0.05% false alarm rate, on a relatively small dataset with 109 SAR and 9 optical images. Merkle et al. (2017) put forward another two-stream, Siamese

CNN to extract features from both optical (PRISM, 2.5m GSD) and SAR (TerraSAR-X) images, followed by a dot-product layer to compute similarities between the extracted features. Their primary goal was to improve the geo-location of optical images, by precisely co-registering them with the corresponding SAR data. Scarpa et al. (2018) used SAR imagery (S1-SAR), in conjunction with CNN-based data fusion, to estimate spectral features for cloudy days where optical data (S2) is unusable.

A dataset "SEN1-2" with image pairs from S1-SAR (VV polarisation) and S2 (only RGB channels) has been proposed by Schmitt et al. (2018) to support research into geo-spatial data fusion. As a follow-up, another curated dataset "SEN12MS" of geo-coded multi-spectral satellite imagery was also made available, which includes patch triplets (dual-polarised S1-SAR, multi-spectral S2 and MODIS land cover) custom tailored to the training of deep learning methods (Schmitt et al., 2019). In another recent work (Bürgmann et al., 2019) ground control points are derived from high spatial resolution TerraSAR-X imagery (1.25m pixel spacing) to improve the absolute geo-location of optical images (Pléiades, 0.5m and 2m GSD for panchromatic and multi-spectral bands, respectively). That method used an adapted version of HardNet (Mishchuk et al., 2017) pre-trained on the SEN1-2 dataset, which was then fine-tuned on the high spatial resolution target data. Hoffmann et al. (2019) proposed a fully convolutional neural network to predict the similarity metric between SAR and optical images, and reported large improvements over standard metrics based on mutual information, for a subset of the SEN1-2 data. Wang et al. (2018b) performed registration of Landsat, Radarsat and SPOT imagery, also with a deep network. They explored transfer learning to save training time. Additionally, they utilised a self-learning trick to work around the lack of labelled training data. Still concerning SAR-to-optical registration, Hughes et al. (2020) designed a three-step framework, consisting of a "goodness" network that localised image regions suitable for matching, followed by a correspondence network that generated a heatmap of matching scores, and a subsequent outlier rejection network.

Finally, there have also been some studies about the potential of Generative Adversarial Networks (GANs). Hughes et al. (2018) matched TerraSAR-X and UltraCAM data. To lower the rate of false positive matches, they mined hard negative samples and trained a variational autoencoder (VAE) with an adversarial strategy, so as to learn the latent distribution of the training data and synthesised realistic "negative" patches. GANs were also adopted to remove clouds from optical satellite images (Gao et al., 2020). In a first step, a CNN estimates an optical image from the cloud-free SAR image ("image-to-image translation"). In the next step, that synthesised image is fused with the original, cloudy optical image to replace the cloud pixels, using a GAN. The method was demonstrated on three different sensor pairs, namely S1/S2, Gaofen-3/2, and airborne SAR/optical images.

Lake ice monitoring

The literature on ice monitoring is vast, as a diverse range of data sources and models have been investigated. Here, we focus on research that used machine learning to derive the relationship between image observations and presence of ice. In our own earlier work (Tom et al., 2018) we have demonstrated semantic segmentation of MODIS and

VIIRS data into frozen and non-frozen pixels, using Support Vector Machines (Cortes and Vapnik, 1995) on the raw channel intensities. The two sensors were then combined via decision-level fusion to extract the ice-on/-off dates for two winters (Tom et al., 2020c). In many cases the critical dates were determined at an accuracy of 1-4 days. We also confirmed that, for the same target lakes, a deep learning model [Deeplab V3+, Chen et al. (2018b)] could perform the frozen/non-frozen segmentation of S1-SAR amplitude images, but still failed to determine the ice-on/-off dates with sufficient accuracy (Tom et al., 2020a).

Hoekstra et al. (2020) combined Iterative Region Growing Segmentation (IRGS) with a Random Forest (RF) classifier (Breiman, 2001) to distinguish lake ice from open water in RADARSAT-2 images of Great Bear Lake (Canada) from 2013 to 2016, attaining an overall segmentation accuracy of ≈96%. Wu et al. (2021b) have recently compared the performance of four popular machine learning classifiers (SVM, RF, Gradient Boosted Trees [GBT], Multinomial Logistic Regression) for lake ice detection from MODIS, for 17 large lakes with areas >1040 km^2, situated across Northern Hemisphere. They reported the best performance with RF and GBT. For our small Alpine lakes, we have also performed a similar comparison with SVM, RF and XGBoost (Chen and Guestrin, 2016), but found that a linear SVM achieved the best generalisation across different lakes and years (Tom et al., 2021b), due to non-linear methods over-fitting to the comparatively small number of training pixels.

Beyond satellite images, there have been attempts to use webcam streams or even crowd-sourced amateur photographs for lake ice monitoring (Xiao et al., 2018, Prabha et al., 2020). To that end, state-of-the-art encoder-decoder network for semantic segmentation of close-range images (Jégou et al., 2016) was retrained (with minor modifications) to distinguish the relevant classes such as ice, snow and water. While ice detection based on webcam images works rather well and largely is unaffected by cloud cover (although a few images may have to be dropped because of excessive fog, rain or snowfall), it is at this point unclear how to scale up such an approach, as only a small portion of all relevant lakes are (often only partially) observed by webcams.

6.1.2. Definitions used

Following Franssen and Scherrer (2008), we define *ice-on* as the first date on which a lake is "almost fully" frozen, and followed by a second day where this remains so; i.e., the end of the freeze-up period, and *ice-off* as the first day on which a significant amount of water re-appears and remains visible for another day; i.e., the start of the break-up period.

All the pixels that fall completely inside the lake boundary (obtained from *openstreetmap.org*, generalised) are considered as *clean* pixels and used for training and inference, so as to sidestep the handling of mixed pixels and the influence of geo-location errors. Moreover, we call *non-transition* days all (>30% cloud-free) days on which the lake is either entirely frozen or entirely non-frozen. The remaining (again, >30% cloud-free) days are referred to as *transition* dates. Only non-transition days are used to train the segmentation, as spatially explicit labels for the transition days are difficult to obtain.

Whereas the regression of the water fraction is trained on all the dates.

6.2. Target lakes, winters, and satellites

6.2.1. Target lakes and winters

For our case study, we monitor lake ice on four (mid-latitude, Alpine) lakes in Switzerland: Sihl, Sils, Silvaplana and St. Moritz. See Fig. 6.1. Lake Sihl is located near the Swiss plateau while the latter three lakes are located close to each other at higher altitude in the Engadin valley, and share similar geographical and environmental conditions. More details about these lakes can be found in Table C.1 (see Appendix C). For each lake, we process all three satellite sources from two winters: 2016–17 and 2017–18. In each winter, all available cloud-free dates between beginning of September and end of May are analysed.

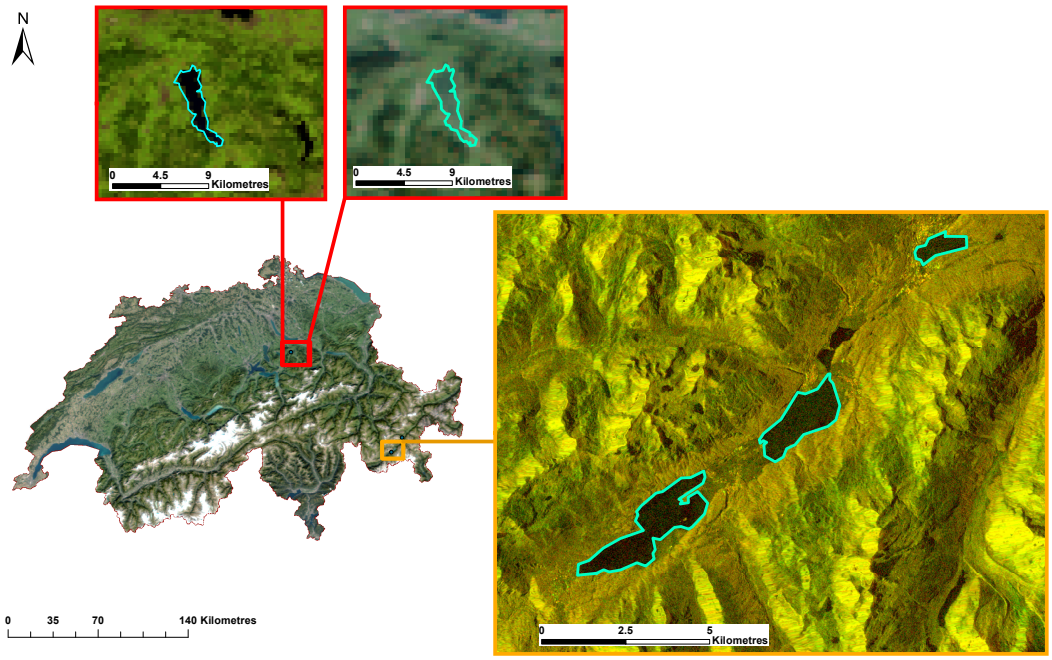


Figure 6.1.: On the orthophoto map of Switzerland, the geographic location of the target Swiss lakes are marked as red and amber rectangles. RGB True Colour Composite (TCC) of VIIRS (top left, R=I_3, G=I_2, B=I_1, captured on 7 September 2016) and MODIS (top right, R=B_1, G=B_4, B=B_3, captured on 7 September 2016) for the region Einsiedeln (lake Sihl) are shown in the zoomed red rectangles. For the region Engadin (lakes Sils, Silvaplana and St. Moritz, from left to right), the RGB TCC (R=VV, G=VH, B=0) of Sentinel-1 SAR (captured on 13 September 2016) is shown in the zoomed amber rectangle. Best if viewed on screen.

6.2.2. Target satellites

Optical satellite data (MODIS and VIIRS)

We use Terra MODIS (`https://terra.nasa.gov/about/terra-instruments/modis`) and Suomi NPP VIIRS (`https://ncc.nesdis.noaa.gov/VIIRS/`) data in our analysis. For both these data types, we follow the download and pre-processing procedures as described in Tom et al. (2020c), which include absolute geolocation correction, backprojection of the generalised lake outlines onto the satellite images to extract clean pixels, cloud filtering, and bilinear interpolation for low-resolution bands (upsampling the 500m and 1km bands to 250m, only for MODIS). All cloud-free lake pixels in acquisitions with >30% non-cloudy pixels are extracted and processed. We note that for VIIRS, after absolute geolocation correction, there is no clean pixel available for lake St. Moritz.[1] Instead of using all the available bands, 12 (5) out of 36 (22) spectral bands are selected for MODIS (VIIRS) as suggested in Tom et al. (2018). For MODIS, this selection was done after a visual check and 12 potentially useful bands without artefacts (stripes) or saturation issues were picked. For VIIRS, all five imagery bands were selected. To generate images with uniform size (to be fed as input to the network), the lake pixels are padded with background pixels to form 12×12 patches (input dimensions of $12\times12\times12$, respectively $12\times12\times5$ for MODIS and VIIRS). Details about the optical images are displayed in Table 6.1.

Radar satellite data (Sentinel-1 SAR)

We download already pre-processed dual polarisation C-band S1-SAR data as level-1 Ground Range Detected (GRD) scenes in Interferometric Wide (IW) swath mode from the Google Earth Engine (GEE) platform (Gorelick et al., 2017). The GEE pre-processing includes border and thermal noise removal, radiometric calibration, terrain correction, absolute geolocation correction, and log scaling. Additionally, we resize the images spatially to 128×128 resolution (covering all lake pixels and some background). Although the instrument can record four polarisations (VV, VH, HV, HH), only VV and VH are available for the regions of interest around Einsiedeln and Engadin, in Switzerland. Hence, the final input size for SAR images is $128\times128\times2$. The temporal resolution varies across regions as well as winters, ranging from 1.5-4.5 days. From winter 2017–18, the data is available every 1.5 days for lakes in region Engadin (Sils, Silvaplana, St. Moritz) and every 3 days for region Einsiedeln (Sihl). On the other hand, in 2016–17, it is 2.3 and 4.5 days respectively. Sentinel-1 (S1) scans region Engadin in four orbits, but region Einsiedeln only in two orbits. This explains why the temporal resolution for Engadin is relatively high for a given winter. In winter 2017–18, the data from both S1A and S1B satellites are available. Additionally, even though the S1B was launched in April 2016, the corresponding data is available only since March 2017, effectively offering a relatively lower temporal resolution for winter 2016–17. More details on the S1-SAR data that we use are also summarised in Table 6.1.

[1] We include this tiny lake mainly to assess the limits of satellite-based ice monitoring for very small lakes.

Satellite combination: effective temporal resolution

For the target lakes and winters, we compute the effective temporal resolution (average number of days between cloud-free image acquisitions) separately for each input sensor and for the combination, see Table 6.2. For the optical satellites, only the cloud-free (at least 30%) days are counted while computing the effective temporal resolution. In any winter, there can be days on which all three satellites have imaged the regions of interest, days on which these regions were scanned more than once by the same satellite (rare), and days with no S1-SAR acquisition and no available cloud-free MODIS and VIIRS data.

Table 6.2.: Effective temporal resolution during the target winters in days, for different sensors. S, M, V denote S1-SAR, MODIS and VIIRS respectively.

Lake	Winter	S	M	V	S+M+V
Sihl	2016–17	4.5	1.9	2.1	1.5
	2017–18	3.0	2.2	2.4	1.5
Sils	2016–17	2.3	1.7	1.7	1.3
	2017–18	1.5	2.0	2.0	1.2
Silvaplana	2016–17	2.3	1.7	1.7	1.3
	2017–18	1.5	2.1	1.9	1.3
St. Moritz	2016–17	2.3	1.7	-	1.4
	2017–18	1.5	2.0	-	1.3

For a given lake and winter, different effective temporal resolution for MODIS and VIIRS is possible since their overpasses occur at different times of day and the cloud patterns may change. Additionally, in the presence of scattered clouds the cloud cover can vary even between the lakes located in the same valley (Sils, Silvaplana, St. Moritz), in the same overpass. Even though both optical satellites have daily acquisitions, the effective temporal resolution for each sensor is greater than 1 day even at a low threshold of <70% clouds. In the best case with zero clouds, both MODIS and VIIRS individually would have daily temporal resolution and no fusion with SAR data is needed. However, in practice, clouds are inevitable in the target regions and the chosen scenario of using (S+M+V) is a realistic minimal setup: For the two winters of interest, the effective temporal resolution after combining the data from three sensors is between 1 and 2 days. See Table 6.2 (a day is counted only once if captured by more than one satellite). It can be noted from Table 6.2 that winter 2017–18 was relatively more cloudy than 2016–17. Still, the temporal resolution of the three sensors combined is better in 2017–18, due to the extra data from the S1B satellite, which was not available in 2016–17.

