

Combining “Deep Learning” and Physically Constrained Neural Networks to Derive Complex Glaciological Change Processes from Modern High-Resolution Satellite Imagery: Application of the GEOCLASS-image System to Create VarioCNN for Glacier Surges

Ute Christina Herzfeld¹, Lawrence John Hessburg¹, Thomas Trantow¹, and Adam N Hayes¹

¹University of Colorado Boulder

April 16, 2024

Abstract

The objectives of this paper are to investigate the tradeoffs between a physically constrained neural network and a deep, convolutional neural network and to design a combined ML approach (“VarioCNN”). Our solution is provided in the framework of a cyberinfrastructure that includes a newly designed ML software, GEOCLASS-image, modern high-resolution satellite image datasets (Maxar WorldView data) and instructions/descriptions that may facilitate solving similar spatial classification problems. Combining the advantages of the physically-driven connectionist-geostatistical classification method with those of an efficient CNN, VarioCNN, provides a means for rapid and efficient extraction of complex geophysical information from submeter resolution