6.2.3. Ground truth

The reference data was generated by visual interpretation of the images from freely available webcams located in the vicinity of the target lakes. The images were labelled independently by a human operator, and independently confirmed by a second one. Whenever there were doubts about the per-day label, the operators used more images from the same webcam, other webcams monitoring the same lake (if available), information from ad-

jacent days, and in some cases also the online media reports and interpretation of S2 satellite images (if available and cloud-free) to corroborate their final choice of label. On non-transition days, only a binary label was assigned: fully-frozen (fraction of water pixels 0) or fully non-frozen (fraction 1). Whereas on transition days the assigned labels have higher granularity and also include the states: more frozen (0.25) and more non-frozen (0.75). There is a price to pay, as these additional states cannot always be labelled reliably and a significant amount of label noise must be expected. Due to interpretation uncertainties and not clearly visible regions, as well as the fact that not the entire lake surface is visible in the webcam images, the actual water area of a "more frozen" lake is likely to fluctuate in the range 60-90% of the lake surface. Moreover, due to interpretation errors caused by bad lighting, compression artefacts of webcam streams, and low spatial resolution especially in the far field, occasional confusions between adjacent states are almost certainly present in the reference data. These issues can be hardly avoided, as webcams are typically placed relatively low on shore, such that the lake surface is viewed at acute angles and the image scale degrades rapidly with distance from the camera.

For training the segmentation task, only the non-transition dates are used, after converting the per-image labels to spatially explicit maps with 100% frozen, respectively non-frozen pixels inside the lake boundary. Pixels outside the lake are labelled as background.

6.3. Methodology

6.3.1. Data fusion model

We propose a *2-step* model for deep satellite data fusion, see Fig. 6.2 for a depiction of the network architecture. The two underlying steps are:

1. learn a satellite "embedding" representation
2. use the learnt embedding as input for the interpretation task, in our case lake ice monitoring

The first step transforms the inputs to a new feature space (a "latent representation"). To train that step, which should preserve and accentuate information about the local state of the lake surface, we use the auxiliary task to explicitly segment the lake into three semantic classes, *frozen*, *non-frozen* and *background*. The second step starts from the resulting feature representation and regresses the fraction of water (non-frozen) pixels on the lake, including a mechanism for multi-temporal analysis over a short period of time around the day in question. In the following we describe each of the two steps in more detail.

Step 1: satellite embedding learning

To perform fusion, the domain gap between the heterogeneous inputs that stem from satellite sensors operating at very different wavelengths (optical, radar) and spatial resolutions (10 to 375m), need to be bridged. We achieve this by learning a shared, in-

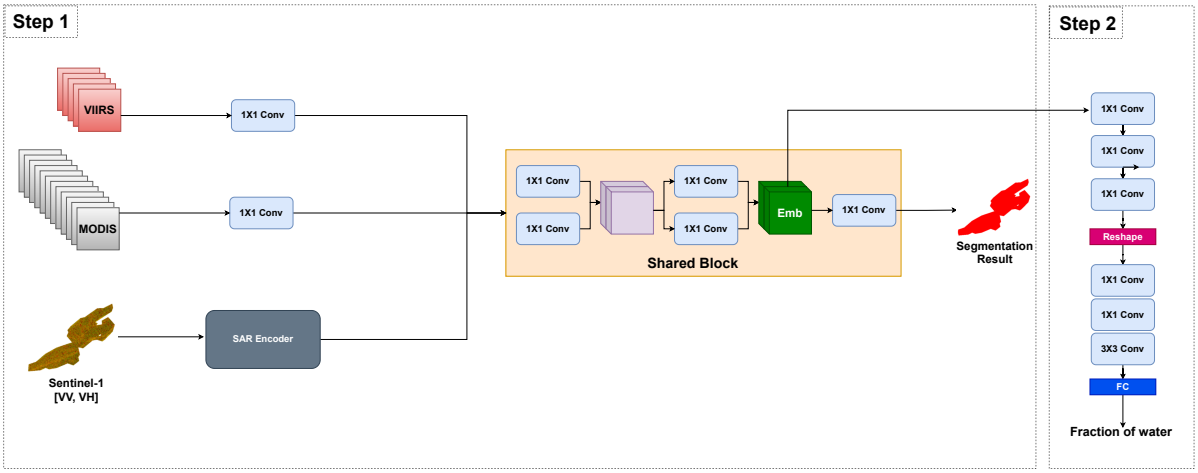


Figure 6.2.: Network structure of the proposed *2-step* model. In step 1, the shared block generates the satellite embedding (Emb). In step 2, the embedding learnt are fused to learn the per-day fraction of non-frozen (water) pixels. Effectively, two outcomes are predicted by the proposed model: 1) in step 1, the per-pixel semantic segmentation of the lake (red=*non-frozen*, blue=*frozen*, here we have shown an example non-frozen result for lake Sils), 2) in step 2, the point estimate of the daily fraction of water pixels. *Conv* and *FC* represent the convolution and fully connected layers respectively. Best if viewed on screen.

termediate representation, along with individual encoders (a.k.a. "embedding function") to transform the data of each sensor to that representation (the common "embedding", abbreviated as "Emb" in Fig. 6.2). The overall architecture consists of an individual branch per sensor to map different inputs to a shared "feature map" and a shared block with the same weights for all sensors to further abstract that feature map into the final representation that serves as the basis for the output predictions.

Encoders: At first, the input image is encoded into intermediate features that, on the one hand, can be derived from any of the input sensors and, on the other hand, preserve the information needed to differentiate the frozen and non-frozen states. For the MODIS and VIIRS branches, with their low spatial resolutions that do not call for texture analysis, we use a simple 1×1 convolution (conv) layer as encoder. On the contrary, the S1-SAR branch has much higher spatial resolution and must be able to learn texture and context features over spatial neighbourhoods of multiple pixels. Our encoder is a CNN inspired by U-net (Ronneberger et al., 2015), with leaky ReLU (Xu et al., 2015) activation. See Fig. 6.3.

Shared block: The per-sensor branches, which one might interpret as a "pre-processing" into a common feature space, are followed by a shared block to optimise the features for the downstream task of lake ice monitoring. That block consists of a shallow network with 1×1 convolution and concatenation layers, see Fig. 6.2. Also, in this block all weights are learnt from data. Its output is the deep embedding (Emb), with a uniform size of $12 \times 12 \times 32$. It serves as input for the second step, described in Section 6.3.1.

Training: The network weights in the first step of the 2-step model are trained on

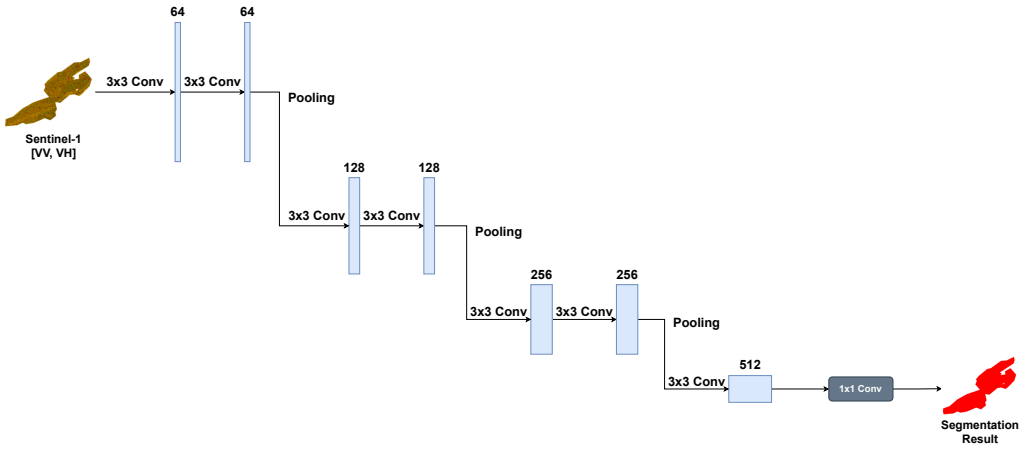


Figure 6.3.: U-net style SAR encoder architecture. Conv and FM represent convolution layer and feature maps respectively. Best if viewed on screen.

the auxiliary task of spatially explicit semantic segmentation, as follows:

- First, the weights in the U-net style SAR encoder are pre-trained, starting from "Glorot" uniform initialisation (Glorot and Bengio, 2010), by feeding the network with S1-SAR data. The leakage rate (slope for values <0) of the leaky ReLU is set as $\alpha = 0.1$.

- Second, the MODIS and VIIRS encoders are pre-trained, along with the shared block, again with "Glorot" uniform initialisation (Glorot and Bengio, 2010). Optical images are fed in alternating fashion, switching between one epoch with MODIS data and one epoch with VIIRS data. This procedure is run for 40 epochs (20 each for MODIS and VIIRS), thus priming the two branches and the shared block for MODIS-VIIRS fusion.

- Third, the shared block weights are fine-tuned to make them compatible also with SAR data. To that end the network is again fed with S1-SAR data, thus jointly updating the pre-trained SAR encoder and shared block. This last step tunes the shared block to also perform optical-SAR (MODIS-SAR, VIIRS-SAR) fusion.

In all above-mentioned training steps, the weights are optimised by minimising the *cross entropy loss* over pixel-wise class labels, using the *Adam* optimiser (Kingma and Ba, 2015). Details about hyper-parameters of the training are given in Table 6.3, columns 2-4.

In the proposed setup, the type and number of input branches is flexible. It is fairly straight-forward to include further satellite sensors, by designing a suitable encoder and fine-tuning the shared block to accommodate the new input data characteristics. In that context note that convolution layer can handle varying input sizes, so input images of varying size, respectively spatial resolution, can be used (in fact our inputs also do not have the same pixel dimensions).

Table 6.3.: Network parameter settings for various training stages in the proposed *2-step* model. Learning Rate (LR) decay refers to exponential scheduling. SGD represents stochastic gradient descent.

	2-step model			
	step 1 uNet-style SAR encoder pre-training	**step 1** MODIS, VIIRS encoder pre-training, Shared block pre-training	**step 1** Shared block fine-tuning	**step 2** Regression
Epochs	500	MODIS (20), VIIRS (20)	250	100
Batch size	16	8	16	4
Loss	cross entropy	cross entropy	cross entropy	L_{reg}
Optimiser	Adam	Adam	Adam	SGD
LR	5e-05	5e-04	1e-05	5e-04
LR decay (steps, rate)	375, 0.9	-	150, 0.9	-
window size	-	-	-	7

Step 2: prediction based on the learnt embedding

The second step in our application task is a regression model that predicts what fraction of the lake surface is covered by open water. This prediction is based on embedding from multiple adjacent days, in order to exploit temporal sequence information and to smooth out noise in the per-day features. The regression network starts with three convolutional layers that operate independently on each per-day feature map, followed by a reshape layer that aligns the information from a *window* of adjacent time steps (i.e., the time and channel dimensions are collapsed into a single dimension). Three more convolution layers then combine the information from all time steps in the window, and a final, fully connected (FC) layer returns a single, scalar prediction, assigned to the central day in the window. See Fig. 6.2.

The *window size* is a hyper-parameter that determines how many per-day embedding are stacked into a local time series. E.g., when it is set to 7, a daily prediction is based on feature maps from the previous 3, current and next 3 days.

For the regression model, the loss function L_{reg} is defined as

$$L_{reg} = L_{mse} + \beta L_{line} + \gamma L_{idc} \tag{6.1}$$

where L_{mse}, L_{line} and L_{idc} correspond to *mean squared error loss*, *line loss*, and *intra-day coherence loss* respectively. We empirically fixed the weights $\beta = 0.25$ and $\gamma = 0.08$. In step 2 also, the training starts from "Glorot" uniform weight initialisation (Glorot and Bengio, 2010). Further training settings for the regression model is shown in Table 6.3 (column 5).

L_{line} penalises the deviations from a linear trend ($mx + c$) computed locally for each mini-batch b. Each batch is carefully selected so that the embedding from adjacent days are present. The line loss is based on the assumption that the predictions of adjacent dates should lie on a straight line, which acts as a smoothness prior and mainly helps to

correct isolated outliers. First, the line is fitted as

$$m = \frac{y^l - y^f}{b}, \quad c = y^f \tag{6.2}$$

with y^f, y^l the first and last prediction in the mini-batch. The deviation d^i of the i^{th} prediction y^i from the line is then

$$d^i = \frac{|mi + c - y^i|}{\sqrt{m^2 + 1}} \tag{6.3}$$

There are a few days where more than one satellite imaged the region of interest. In such cases, multiple embedding (one per satellite) will be generated by step 1. The loss term L_{idc} computes the variance between predictions from the same day and penalises it, thereby encouraging coherent outputs. The final output (fraction of water) for a day with more than one sensor observation is the mean of the individual predictions.

6.4. Experiments, results and discussion

We perform experiments in the leave-one-out setting: the dataset is sub-divided into different portions, and in each run one portion is left out and a model is trained on the remaining data. The model thus trained is then tested using the left out portion. We use two settings: *Leave One Winter Out* (LOWO) and *Leave One Lake Out* (LOLO), to assess the model's capability to generalise across time and space. For instance, "LOLO-Sihl" means training on the data of all lakes except Sihl (from both winters 2016–17 and 2017–18) and testing on the data of Sihl from both winters, and similar for other combinations.

Using the proposed *2-step* model, we generate embedding on all available dates, from both non-transition and transition periods. To train the step 1 (semantic segmentation) which require per-pixel labels, we use only data from non-transition days, because pixel-wise ground truth is not available for the transition days. Step 2 (regression of water fraction) needs only scalar per-day labels, hence we include data from all days in the training set, despite the inevitably higher label noise during transitions (see Section 6.2.3).

Although an auxiliary task in our setup, we quantitatively assess segmentation performance, in terms of overall classification accuracy and mean per-class Intersection-over-Union (mIoU) score, see Section 6.4.1. For qualitative assessment, we visualise the learnt embedding in Section 6.4.3. As an additional check we plot and visually inspect the time series of predicted water fractions for each lake throughout the entire winter, see Section 6.4.2. From those time series, we also derive the ice-on/-off dates and assess the model's ability to retrieve lake ice phenology, in Section 6.4.4.

6.4.1. Quantitative results: semantic segmentation

To learn the embedding, our model is trained to perform pixel-wise semantic segmentation, which we assume to be a good proxy task for the retrieval of ice cover, but with a stronger, spatially more explicit and more detailed supervision signal. This experiment inspects the performance of the segmentation step (step 1). After training the model, the data from each satellite sensor is fed to the respective input branch and is analysed independently, for two reasons. On the one hand it allows us to separately quantify the performance of the model for each satellite sensor. On the other hand the embedding is trained jointly, but still computed features separately for each input image, and thus also for each sensor – the mixing of features extracted from different sources only happens when aggregating over time windows in the regression step. LOWO and LOLO results are shown in Tables 6.4 and 6.5, respectively. To quantify the epistemic uncertainty (model uncertainty due to imperfect training data), we also estimate the standard deviation of the predictions, using an ensemble of five models with independent random initialisations. The results support our assumption that a joint model with a single, shared embedding can handle inputs from any single sensor. In more detail, it can be seen that the proposed model achieves very good generalisation across different winters (mIoU >76.1%, accuracy>94.6%). Overall, also the generalisation across different lakes is very good for optical images (mIoU >79.8%, accuracy>91.6%), with a small decrease for the S1-SAR data (mIoU >60.7%, accuracy>91.3%).

Table 6.4.: Semantic segmentation results (in %) of *leave one winter out* experiment. μ and σ denote the mean and standard deviation for the results obtained using five random initialisation.

Data	**Winter 2016–17**				**Winter 2017–18**			
	Accuracy		*mIoU*		*Accuracy*		*mIoU*	
	μ	σ	μ	σ	μ	σ	μ	σ
MODIS	96.1	0.1	81.2	0.4	97.5	0.5	84.9	1.7
VIIRS	98.9	0.2	87.9	1.9	99.1	0.1	87.4	1.1
S1-SAR	94.6	0.3	76.1	1.0	95.2	0.1	79.1	0.4

Table 6.5.: Semantic segmentation results (in %) of *leave one lake out* experiment. μ and σ denote the mean and standard deviation for the results obtained using five random initialisation.

Data	**Sihl**				**Sils**				**Silvaplana**				**St. Moritz**			
	Accuracy		*mIoU*		*Accuracy*		*mIoU*		*Accuracy*		*mIoU*		*Accuracy*		*mIoU*	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
MODIS	91.6	2.5	79.8	3.3	98	0.2	89.2	0.9	98.1	0.3	85.7	2.0	99.8	0	80.9	1.6
VIIRS	98.3	0.3	87.8	1.8	99.5	0.1	90.9	1.3	99.5	0.1	89.3	1.6	-	-	-	-
S1-SAR	91.3	0.6	60.7	1.4	92.2	0.5	70.3	1.4	92.5	0.4	72.6	1.1	93.9	0.3	72.8	1.1

6.4.2. Qualitative results: full winter time series

In each winter, after fusing the embedding from all three satellite sources, daily prediction of the fraction of water (non-frozen pixels) are obtained for each lake and the time series

is generated. Fig. 6.4 shows the sample results for lakes Sihl and Sils. Here, the model is trained using the data from all the lakes from winter 2017–18. In each sub-figure, the predicted fraction of water is shown on the y-axis against the acquisitions (dates) on the x-axis displayed in chronological order from beginning of September until end of May. Reference annotations are shown as a blue line, predictions are displayed as points with a different colour per calendar month. Results for the remaining two lakes are shown in Appendix C.2.

It can be inferred from Figs. 6.4 and C.1 (Appendix C) that the time series results are good for the bigger lakes Sihl and Sils, satisfactory for Silvaplana, but not very good for lake St. Moritz, especially during freeze-up. As said earlier, that lake is tiny (not a single clean pixel in VIIRS, four clean pixels in MODIS) and has been included as an extreme case to test the limits of high temporal-resolution satellite data for small lakes. As expected, the results generally improve with increasing lake size, see also Table C.1 (Appendix C). There are several possible reasons. For low resolution sensors, there are simply few pixels to learn the model, whereas for higher-resolution sensors like S1-SAR there may be just enough pixels, but still only little context. Moreover, the portion covered by mixed pixels along the lake boundary is relatively larger, and such boundary effects are compounded by the fact that the mixed pixels are most affected by residual geo-localisation errors. Moreover, quantisation and correspondence issues may play a role when transferring ground truth annotations from webcams to satellite images, see also Section 6.2.3. Finally, note that for St. Moritz only two sensors were fused as there are no clean VIIRS pixels.

6.4.3. Qualitative visualisation: learnt embedding

We visualise the learnt embedding after dimension reduction with t-distributed Stochastic Neighbour Embedding (Maaten and Hinton, 2008, t-SNE), for a visual impression of how the data cluster in high-dimensional feature space. See Figs. 6.5 and C.2 (Appendix C). In the illustration, the blue corresponds to a high likelihood of ice/snow, while yellow means a high likelihood of water. It can be seen that the learnt embedding looks reasonable, in the sense that the classes are separable. Reducing the dimension of the projection to only 2 (right column) reveals that clusters for different sensors can still be restored. I.e., the joint feature spaces allows one to construct decision boundaries, respectively regression coefficients, that are invariant w.r.t. the input type, but the features themselves are not completely invariant.

6.4.4. Ice-on/off results

To compare the performance of the proposed approach with previous works which analysed the target lakes (Tom et al., 2020a,c, 2019), we extract the ice-on/off dates from the daily predictions of water fraction, see Table 6.6. For robustness, those daily predictions are again ensemble over five independently initialised networks. Furthermore, we perform a comparison with the in-situ temperature analysis reported in Tom et al. (2019). The comparisons are only possible for the winter 2016–17, as no other estimates are available for 2017–18.

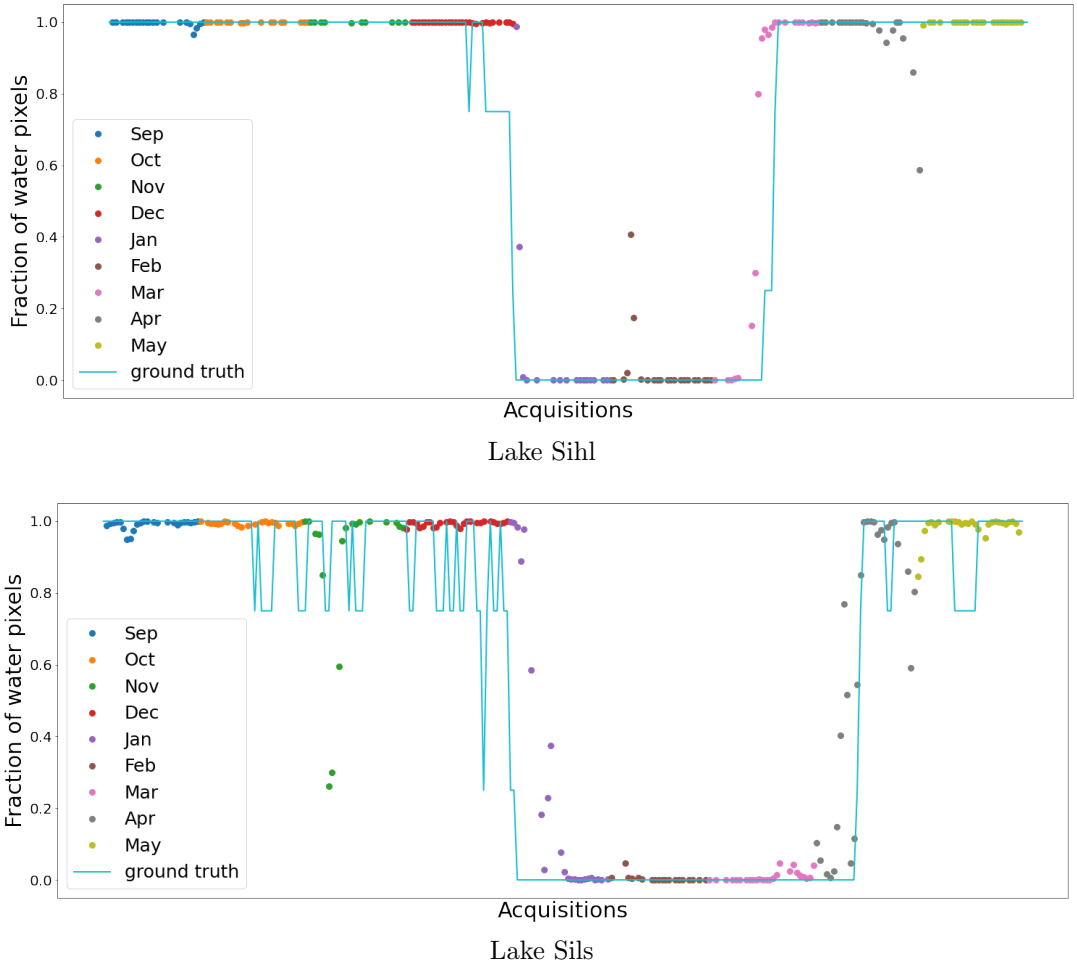
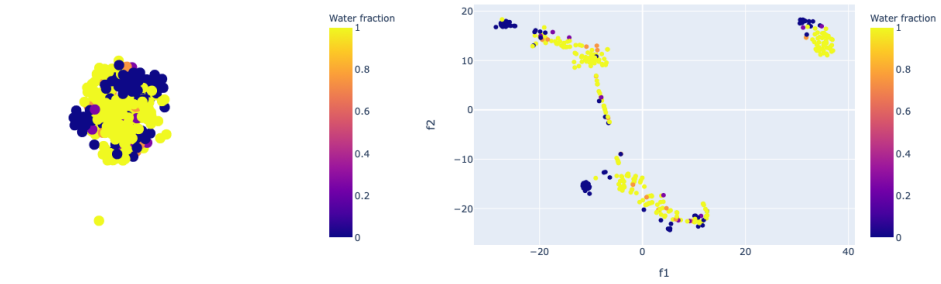


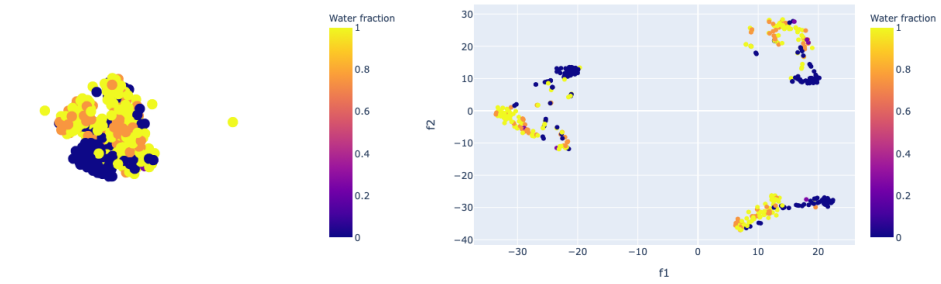
Figure 6.4.: Time series plots for lakes Sihl and Sils from winter 2016–17 using a model trained on all the data (all four lakes) from winter 2017–18. Best if viewed on screen.

Following Tom et al. (2021b), we opt for a higher freeze threshold of 30% to estimate the two LIP events. I.e., ice-on is detected when <30% of the lake is not frozen, and ice-off when >30% of it return to the non-frozen state after the freeze period. The choice of threshold generally does not make a big difference, but the more conservative setting of 30% is sometimes beneficial to compensate the higher uncertainty in remote sensing estimates, as opposed to in-situ observations. For completeness we also show the results for a threshold value of 10%. See Table 6.6. In the Table, "M+V" refers to a decision-level fusion of MODIS and VIIRS predictions obtained with Support Vector Machines, as reported in Tom et al. (2020c). Similarly, "Webcam" means the prediction results from webcam images, obtained with the *Deep-U-Lab* network (Tom et al., 2020c). The ground truth is determined by interactive visual interpretation. Outputs that meet the GCOS target of ±2 days (relative to the ground truth annotations) are printed bold.

When comparing the results on Table 6.6 from different satellite remote sensing methods (Tom et al., 2020a,c) methods, the joint embedding model, on average, deviates least from the ground truth. However, the results of in-situ temperature analysis (Tom et al., 2019), are better. As before, the predictions for the two (relatively) larger lakes Sihl



Lake Sihl, dimension=3 Lake Sihl, dimension=2



Lake Sils, dimension=3 Lake Sils, dimension=2

Figure 6.5.: *t*-SNE representation of the embedding learnt (lakes Sihl and Sils, winter 2016–17) using the proposed *2-step* approach with a model trained on all the data from winter 2017–18. Water fraction refers to the predicted fraction of non-frozen pixels. Best if viewed on screen.

and Sils are more accurate. The accuracy is inversely proportional to the lake area, see also Table C.1 (Appendix C). For the tiny lake St. Moritz, all satellite-based predictions of ice-on are wildly off, again showing the limits of satellite remote sensing and/or statistical machine learning for very small geographic objects.

6.4.5. Runtime

On a computer equipped with NVIDIA GeForce GTX 1080 Ti graphics card (12 GB), it takes ≈1 hour to train the whole 2-step model (including all pre-training steps) for the LOWO 2016–17 experiment. Predictions for a complete winter take only a couple of minutes.

6.5. Discussion and conclusion

We have developed a deep learning framework that learns a joint embedding in order to fuse MODIS, VIIRS and S1-SAR satellite data into a task-specific, homogenised time series. We have employed the proposed framework for lake ice monitoring, to estimate the freezing state of four Swiss lakes and to determine their ice-on and ice-off dates.

Table 6.6.: Ice-on/off dates (winter 2016–17) estimated with the *2-step* model, with two different thresholds. For comparison we show the ground truth (in chronological order when more than one candidate exists), the earlier remote sensing results: M+V (Tom et al., 2020c), Webcam (Tom et al., 2020c), S1-SAR (Tom et al., 2020a), and the results of in-situ (temperature, T) analysis (Tom et al., 2019). Results that meet the GCOS requirement are printed bold.

Date	**Ground truth**	***2-step* (10%)**	***2-step* (30%)**	**M+V**	**Webcam**	**S1-SAR**	**In-situ (T)**
Ice-on (Sihl)	1 Jan	**3 Jan**	**3 Jan**	**3 Jan**	4 Jan	11 Jan	28-29 Dec
Ice-off (Sihl)	14 Mar 15 Mar	6 Mar	10 Mar	10 Mar	14 Feb	28 Feb	**16 Mar**
Ice-on (Sils)	2 Jan 5 Jan	15 Jan	9 Jan 15 Jan	**6 Jan**	-	**6 Jan**	**31 Dec**
Ice-off (Sils)	8 Apr 11 Apr	31 Mar **6 Apr** **11 Apr**	**7 Apr** **12 Apr**	31 Mar	-	**12 Apr**	**10 Apr**
Ice-on (Silvaplana)	12 Jan	16 Jan	15 Jan	15 Jan	-	15 Jan	**14 Jan**
Ice-off (Silvaplana)	11 Apr	7 Apr **12 Apr**	7 Apr **13 Apr**	30 Mar	-	18 Mar	14 Apr
Ice-on (St. Moritz)	15-17 Dec	17 Jan 27 Jan	6 Jan 16 Jan 25 Jan	1 Jan	**14 Dec**	25 Dec	**17 Dec**
Ice-off (St. Moritz)	30 Mar - 6 Apr	**7 Apr**	**7 Apr**	**7 Apr**	18 Mar-26 Apr	**30 Mar**	**5-8 Apr**

In our experiments the method has shown good generalisation across different winters and across different lakes.

Notably, the data we fuse for our application example not only come from two different imaging modalities, but also differ in resolution by an order of magnitude. At least conceptually, the proposed multi-branch architecture is very generic. It should be straightforward to extend it to further satellite sensors, by pre-training the corresponding low-level branches and fine-tuning the shared block.

Although the proposed approach in many cases achieves promising results, there is still room for improvement. In particular, we have used only rough, approximate reference annotations for the transition days. Better ground truth will likely improve the performance of the regression. Moreover, we expect that adding optical satellite images with higher spatial resolution and not too long revisit times could improve the results. Given the statistical nature of our approach, it is also likely that a larger and more diverse training set would bring performance gains, but we leave this for future work, as it will require annotations for additional lakes and/or winters. An interesting, but challenging direction may be to fuse in also data from non-satellite sources (webcams, UAVs, etc.).

A caveat of our approach is that the the learnt embedding at this stage is perhaps better described as *equivariant*, in the sense of delivering comparable output for different image sources. But it is not truly invariant, as the latent vectors to some degree still appear to cluster according to the different sensors.

In terms of application, while our focus was on lake ice, the proposed data fusion methodology is generic and should be easily transferable to other geo-spatial data analysis and Earth science applications.

Finally, at the methodological level, it may in the longer term be necessary to shift the focus to techniques that learn task-oriented data representations (i.e., embedding) with minimal supervision, such as unsupervised, weakly supervised or self-supervised learning. There exists a huge amount of unlabelled remotely sensed data, from various sources (spaceborne, airborne, webcams, amateur images, etc.), and labelling reference data for supervised learning is a critical bottleneck.

6.6. Acknowledgements

The authors would like to thank Roberto Aguilar (Aerialytics, Costa Rica) for his support.

7 Conclusions

7.1. Summary

This thesis proposed a future-oriented system for automated visual monitoring of lake ice from space (satellites) and Earth (terrestrial webcams). The analysis primarily focused on four selected small-to-medium-sized mid-latitude lakes located in the Swiss Alps. Data from different sensors were integrated into the proposed system and analysed, such as optical (MODIS, VIIRS) and radar (Sentinel-1 SAR) satellites, and close-range webcams, which constituted a stable data source. A comparison of the various attributes of these data is shown in Table 7.1.

Table 7.1.: Intercomparison of the properties of different input data.

	MODIS	**VIIRS**	**Sentinel-1 SAR**	**Webcams**
Spatial resolution	250–1000m	375–750m	10m (pixel spacing)	ca. 4mm to 4m
Temporal resolution	1d	1d	1.5–4.5d (for Switzerland)	1 hour (typically)
Spectral resolution	36 bands (0.4–14.2 μm)	22 bands (0.4–12.0 μm)	C-band, (around 5.6 cm) 4 polarisations (mainly VV, VH)	RGB
Cloud problems	severe	severe	nil	negligible
Cloud mask issues	slight	slight	NA	NA
Costs	free	free	free	free
Availability	very good (via VIIRS continuity)	very good	very good (HV / HH only partially available)	depending on location
Sensor type	optical	optical	radar	optical

In the processing chain, state-of-the-art machine learning methodologies were deployed for lake ice monitoring. Such learning-based techniques have rarely been investigated in the past, especially for monitoring various small mountain lakes on the planet, which are the ones that freeze besides the Arctic lakes. Additionally, the proposed approaches have a big advantage given the fact that they are largely independent of in-situ observations and models of the lake freezing and thawing processes. Spatio-temporal ice maps were derived by casting lake ice retrieval as a supervised semantic segmentation problem, and the critical LIP events were estimated using these maps. SVM-based classical ML methodology was applied on MODIS and VIIRS sensors, while deep CNN was employed for S1-SAR and webcam data. Different aspects of the various single-sensor methodologies that were proposed are compared in Table 7.2. Furthermore, the fusion

of several satellite data (SAR and optical) was also studied. Detailed experimentation was performed, including cross-validation and spatio-temporal generalisation tests. Reference data was primarily generated by interactive visual interpretation of freely available webcam images. Technically, a huge amount of labelled data is needed to train all proposed DL models from scratch. However, the training load was reduced with the transfer learning trick, and only fine-tuning on the relatively smaller dataset was needed.

Table 7.2.: Intercomparison of the proposed methodologies for each input data type.

	MODIS, VIIRS	**Sentinel-1 SAR**	**Webcams**
Training complexity	low	medium (transfer learning decreases training load)	medium (transfer learning decreases training load)
Computation complexity	very low	high, GPU required	high, GPU required
Per-pixel ground-truth	not necessary (fully-frozen or fully-non-frozen days used)	not necessary (fully-frozen or fully-non-frozen days used)	necessary
Pre-training on external data	not required	required	required
Near-real time performance	yes	feasible (excluding training)	feasible (excluding training)

For both MODIS and VIIRS sensors, the proposed SVM-based approach (Chapter 3) achieved mIoU scores $> 93\%$. It also marked good generalisation performance with mIoU scores $> 78\%$ and $> 80\%$ across lakes and winters, respectively. Additionally, using the same approach, a large time series (since 2000) of MODIS data was established to study the LIP trends in two selected lakes in Switzerland (Chapter 4). As expected, later freeze-up, earlier break-up, and shorter freeze duration trends were estimated from this 20-winter time series analysis. For the lakes Sils and Silvaplana, the change in complete freeze duration were determined as -0.76 d/a and -0.89 d/a, respectively. In addition, plausible correlations of LIP events and climate indicators such as temperature, sunshine, and precipitation were observed. It was noticed that the mean winter temperature and total winter sunshine have a strong negative correlation with the freeze duration events. With empirical evidence, Chapters 3 and 4 showed that the classical ML models are sufficient to accurately extract the spatio-temporal extent of lake ice in small-to-medium-sized lakes from low-spatial resolution optical satellite imagery.

It was shown in Chapter 3 that terrestrial webcams could efficiently monitor very small lakes such as St. Moritz that cannot easily be surveyed using VIIRS type data. A state-of-the-art has been established for webcam monitoring using the proposed *Deep-U-Lab* network. On average, the *Deep-U-Lab* methodology reached mIoU values of $\approx 87\%$ with generalisation scores of $\approx 71\%$ and $\approx 69\%$ across different cameras and winters, respectively. Furthermore, automated detection of lake outlines from webcams was presented using the same network. Moreover, a new freely available webcam dataset with per-pixel ground truth information has been published.

Chapter 5 investigated and demonstrated that deep neural networks could learn to observe lake ice from SAR satellite data. The *deeplab v3+* model on S1-SAR data produced good generalisation performance across lakes and winters and attained mIoU scores $> 84\%$ for the most challenging lake, and $> 90\%$ on average. One key finding of this chapter is that the weights of a deep neural network trained on close-range optical amateur photographs can be reused as a starting point to train a network that works well on radar satellite data.

Besides single-sensor approaches, a data fusion methodology was presented in Chapter 6 which showed that optical satellite data, supplemented by SAR imagery, may be a good trade-off for operational estimation of lake ice from space. The proposed deep satellite data fusion technique achieved scores that are of practical interest (accuracy and mIoU scores $>91\%$ and $>60\%$ respectively) in addition to attaining good cross-winter (mIoU $>76\%$) and promising cross-lake (mIoU $>60\%$) generalisation results. Furthermore, it marked a new state-of-the-art in estimating the critical ice-on and ice-off dates, in many cases meeting the GCOS temporal resolution criteria.

7.2. Limitations

The quantitative and qualitative analysis presented in this dissertation has shown that machine learning is a valuable and practical tool for lake ice monitoring from terrestrial and satellite sensors. However, there are some shortcomings too, and this thesis would be incomplete without discussing them. The following sub-sections discuss the limitations of the input data and proposed methodologies.

Optical satellite data analysis. Compared to the non-transition days, the performance of the proposed optical satellite methodology (Chapters 3, 4) is less reliable during the transition days, which is also reflected in the accuracy of estimation of the ice-on and ice-off dates. The main reason is that the training set has only data from non-transition days due to the unavailability of pixel-accurate annotations on the transition days. When clouds occur during the transition days, this issue stands out. A second weakness is that, to some extent, the errors in MODIS and VIIRS cloud masks could introduce discrepancies in the final results. Furthermore, for small lakes with very few clean pixels, absolute geolocation errors could become critical. Note that, even after correction, an error of up to half a pixel remains, which is not negligible when the lake itself is just a handful of pixels.

Webcam image analysis. The proposed webcam-based solution has some caveats, too. A major drawback is that it is very difficult or nearly impossible to scale up this technology to country- or planet-scale, unlike the satellite-based methodology. Though the webcam network is quite good for countries like Switzerland, it is not always the case for other parts of the globe. Even in Switzerland, at present, not all lakes are monitored by publicly available webcams. Even for the lakes with webcam surveillance, the number of cameras installed might be very few, and it is not ensured that the complete lake

area is covered, making it challenging to observe large lakes with webcams. Furthermore, since the public webcams were not installed targeting lake ice monitoring, there is limited control over their position, orientation, lake area coverage, image quality etc. Moreover, the spectral resolution of these freely available cameras is just satisfactory compared to satellites. Besides that, webcams are not useful in heavy fog situations. However, for the target lakes, intense fog scenarios occur only during very few days in a winter. Furthermore, on the two tested lakes, the cross-lake generalisation performance of *Deep-U-Lab* methodology is below par.

As an extension of the webcam methodology, we have published some preliminary results regarding the usage of crowd-sourced social media images, refer to Prabha et al. (2020), however, concluded that it is pretty hard and perhaps unrealistic for operational monitoring. The larger the lake area the worse since only the area close to the lake shore is (usually) visible in these images. In addition, it is difficult to spot the lake shoreline in such images, especially on snowy days. Furthermore, the field of view and lake coverage vary significantly from one image to another. It was identified that ensuring good lake coverage and usable geo-reference are probably the show-stoppers. Such crowd-sourced images are available mainly for touristic lakes and not necessarily for all the lakes that freeze. Additionally, it was noticed that the geotags of these images are not always available[1], and even if it is, they are not always correct[2].

Snow-free ice. In the two winters with webcam-based ground truth available, it was observed that there are only very few days with snow-free ice compared to snow and open water, effectively creating class imbalance. If no class balancing measures are taken in such a scenario, a learning-based system that assumes that the observed distribution is representative tends to be biased to the well-represented classes. The presence of some noise in the ground truth could make this problem worse. Consequently, the proposed models sometimes find it difficult to distinguish between thin ice and water. Note that, even for human experts, it is difficult to differentiate between thin snow-free ice and open water in some cases.

7.3. Lessons learnt

The methodologies, experiments and results presented in this dissertation are the outcomes of several years of research work. Some lessons learnt during these years are discussed in the following sub-sections.

Lake ice detection using low spatial resolution optical satellites. One of the main advantages of satellite-based monitoring, as opposed to webcam-based approaches, is the ability to survey much larger areas in a single scan. However, in both MODIS and VIIRS imagery, the target lakes are represented only by very few clean pixels. Hence,

[1] mountains silhouettes or skylines may be used for accurate geo-referencing, but this is still an active area of research

[2] usually provides coordinates but not the orientation of the optical axis

there is not much spatial context to learn. Therefore, sophisticated networks such as CNNs that could take advantage of the spatial neighbourhood relationships were not needed, and hence shallow classifiers that operate per-pixel were used.

20-winter time series analysis. One key advantage of optical satellite imagery is the availability of a large MODIS data archive with continuity guaranteed by VIIRS. This thesis has shown that it is possible to learn the LIP trends from the 20-year long time series of MODIS data. Some meaningful correlations were found between the derived LIP trends and the regional climate variables. Even though a two-decade period is not enough to draw solid conclusions, some preliminary inference can be made on the freeze patterns in Swiss lakes. An interesting observation from Chapter 4 is that the proposed, dedicated monitoring system was able to retrieve lake ice more accurately than the operational MODIS and VIIRS snow products, despite the relatively high correlation between freezing of lakes and snow cover. Though the analysis was limited to two lakes in the Oberengadin valley located in the south-eastern part of Switzerland, the trends are not expected to be too different for other lakes with similar characteristics situated in the Swiss and European Alps.

Lake ice monitoring with webcams. The proposed webcam-based *Deep-U-Lab* approach showed very good ice and snow detection performance, especially during the transition periods by capturing the finer lake ice dynamics, which were difficult to achieve using satellite approaches. Cloud problems with webcams are negligible as opposed to optical satellites. Additionally, the spatial (if not placed far from the lake) and temporal resolutions are higher than S1-SAR. Webcams are valuable sensors particularly useful to monitor very small lakes that cannot be observed well using low- and medium-spatial resolution optical satellite sensors. Regardless of the meagre image quality, the CNN-based webcam algorithm presented in Chapter 3 attained good performance even in challenging scenarios such as variable solar illumination conditions and in the presence of shadows (from clouds, mountains). In a nutshell, though satellites are the best choice for country-wide, planet-scale coverage for lake ice retrieval, webcams are valuable for meticulous, regional monitoring purposes.

Lake ice detection from Sentinel-1 SAR. Among the satellite sensors investigated in this thesis, the results suggest that S1-SAR is the best solution for lake ice monitoring as a stand-alone sensor. Firstly, the microwave sensor can penetrate clouds, preventing data loss on cloudy days compared to any optical satellite. Secondly, due to the low GSD of the S1-SAR sensor, unlike MODIS and VIIRS, there is no scarcity of pixels to deploy the *deeplab v3+* network. Thirdly, though the temporal resolutions of MODIS and VIIRS are good enough for lake ice monitoring, the corresponding spatial resolutions are just satisfactory. On the contrary, S1-SAR images have very good spatial resolution. Hence, the CNN learns better context and texture features by grasping the pixel neighbourhood relationships, effectively marking excellent performance for lake ice monitoring. However, note that, the 1.5 days (best case) repeat cycle for the target regions involves different

incidence angles. Moreover, not all regions on Earth are captured from both ascending and descending orbits.

Satellite embedding learning. This thesis proposed various single satellite sensor approaches for lake ice observation using MODIS, VIIRS (Chapters 3, 4) and S1-SAR (Chapter 5). However, one key conclusion is that it is difficult to satisfy the GCOS temporal resolution criteria using such approaches. A hybrid approach that fuses data from multiple satellites is one of the possible solutions to this problem. However, fusing data from multiple optical satellites alone is risky, especially if clouds occur near the critical dates during the vital freeze-up and break-up periods. Such cases can compromise the estimation accuracy of the critical LIP dates. Clouds are unavoidable in the Alps and many other mountain areas during the winter season when the lakes freeze, and their occurrence can be random. Hence, multi-sensor data fusion is not just a recommendation but a necessity. Perhaps the most reliable solution for satellite-based lake ice monitoring is a combination of SAR and optical data. Therefore, this thesis has presented an integrated method (Chapter 6) that compensated for the shortcomings of each satellite sensor by learning a joint feature space representation.

7.4. Outlook

The following paragraphs discuss the outlook and promising future directions to drive the lake ice monitoring research based on the findings of this dissertation.

Additional training data and better generalisation. The approaches presented in this thesis attained good generalisation performance across lakes, winters and cameras (for webcams). However, there is still room for improvement, especially across lakes for webcams. Re-training (or fine-tuning) of the models might be needed when the proposed methods are applied on the data from new lakes (or winters, but probably less tuning might be required for winters as opposed to lakes). The cross-winter variations are assumed to be barely correlated. Hence, the data from probably a handful of winters might represent the expected range of conditions. On the other hand, any small set of lakes may still be biased by the geographic conditions and not generalise well to different locations. To some extent, the generalisation error might prevail in all bottom-up data-driven learning-based systems unless the learnt model fully covers the entire input distribution (including corner cases). However, this is not easily attainable. This issue can be minimised by incorporating multi-winter data (including also pixel-accurate annotated data from transition dates) in the training set (also from more lakes across the globe). Another workaround is to strengthen the approaches for better domain adaptation capabilities.

Reduced supervision. Though various ML approaches were used in this thesis, the analysis was limited to supervised learning methodologies which needed labelled training data. As already discussed, the ground truth generation process based on visual image

interpretation is laborious and time-consuming. Reducing the supervision with the help of semi-supervised learning, active learning, or data-driven domain adaptation will abate the need for large amounts of annotated data (especially to train deep networks) and hence looks promising. With such an extension, the untapped potential of huge unlabelled archives of remote sensing data can be utilised to improve the quality of the trained model. In addition, data from meteorological stations (e.g., temperature) in the vicinity of the lakes can be used to perform further checks and to fill gaps in the existing reference data.

More hybrid models. The hybrid model proposed in Chapter 6 that learns satellite embedding could potentially serve as a basis for an operational product that closely meets the GCOS specification. It would be intriguing to fuse even more satellite data such as Sentinel-2, Landsat-8 etc. Though such freely available satellites will not be effective as stand-alone data sources due to their relatively poor temporal resolution, they might still be useful for closing gaps in the existing time series. Another possible option for data fusion is the Planet Cubesat satellite constellation. An even more challenging but fascinating research direction is to extend the proposed satellite embedding learning approach to fuse images from webcams and satellite imagery.

UAVs for monitoring. The addition of more sensors such as UAVs (furnished with both RGB and thermal cameras) for lake ice monitoring would be exciting. Drones could be considered as a more professional version of webcams with better control, viewpoint advantages etc. It would be straightforward to adapt the proposed *Deep-U-Lab* network for UAV-captured images. However, there are certain bottlenecks such as reliability issues (especially with accurate geo-referencing), reduced flight time in cold weather, area coverage (only feasible for small lakes), permission/security/privacy issues etc. Additionally, even though the data acquisition flights are largely automated, there is still a lot of manual work involved during the field campaigns.

Global-scale analysis. In Chapter 4, the MODIS data from 20 winters was used to derive the LIP trends in two Swiss mountain lakes. Even though this thesis focused on Swiss regional lakes, the proposed optical satellite-based approach is straightforward. Hence, it could be easily adapted to similar lakes in Switzerland and nearby countries (and probably lakes from other geographical regions). One possible future direction is the automated extraction of the 20-winter LIP patterns from more lakes that freeze in the Northern Hemisphere (e.g. lakes in Canada and Finland, high mountain Asia etc.). It would be interesting to observe whether the global trends also align with the Swiss LIP patterns. An added advantage is that it will test the robustness of the proposed ML-based approach for operational monitoring. However, more training might be needed to adapt the proposed models on totally new data from unseen geographic regions.

Physics-informed ML. One of the main problems with the current artificial intelligence approaches for Earth observation applications is the lack of explainability and

interpretability, which primarily stem from the bottom-up modelling approach. Inducing a balance between the top-down knowledge-driven numerical approaches and the bottom-up data-driven learning-based approaches has so far been less explored. A good trade-off is physics-inspired ML that incorporates the advantages of both top-down and bottom-up methodologies (Camps-Valls et al., 2018, 2020). It will be beneficial if such approaches could lead to more efficient designs which either reduce the amount of training data needed or achieve better domain adaptation. For the specific application of lake ice monitoring, the physics of ice formation and melting, if incorporated in the loss function or as additional constraints, could improve the accuracy during the transition periods.

Climate change quantification. One main goal of this thesis was to accurately quantify lake ice, as a basis for trend analysis and climate studies, using machine learning. In Chapter 4, strong connections between the LIP trends and regional climate variables were identified. Such trends could be useful as a complementary cue to confirm or question quantitative climate change analysis, especially at the local to regional scale. Accurate climate change and global warming assessment will be useful in the effective design of climate change mitigation strategies and policies.

A | Bibliography

Aksakal, S.K., 2013. Geometric Accuracy Investigations of SEVIRI High Resolution Visible (HRV) Level 1.5 Imagery. Remote Sens. 5, 2475–2491.

Antonova, S., Duguay, C., Kääb, A., Heim, B., Langer, M., Westermann, S., Boike, J., 2016. Monitoring bedfast ice and ice phenology in lakes of the Lena river delta using TerraSAR-X backscatter and coherence time series. Remote Sens. 8, 903.

Arp, C., Jones, B., Grosse, G., 2013. Recent lake ice-out phenology within and among lake districts of Alaska, U.S.A. Limnol. Oceanogr. 58, 2013–2028.

Barbieux, K., Charitsi, A., Merminod, B., 2018. Icy lakes extraction and water-ice classification using Landsat 8 OLI multispectral data. Int. J. Remote Sens. 39, 3646–3678.

Beyene, M.T., Jain, S., 2018. Freezing degree-day thresholds and Lake ice-out dates: Understanding the role of El Niño conditions. Int. J. Climatol. 38, 4335–4344.

Boureau, Y.L., Ponce, J., Lecun, Y., 2010. A Theoretical Analysis of Feature Pooling in Visual Recognition, in: International Conference on Machine Learning, Haifa, Israel, 21 - 24 June.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Brown, L.C., Duguay, C.R., 2010. The response and role of ice cover in lake-climate interactions. Prog. Phys. Geogr. 34, 671–704.

Brown, L.C., Duguay, C.R., 2011. The fate of lake ice in the North American Arctic. Cryosphere 5, 869–892.

Brown, L.C., Duguay, C.R., 2012. Modelling Lake Ice Phenology with an Examination of Satellite-Detected Subgrid Cell Variability. Adv. Meteorol. 6, 431–446.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language Models are Few-Shot Learners, in: Advances in Neural Information Processing Systems, Virtual, 6-14 December.

Burda, N., 1999. Computer simulation for Ladoga Lake ice dynamics based on remotely sensed data, in: International Geoscience and Remote Sensing Symposium, Hamburg, Germany, 28 June - 2 July.

Bürgmann, T., Koppe, W., Schmitt, M., 2019. Matching of TerraSAR-X derived ground control points to optical image patches using deep learning. ISPRS J. Photogramm. Remote Sens. 158, 241–248.

Cai, Y., Ke, C.Q., Li, X., Zhang, G., Duan, Z., Lee, H., 2019. Variations of lake ice phenology on the Tibetan Plateau from 2001 to 2017 based on MODIS data. J. Geophys. Res. Atmos. 124, 825–843.

Cai, Y., Ke, C.Q., Yao, G., Shen, X., 2020. MODIS-observed variations of lake ice phenology in Xinjiang, China. Clim. Change 158, 575–592.

Camps-Valls, G., Martino, L., Svendsen, D.H., Campos-Taberner, M., Muñoz-Marí, J., Laparra, V., Luengo, D., García-Haro, F.J., 2018. Physics-aware Gaussian processes in remote sensing. Appl. Soft Comput. 68, 69–82.

Camps-Valls, G., Svendsen, D.H., Cortés-Andrés, J., Álvaro Moreno-Martínez, Pérez-Suay, A., Adsuara, J., Martín, I., Piles, M., Muñoz-Marí, J., Martino, L., 2020. Living in the Physics and Machine Learning Interplay for Earth Observation. arXiv preprint, arXiv:2010.09031.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2015. Semantic image segmentation with deep convolutional nets and fully connected CRFs, in: International Conference on Learning Representations, San Diego, USA, 7-9 May.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40, 834–848.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint, arXiv:1706.05587.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018b. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: European Conference on Computer Vision, Munich, Germany, 8 – 14 September.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 13 – 17 August.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions, in: International Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 21 - 26 July.

Clevert, D.A., Unterthiner, T., Hochreiter, S., 2016. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), in: International Conference on Learning Representations, San Juan, Puerto Rico, 2 - 4 May.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding, in: International Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 26 June - 1 July.

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Mach. Learn. 20, 273–297.

Crétaux, J.F., Merchant, C., Duguay, C., Simis, S., Calmettes, B., Bergé-Nguyen, M., Wu, Y., Zhang, D., Carrea, L., Liu, X., Selmes, N., Warren, M., 2020. ESA Lakes Climate Change Initiative (Lakes_cci): Lake products, Version 1.0. Centre for Environmental Data Analysis. Available online: `http://dx.doi.org/10.5285/3c324bb4ee394d0d876fe2e1db217378` (accessed 10 July 2021).

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: International Conference on Computer Vision and Pattern Recognition, Miami, USA, 20 - 25 June.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, USA, 2 - 7 June.

Douglas, D.H., Peucker, T.K., 1973. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. Cartographica 10, 112–122.

Du, J., Kimball, J.S., Duguay, C., Kim, Y., Watts, J.D., 2017. Satellite microwave assessment of Northern Hemisphere lake ice phenology from 2002 to 2015. Cryosphere 11, 47–63.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12, 2121–2159.

Duguay, C., Bernier, M., Gauthier, Y., Kouraev, A., 2015. Remote sensing of lake and river ice, in: Tedesco, M. (Ed.), Remote Sensing of the Cryosphere. Wiley-Blackwell, Oxford, UK, pp. 273–306.

Duguay, C., Prowse, T., Bonsal, B., Brown, R., Lacroix, M., Menard, P., 2006. Recent trends in Canadian lake ice cover. Hydrol. Process. 20, 781–801.

Duguay, C., Wang, J., 2019. Advancement in bedfast lake ice mapping from Sentinel-1 SAR data, in: International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July - 2 August.

Duguay, C.R., Lafleur, P.M., 2003. Determining depth and ice thickness of shallow sub-Arctic lakes using space-borne optical and SAR data. Int. J. Remote Sens. 24, 475–489.

Ehlers, M., Klonus, S., rAstrand, P.J., Rosso, P., 2010. Multi-sensor image fusion for pansharpening in remote sensing. Int. J. Image Data Fusion 1, 25–45.

Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2015. The Pascal Visual Object Classes Challenge: A Retrospective. Int. J. Comput. Vis. 111, 98–136.

Forster, P.M., Maycock, A.C., McKenna, C.M., Smith, C.J., 2020. Latest climate models confirm need for urgent mitigation. Nat. Clim. Chang. 10, 7–10.

Franssen, H.H., Scherrer, S., 2008. Freezing of lakes on the Swiss Plateau in the period 1901-2006. Int. J. Climatol. 28, 421–433.

Gao, J., Yuan, Q., Li, J., Zhang, H., Su, X., 2020. Cloud Removal with Fusion of High Resolution Optical and SAR Images Using Generative Adversarial Networks. Remote Sens. 12, 191.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J., 2017. A Review on Deep Learning Techniques Applied to Semantic Segmentation. arXiv preprint, arXiv:1704.06857.

Geldsetzer, T., van der Sanden, J., Brisco, B., 2010. Monitoring lake ice during spring melt using RADARSAT-2 SAR. Can. J. Remote Sens. 36, S391–S400.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13 - 15 May.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep Sparse Rectifier Neural Networks, in: International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, USA, 11 - 13 April.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. `http://www.deeplearningbook.org`.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Gou, P., Ye, Q., Che, T., Feng, Q., Ding, B., Lin, C., Zong, J., 2017. Lake ice phenology of Nam Co, Central Tibetan Plateau, China, derived from multiple MODIS data products. J. Great Lakes Res. 43, 989–998.

Gou, P., Ye, Q., Wei, Q., 2015. Lake ice change at the Nam Co Lake on the Tibetan Plateau during 2000-2013 and influencing factors. Prog. Geogr. 34, 1241–1249.

Gunn, G., Duguay, C.R., Atwood, D.K., King, J., Toose, P., 2018. Observing scattering mechanisms of bubbled freshwater lake ice using polarimetric RADARSAT-2 (C-Band) and UW-Scat (X- and Ku-Bands). IEEE Trans. Geosci. Remote Sens. 56, 2887–2903.

Hall, D.K., Riggs, G.A., 2016. MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6, Boulder, Colorado, USA, NASA National Snow and Ice Data Center Distributed Active Archive Center. Available online: `https://doi.org/10.5067/MODIS/MOD10A1.006` (accessed 10 July 2021).

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: International Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 27 - 30 June.

Hendricks Franssen, H.J., Scherrer, S.C., 2008. Freezing of lakes on the Swiss plateau in the period 1901–2006. Int. J. Climatol. 28, 421–433.

Hoekstra, M., Jiang, M., Clausi, D.A., Duguay, C., 2020. Lake Ice-Water Classification of RADARSAT-2 Images by Integrating IRGS Segmentation with Pixel-Based Random Forest Labeling. Remote Sens. 12, 1425.

Hoffmann, S., Brust, C.A., Shadaydeh, M., Denzler, J., 2019. Registration of High Resolution SAR and Optical Satellite Imagery Using Fully Convolutional Networks, in: International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July - 2 August.

Howell, S., Brown, L., Kang, K., Duguay, C., 2009. Variability in ice phenology on Great Bear Lake and Great Slave Lake, Northwest Territories, Canada, from SeaWinds/QuikSCAT: 2000–2006. Remote Sens. Environ. 113, 816–834.

Huang, B., Song, H., 2012. Spatiotemporal Reflectance Fusion via Sparse Representation. IEEE Trans. Geosci. Remote Sens. 50, 3707–3716.

Huang, B., Song, H., Cui, H., Peng, J., Xu, Z., 2013a. Spatial and Spectral Image Fusion Using Sparse Matrix Factorization. IEEE Trans. Geosci. Remote Sens. 52, 1693–1704.

Huang, B., Zhang, H., Song, H., Wang, J., Song, C., 2013b. Unified fusion of remote-sensing imagery: Generating simultaneously high-resolution synthetic spatial–temporal–spectral earth observations. Remote Sens. Lett. 4, 561–569.

Hughes, L., Schmitt, M., Zhu, X., 2018. Mining Hard Negative Samples for SAR-Optical Image Matching Using Generative Adversarial Networks. Remote Sens. 10, 1552.

Hughes, L.H., Marcos, D., Lobry, S., Tuia, D., Schmitt, M., 2020. A deep learning framework for matching of SAR and optical imagery. ISPRS J. Photogramm. Remote Sens. 169, 166–179.

Jacobs, N., Burgin, W., Fridrich, N., Abrams, A., Miskell, K., Braswell, B.H., Richardson, A.D., Pless, R., 2009. The global network of outdoor webcams: Properties and applications, in: International Conference on Advances in Geographic Information Systems, Seattle, USA, 4 - 6 November.

Jarrett, K., Kavukcuoglu, K., Ranzato, M., , LeCun, Y., 2009. What is the best multistage architecture for object recognition?, in: International Conference on Computer Vision, Kyoto, Japan, 29 September - 2 October.

Jégou, S., Drozdzal, M., Vázquez, D., Romero, A., Bengio, Y., 2016. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, in: International Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, USA, 26 June – 1 July.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M.R., Kuemmerle, T., Meyfroidt, P., Mitchard, E.T., Reiche, J., Ryan, C.M., Waske, B., 2016. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. Remote Sens. 8, 70.

Kang, K., Duguay, C., Howell, S., 2012. Estimating ice phenology on large northern lakes from AMSR-E: algorithm development and application to Great Bear Lake and Great Slave Lake, Canada. Cryosphere 6, 235–254.

Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization, in: International Conference on Learning Representations, San Diego, USA, 7 - 9 May.

Kirillin, G., Leppäranta, M., Terzhevik, A., Granin, N., Bernhardt, J., Engelhardt, C., Efremova, T., Golosov, S., Palshin, N., Sherstyankin, P., Zdorovennova, G., Zdorovennov, R., 2012. Physics of seasonally ice-covered lakes: a review. Aquat. Sci 74, 659–682.

Knoll, L.B., Sharma, S., Denfeld, B.A., Flaim, G., Hori, Y., Magnuson, J.J., Straile, D., Weyhenmeyer, G.A., 2019. Consequences of lake and river ice loss on cultural ecosystem services. Limnol. Oceanogr. Lett. 4, 119–131.

Kondratyev, K., Filatov, N., 1999. in: Limnology and remote sensing. A contemporary approach. Springer-Praxis, Chichester.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems, Lake Tahoe, USA, 3 - 8 December.

Kropáček, J., Maussion, F., Chen, F., Hoerz, S., Hochschild, V., 2013. Analysis of ice phenology of lakes on the Tibetan Plateau from MODIS data. Cryosphere 7, 287–301.

Kulkarni, S.C., Rege, P.P., 2020. Pixel level fusion techniques for SAR and optical images: A review. Inf. Fusion 59, 13–29.

Lanaras, C., Baltsavias, E., Schindler, K., 2017a. Hyperspectral Super-Resolution with Spectral Unmixing Constraints. Remote Sens. 9, 1196.

Lanaras, C., Bioucas-Dias, J., Baltsavias, E., Schindler, K., 2017b. Super-Resolution of Multispectral Multiresolution Images From a Single Sensor, in: International Conference on Computer Vision and Pattern Recognition Workshops, Hawaii, USA, 21 - 26 July.

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. ISPRS J. Photogramm. Remote Sens. 146, 305–319.

Latifovic, R., Pouliot, D., 2007. Analysis of climate change impacts on lake ice phenology in Canada using the historical satellite data record. Remote Sens. Environ. 106, 492–507.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521, 436–444.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.

Leigh, S., Wang, Z., Clausi, D.A., 2014. Automated ice–water classification using dual polarization SAR satellite imagery. IEEE Trans. Geosci. Remote Sens. 52, 5529–5539.

Lemmetyinen, J., Derksen, C., Pulliainen, J., Strapp, W., Toose, P., Walker, A., Tauriainen, S., Pihlflyckt, J., Kärnä, J., Hallikainen, M., 2009. A comparison of airborne microwave brightness temperatures and snowpack properties across the boreal forests of Finland and Western Canada. IEEE Trans. Geosci. Remote Sens. 47, 965–978.

Lenormand, F., Duguay, C.R., Gauthier, R., 2002. Development of a historical ice database for the study of climate change in Canada. Hydrol. Process. 16, 3707–3722.

Leppäranta, M., Kosloff, P., 2000. The thickness and structure of Lake Pääjärvi ice. Geophysica 36, 233–248.

Liu, Y., Key, J., Mahoney, R., 2016. Sea and Freshwater Ice Concentration from VIIRS on Suomi NPP and the Future JPSS Satellites. Remote Sens. 8, 523.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. of Photogramm. Remote Sens. 152, 166–177.

Maas, A.L., Hannun, A.Y., Ng, A.Y., 2013. Rectifier Nonlinearities Improve Neural Network Acoustic Models, in: International Conference on Machine Learning Workshop on Deep Learning for Audio, Speech, and Language Processing, Atlanta, USA, 16 - 21 June.

Maaten, L.v.d., Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. IEEE Trans. Geosci. Remote Sens. 55, 645–657.

Magnuson, J.J., Robertson, D.M., Benson, B.J., Wynne, R.H., Livingstone, D.M., Arai, T., Assel, R.A., Barry, R.G., Card, V., Kuusisto, E., Granin, N.G., Prowse, T.D., Stewart, K.M., Vuglinski, V.S., 2000. Historical Trends in Lake and River Ice Cover in the Northern Hemisphere. Science 289, 1743–1746.

Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. IEEE Geosci. Remote Sens. Lett. 13, 105–109.

Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: Improving semantic image segmentation with boundary detection. ISPRS J. of Photogramm. Remote Sens. 135, 158–172.

Masson-Delmotte, V., Zhai, P., Pörtner, H.O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M., I., Lonnoy, E., Maycock, T., Tignor, M., Waterfield, T., 2018. IPCC: Summary for Policymakers, in: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and

efforts to eradicate poverty, World Meteorological Organization, Geneva, Switzerland, 32 pp, Available online: https://www.ipcc.ch/sr15/ (accessed 10 July 2021).

Melgani, F., Serpico, S.B., 2002. A statistical approach to the fusion of spectral and spatio-temporal contextual information for the classification of remote-sensing images. Pattern Recognit. Lett. 23, 1053–1061.

Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. ISPRS J. Photogramm. Remote Sens. 166, 333–346.

Merkle, N., Luo, W., Auer, S., Müller, R., Urtasun, R., 2017. Exploiting Deep Matching and SAR Data for the Geo-Localization Accuracy Improvement of Optical Satellite Images. Remote Sens. 9, 586.

Michel, B., Ramseier, R.O., 1971. Classification of river and lake ice. Can. Geotech. J. 8, 36–45.

Miles, K.E., Willis, I.C., Benedek, C.L., Williamson, A.G., Tedesco, M., 2017. Toward Monitoring Surface and Subsurface Lakes on the Greenland Ice Sheet Using Sentinel-1 SAR and Landsat-8 OLI Imagery. Front. Earth Sci. 5, 58.

Mishchuk, A., Mishkin, D., Radenovic, F., Matas, J., 2017. Working hard to know your neighbor's margins: local descriptor learning loss, in: Advances in Neural Information Processing Systems, Long Beach, USA, 4 - 9 December.

Mou, L., Schmitt, M., Wang, Y., Zhu, X.X., 2017. A CNN for the identification of corresponding patches in SAR and optical imagery of urban scenes, in: Joint Urban Remote Sensing Event, Dubai. United Arab Emirates, 6 - 8 March.

Murfitt, J., Brown, L., 2017. Lake ice and temperature trends for Ontario and Manitoba: 2001 to 2014. Hydrol. Process. 31, 3596–3609.

Murfitt, J., Brown, L., Howell, S., 2018. Evaluating RADARSAT-2 for the monitoring of lake ice phenology events in mid-latitudes. Remote Sens. 10, 1641.

Murfitt, J., Duguay, C.R., 2021. 50 years of lake ice research from active microwave remote sensing: Progress and prospects. Remote Sens. Environ 264, 112616.

Nair, V., Hinton, G., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines, in: International Conference on Machine Learning, Haifa, Israel, 21 - 24 June.

Ogutu, J.O., Piepho, H., Schulz-Streeck, T., 2011. A comparison of random forests, boosting and support vector machines for genomic selection. BMC Proc. 5, S11.

Oke, T., 1987. in: Boundary Layer Climates (2nd edition). Routledge, London, UK.

Owen, A.B., 2006. A robust hybrid of lasso and ridge regression, Technical Report, Stanford University. Available online: https://statweb.stanford.edu/~owen/reports/hhu.pdf (accessed 10 July 2021).

Palosuo, E., 1965. Frozen slush on lake ice. Geophysica 9, 36–45.

Pham, T., Yokoya, N., Xia, J., Ha, N., Le, N., Nguyen, T., Dao, T., Vu, T., Pham, T., Takeuchi, W., 2020. Comparison of Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam. Remote Sens. 12, 1334.

Pointner, G., Bartsch, A., Forbes, B.C., Kumpula, T., 2018. The Role of Lake Size and Local Phenomena for Monitoring Ground-Fast Lake Ice. Int. J. Remote Sens. 40, 832–858.

Polyak, B.T., 1964. Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. 4, 1–17.

Prabha, R., Tom, M., Rothermel, M., Baltsavias, E., Leal-Taixe, L., Schindler, K., 2020. Lake Ice Monitoring with Webcams and Crowd-Sourced Images. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. V-2-2020, 549–556.

Qi, M., Liu, S., Yao, X., Xie, F., Gao, Y., 2020. Monitoring the Ice Phenology of Qinghai Lake from 1980 to 2018 Using Multisource Remote Sensing Data and Google Earth Engine. Remote Sens. 12, 2217.

Qi, M., Yao, X., Li, X., Duan, H., Gao, Y., Liu, J., 2019. Spatiotemporal characteristics of Qinghai Lake ice phenology between 2000 and 2016. J. Geogr. Sci. 29, 115–130.

Qiu, Y., Xie, P., Leppäranta, M., Wang, X., Lemmetyinen, J., Lin, H., Shi, L., 2019. MODIS-based Daily Lake Ice Extent and Coverage dataset for Tibetan Plateau. Big Earth Data 3, 170–185.

Ranchin, T., Wald, L., 2000. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. Photogramm. Eng. Remote Sens. 66, 49–61.

Rebetez, M., 1996. Seasonal relationship between temperature, precipitation and snow cover in a mountainous region. Theor. Appl. Climatol. 54, 99–106.

Reed, B., Budde, M., Spencer, P., Miller, A.E., 2009. Integration of MODIS-derived metrics to assess interannual variability in snowpack, lake ice, and NDVI in southwest Alaska. Remote Sens. Environ. 113, 1443–1452.

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J., Bengio, Y., 2019. Tackling Climate Change with Machine Learning. arXiv preprint, arXiv:1906.05433v2.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 5 - 9 October.

Roy, D., Borak, J., Devadiga, S., Wolfe, R., Zheng, M., Descloitres, J., 2002. The MODIS Land Product Quality Assessment Approach. Remote Sens. Environ. 83, 62–76.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks, in: International Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 18 - 23 June.

Sanli, F.B., Abdikan, S., Esetlili, M.T., Ustuner, M., Sunar, F., 2013. Fusion of TerraSAR-X and rapideye data: a quality analysis. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL-7/W2, 27–30.

Scarpa, G., Gargiulo, M., Mazza, A., Gaetano, R., 2018. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. Remote Sens. 10, 236.

Schindler, D.W., Beaty, K.G., Fee1, E.J., Cruikshank, D.R., DeBruyn, E.R., Findlay, D.L., Linsey, G.A., Shearer, J.A., Stainton, M.P., Turner, M.A., 1990. Effects of Climatic Warming on Lakes of the Central Boreal Forest. Science 250, 967–970.

Schmitt, M., Hughes, L.H., Qiu, C., Zhu, X.X., 2019. SEN12MS – A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-2, 153–160.

Schmitt, M., Hughes, L.H., Zhu, X.X., 2018. The SEN1-2 Dataset for Deep Learning in SAR-Optical Data Fusion. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-1, 141–146.

Schmitt, M., Tupin, F., Zhu, X.X., 2017. Fusion of SAR and optical remote sensing data – Challenges and recent trends, in: International Geoscience and Remote Sensing Symposium, Fort Worth, USA, 23 - 28 July.

Schmitt, M., Zhu, X., 2016. Data Fusion and Remote Sensing: An ever-growing relationship. IEEE Geosci. Remote Sens. Mag 4, 6–23.

Sharma, S., Blagrave, K., Magnuson, J.J., O'Reilly, C.M., Oliver, S., Batt, R.D., Magee, M.R., Straile, D., Weyhenmeyer, G.A., Winslow, L., Woolway, R.L., 2019. Widespread loss of lake ice around the Northern Hemisphere in a warming world. Nat. Clim. Chang. 9, 227–231.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition, in: International Conference on Learning Representations, San Diego, USA, May 7 - 9.

Slater, T., Lawrence, I., R., Otosaka, I., N., Shepherd, A., Gourmelen, N., Jakob, L., Tepes, P., Gilbert, L., Nienow, P., 2021. Review article: Earth's ice imbalance. Cryosphere 15, 233–246.

Šmejkalová, T., Edwards, M., Dash, J., 2016. Arctic lakes show strong decadal trend in earlier spring ice-out. Sci. Rep. 6, 38449.

Song, H., Liu, Q., Wang, G., Hang, R., Huang, B., 2018. Spatiotemporal satellite image fusion using deep convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 821–829.

Spencer, P., Miller, A.E., Reed, B., Budde, M., 2008. Monitoring lake ice seasons in southwest Alaska with MODIS images, in: Pecora Conference, Denver, USA, 18 - 20 November.

Surdu, C.M., Duguay, C.R., Brown, L.C., Fernández Prieto, D., 2014. Response of ice cover on shallow lakes of the North Slope of Alaska to contemporary climate conditions (1950–2011): Radar remote-sensing and numerical modeling data analysis. Cryosphere 8, 167–180.

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 Sens. 7, 6133–6159.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning, in: International Conference for Machine Learning, Atlanta, USA, 16 - 21 June.

Sütterlin, M., Duguay-Tetzlaff, A., Wunderle, S., 2017. Toward a Lake Ice Phenology Derived from VIIRS Data, in: EGU General Assembly, Vienna, Austria, 23 - 28 April.

Svacina, N., Duguay, C., King, J., 2014. Modelled and satellite-derived surface albedo of lake ice - part II: evaluation of MODIS albedo products. Hydrol. Process. 28, 4562–4572.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: International Conference on Computer Vision and Pattern Recognition, Boston, USA, 7 - 12 June.

Thoma, M., 2016. A survey of semantic segmentation. arXiv preprint, arXiv:1602.06541.

Tom, M., Aguilar, R., Imhof, P., Leinss, S., Baltsavias, E., Schindler, K., 2020a. Lake Ice Detection from Sentinel-1 SAR with Deep Learning. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. V-3-2020, 409–416.

Tom, M., Baltsavias, E., Schindler, K., 2020b. Integrated Lake Ice Monitoring and Generation of Sustainable, Reliable, Long Time-Series. Available online: `https://ethz.ch/content/dam/ethz/special-interest/baug/igp/photogrammetry-remote-sensing-dam/documents/pdf/Misc/Lake_Ice_Project_Final_Report.pdf` (accessed 10 July 2021).

Tom, M., Jiang, Y., Baltsavias, E., Schindler, K., 2021a. Learning a Sensor-invariant Embedding of Satellite Data: A Case Study for Lake Ice Monitoring. arXiv preprint, arXiv:2107.09092.

Tom, M., Kälin, U., Sütterlin, M., Baltsavias, E., Schindler, K., 2018. Lake ice detection in low-resolution optical satellite images. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-2, 279–286.

Tom, M., Lanaras, C., Baltsavias, E., Schindler, K., 2017. Ice Detection in Swiss Lakes using MODIS Data, in: Asian Conference on Remote Sensing, New Delhi, India, 23 - 27 October.

Tom, M., Prabha, R., Wu, T., Baltsavias, E., Leal-Taixé, L., Schindler, K., 2020c. Ice Monitoring in Swiss Lakes from Optical Satellites and Webcams Using Machine Learning. Remote Sens. 12, 3555.

Tom, M., Suetterlin, M., Bouffard, D., Rothermel, M., Wunderle, S., Baltsavias, E., 2019. Integrated monitoring of ice in selected Swiss lakes, Final Project Report. Available online: `https://arxiv.org/abs/2008.00512` (accessed 10 July 2021).

Tom, M., Wu, T., Baltsavias, E., Schindler, K., 2021b. Recent Ice Trends in Swiss Mountain Lakes: 20-year Analysis of MODIS Imagery. arXiv preprint, arXiv:2103.12434.

Trishchenko, A.P., 2019. Clear-Sky Composites over Canada from Visible Infrared Imaging Radiometer Suite: Continuing MODIS Time Series into the Future. Can. J. Remote Sens. 45, 276–289.

Trishchenko, A.P., Ungureanu, C., 2017. Intercomparison of MODIS and VIIRS Results for Mapping Summer Minimum of Snow and Ice (MSI) Extent Over Canadian Landmass, in: EARSeL workshop on Land Ice and Snow, Bern, Switzerland, 7 - 9 February.

Trishchenko, A.P., Ungureanu, C., 2018. Warm Season Snow/Ice Probability Maps from MODIS and VIIRS Sensors over Canada, in: International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22 - 27 July.

Volpi, M., Tuia, D., 2017. Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks. IEEE Trans. Geosci. Remote Sens. 55, 881–893.

Wainer, J., 2016. Comparison of 14 different families of classification algorithms on 115 binary datasets. arXiv preprint, arXiv:1606.00930v1.

Wang, J., Duguay, C.R., Clausi, D.A., Pinard, V., Howell, S.E.L., 2018a. Semi-automated classification of lake ice cover using dual polarization RADARSAT-2 imagery. Remote Sens. 10, 1727.

Wang, K., Leppäranta, M., A, R., 2006. Modeling ice dynamics in lake peipsi. Verh. Int. Verein. Limnol. 29, 1443–1446.

Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., Jiao, L., 2018b. A deep learning framework for remote sensing image registration. ISPRS J. Photogramm. Remote Sens. 145, 148–164.

Wang, W., Lee, X., Xiao, W., Liu, S., Schultz, N., Wang, Y., Zhang, M., Zhao, L., 2018c. Global lake evaporation accelerated by changes in surface energy allocation in a warmer climate. Nature Geosci. 11, 410–414.

Williamson, A.G., Banwell, A.F., Willis, I.C., Arnold, N.S., 2018. Dual-satellite (Sentinel-2 and Landsat 8) remote sensing of supraglacial lakes in Greenland. Cryosphere 12, 3045–3065.

Wu, Y., Duguay, C.R., Xu, L., 2021a. Assessment of machine learning classifiers for global lake ice cover mapping from MODIS TOA reflectance data. Remote Sens. Environ. 253, 112206.

Wu, Y., Duguay, C.R., Xu, L., 2021b. Assessment of machine learning classifiers for global lake ice cover mapping from MODIS TOA reflectance data. Remote Sens. Environ. 253, 112206.

Wynne, R.H., Lillesand, T.M., 1993. Satellite observation of lake ice as a climate indicator: Initial results from statewide monitoring in Wisconsin. Photogramm. Eng. Remote Sens. 59, 1023–1031.

Xiao, M., Rothermel, M., Tom, M., Galliani, S., Baltsavias, E., Schindler, K., 2018. Lake ice monitoring with webcams. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. IV-2, 311–317.

Xu, B., Wang, N., Chen, T., Li, M., 2015. Empirical Evaluation of Rectified Activations in Convolutional Network. arXiv preprint, arXiv:1505.00853v2.

Yang, Q., Song, K., Wen, Z., Hao, X., Fang, C., 2019. Recent trends of ice phenology for eight large lakes using MODIS products in Northeast China. Int. J. Remote Sens. 40, 5388–5410.

Yao, X., Li, L., Zhao, J., Sun, M., Li, J., Gong, P., An, L., 2016. Spatial-temporal variations of lake ice phenology in the Hoh Xil region from 2000 to 2011. J. Geogr. Sci. 26, 70–82.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716.

Zeiler, M.D., 2012. ADADELTA: An Adaptive Learning Rate Method. arXiv preprint, arXiv:1212.5701.

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. Int. J. Image Data Fusion 1, 5–24.

Zhang, S., Pavelsky, T.M., 2019. Remote Sensing of Lake Ice Phenology across a Range of Lakes Sizes, ME, USA. Remote Sens. 11, 1718.

Zhang, S., Pavelsky, T.M., Arp, C.D., Yang, X., 2021. Remote sensing of lake ice phenology in Alaska. Environ. Res. Lett. 16, 064007.

Zhu, H., Meng, F., Cai, J., Lu, S., 2016. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. J. Vis. Commun. Image Represent. 34, 12–27.

Zhu, X., Cai, F., Tian, J., Williams, T.K.A., 2018. Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. Remote Sens. 10, 527.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.

B | 20-winter time series

B.1. Operational lake ice products

Details of operational lake ice products are shown in Table B.1.

Table B.1.: Comparison of the operational lake ice / snow products. Note that lake ice extent is still a pre-operational product. Res indicates resolution.

Product	**Availability**	**Spatial res**	**Temporal res**	**Input sensor(s)**
CCI Lake Ice Cover	from 2000	250m	daily	MODIS, VIIRS Sentinel-1,-3
MODIS Snow Product	from 2000	500m	daily	MODIS
VIIRS Snow Product	from 2012	375m	daily	VIIRS
Lake Ice Extent	from 2017	250m	daily	MODIS

B.2. Characteristics of the target lakes

Details of the target lakes are shown in Table B.2

Table B.2.: Details of the lakes. (primary source: Wikipedia). Maximum and average depths are shown in m. Last three rows display information about the nearest meteorological stations.

	Sihl	**Sils**	**Silvaplana**	**St. Moritz**
Lat (°N), Long (°E)	47.14, 8.78	46.42, 9.74	46.45. 9.79	46.49, 9.85
Altitude (m)	889	1797	1791	1768
Depth (Max, Avg)	23, 17	71, 35	77, 48	42, 26
Area (km^2)	11.3	4.1	2.7	0.78
Volume (Mm^3)	96	137	140	20
Meteo station	EIN	SIA	SIA	SAM
Lat (°N), Long (°E)	47.13, 8.75	46.43, 9.77	46.43, 9.77	46.53, 9.88
Altitude (m)	910	1804	1804	1708

B.3. MODIS and VIIRS bands

MODIS and VIIRS band spectrum are displayed in Fig. B.1

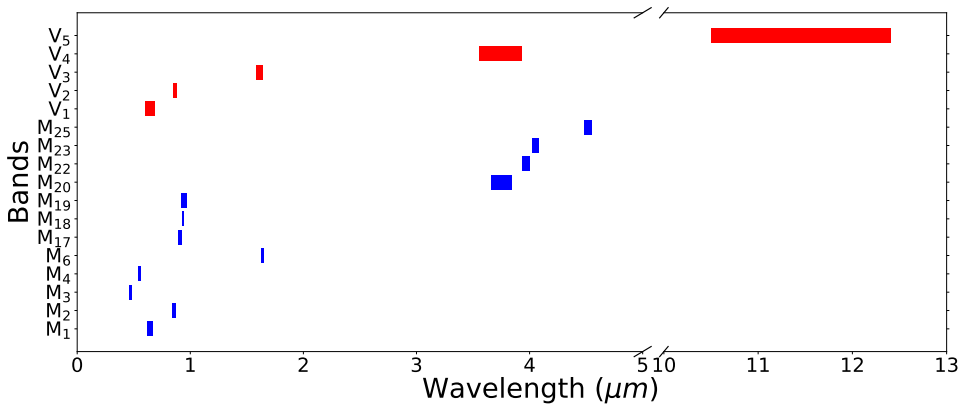


Figure B.1.: Spectrum of the MODIS (M, blue) and VIIRS (V, red) bands used in our analysis.

B.4. Class imbalance in our dataset

Details on class imbalance in our dataset are shown in Fig. B.2

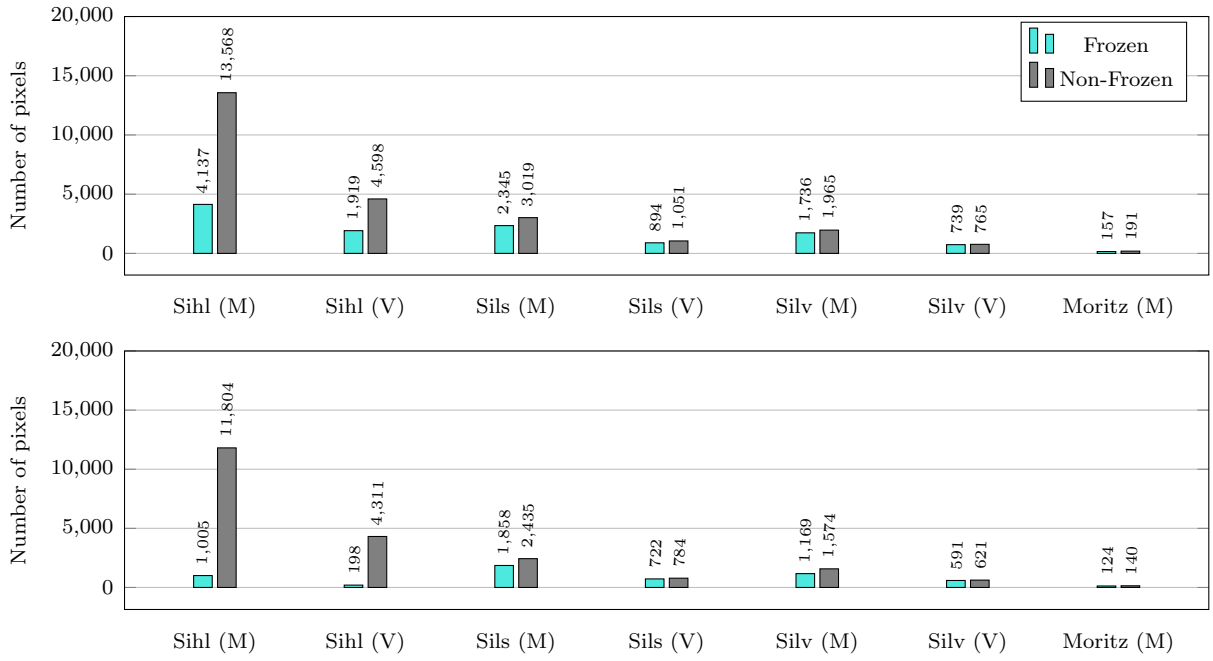


Figure B.2.: Bar graphs showing the class distribution in our dataset from the winters 2016–17 (top) and 2017–18 (bottom). The total number of clean, cloud-free pixels from the non-transition dates that are at least 30% cloud-free are shown. M and V denote MODIS and VIIRS respectively. Silv and Moritz represent lakes Silvaplana and St. Moritz respectively.

B.5. Correlation of weather data and LIP events

Correlation of LIP events with temperature variables are shown in. Fig. B.3. Similarly, correlations with sunshine and precipitation are displayed in Fig. B.4.

Figure B.3.: Correlation of MWT (°C) with ICD (days) and BUS are shown in rows 1 and 2 respectively, AFDD (°C) with ICD, CFD (days) and FUE are displayed in rows 3, 4 and 5 respectively. Results for lakes Sils and Silvaplana are displayed in left and right columns respectively.

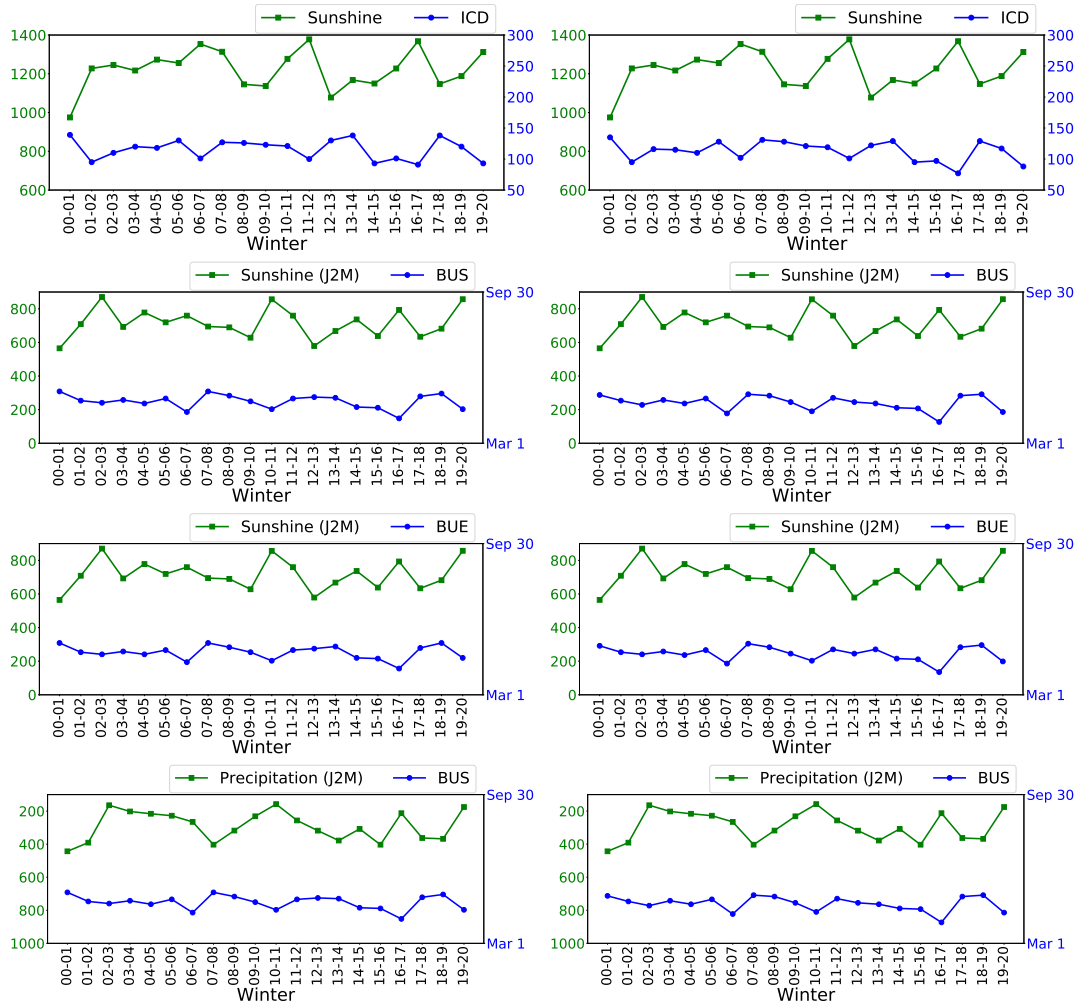


Figure B.4.: Correlation of total winter sunshine (hours) with ICD (days) is shown in row 1, total sunshine from January to May (J2M) with BUS and BUE are displayed in rows 2 and 3 respectively. Correlation of total precipitation (mm) from January to May (J2M) and BUS is shown in last row. Results for lakes Sils and Silvaplana are displayed in left and right columns respectively.

C | Embedding learning

C.1. Details of target lakes

Properties of the four target lakes are displayed in Table C.1.

Table C.1.: Physical properties of target lakes.

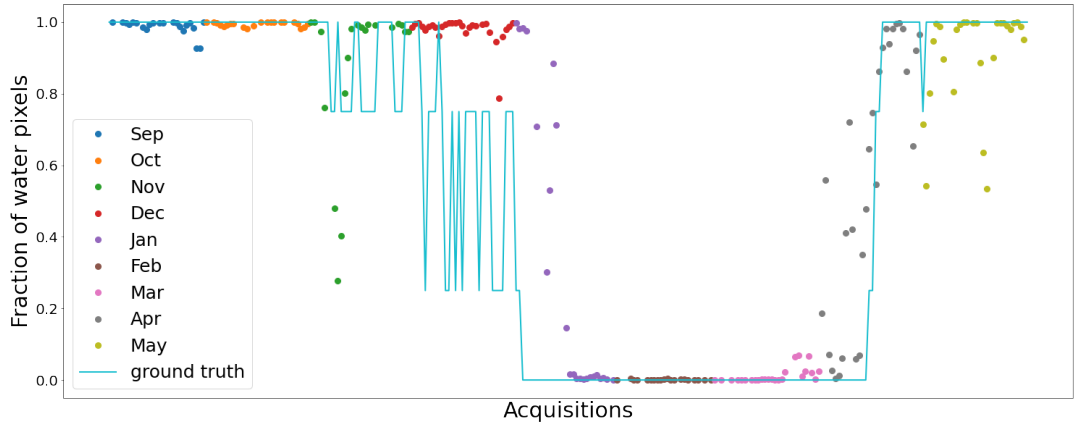
	Sihl	**Sils**	**Silvaplana**	**St. Moritz**
Area (km^2)	11.3	4.1	2.7	0.78
Altitude (m)	889	1797	1791	1768
Max. depth (m)	23	71	77	42
Avg. depth (m)	17	35	48	26
Volume (Mm^3)	96	137	140	20

C.2. Full winter time-series

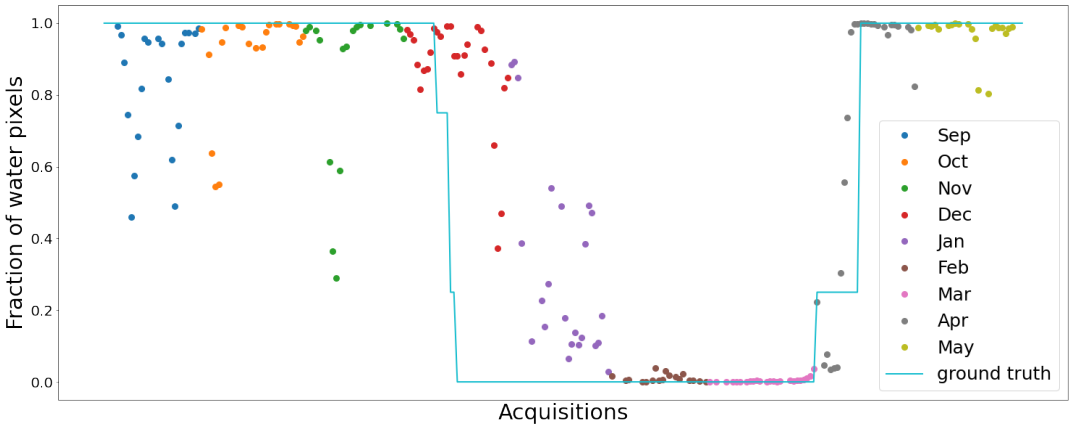
Full winter time-series results for lakes Silvaplana and St. Moritz are displayed in Fig. C.1.

C.3. Embedding visualisation

t-SNE representation for the learnt embedding for lakes Silvaplana and St. Moritz are displayed in Fig C.2.

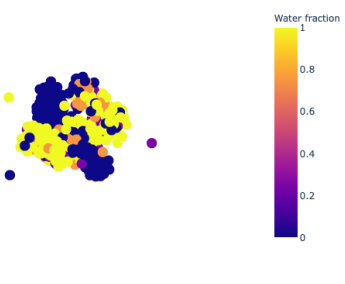


Lake Silvaplana

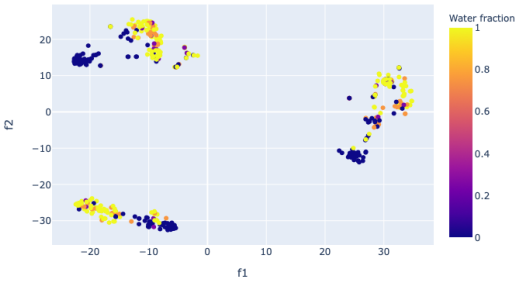


Lake St. Moritz

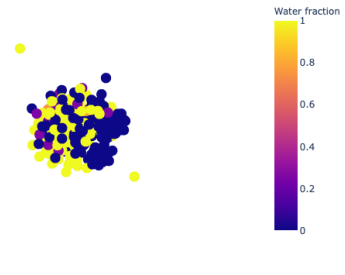
Figure C.1.: Time-series plots for lakes Silvaplana and St. Moritz from winter 2016–17 using a model trained on all the data (all four lakes) from winter 2017–18. Best if viewed on screen.



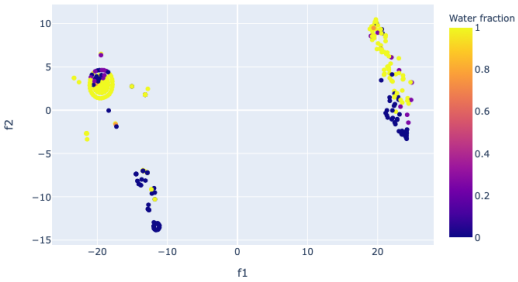
Lake Silvaplana, dimension=3



Lake Silvaplana, dimension=2



Lake St. Moritz, dimension=3



Lake St. Moritz, dimension=2

Figure C.2.: t-SNE representation of the embedding learnt (lakes Silvaplana and St. Moritz, winter 2016–17) using the proposed *2-step* approach with a model trained on all the data from winter 2017–18. Water fraction refers to the predicted fraction of non-frozen pixels. Best if viewed on screen.

D List of publications

[1] Prabha, R., Tom, M., Rothermel, M., Baltsavias, E., Leal-Taixe, L., Schindler, K., 2020. Lake Ice Monitoring with Webcams and Crowd-Sourced Images, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-2-2020, 549–556.

[2] Rothermel, M., Xiao, M., Tom, M., Baltsavias, E., Schindler, K., 2018. Monitoring der Vereisung von Schweizer Seen mit Webcams, Geomatik Schweiz, 9, 268-271.

[3] Tom, M., Lanaras, C., Baltsavias, E., Schindler, K., 2017. Ice Detection in Swiss Lakes using MODIS Data, Asian Conference on Remote Sensing, New Delhi, India, 23-27 October.

[4] Tom, M., Kälin, U., Sütterlin, M., Baltsavias, E., Schindler, K., 2018. Lake Ice Detection in Low-Resolution Optical Satellite Images, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2, 279-286.

[5] Tom, M., Sütterlin, M., Bouffard, D., Rothermel, M., Hamann, U., Duguay-Tetzlaff, A., Wunderle, S., Baltsavias, E., 2018. Integrated Lake Ice Monitoring in Swiss Lakes, EUMETSAT Meteorological Satellite Conference, Tallinn, Estonia, 17-21 September.

[6] Tom, M., Rothermel, M., Baltsavias, E., Schindler, K., 2019. Semantic Segmentation of Ice in selected Swiss Lakes, Swiss Workshop on Machine Learning for Environmental and Geosciences, Dübendorf, Switzerland, 16-17 January.

[7] Tom, M., Aguilar, R., Imhof, P., Leinss, S., Baltsavias, E., Schindler, K., 2020. Lake Ice Detection from Sentinel-1 SAR with Deep Learning, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2020, 409–416.

[8] Tom, M., Prabha, R., Wu, T., Baltsavias, E., Leal-Taixe, L., Schindler, K., 2020. Ice Monitoring in Swiss Lakes from Optical Satellites and Webcams using Machine Learning, Remote Sens., 12, 3555.

[9] Tom, M., Jiang, Y., Baltsavias, E., and Schindler, K., 2021. Learning a Sensor-invariant Embedding of Satellite Data: A Case Study for Lake Ice Monitoring, arXiv preprint, arXiv:2107.09092

[10] Tom, M., Wu, T., Baltsavias, E., and Schindler, K., 2021. Recent Ice Trends in Swiss Mountain Lakes: 20-year Analysis of MODIS Imagery, arXiv preprint, arXiv:2103.12434.

[11] Xiao, M., Rothermel, M., Tom, M., Galliani, S., Baltsavias, E., Schindler, K., 2018. Lake Ice Monitoring with Webcams, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2, 311-317.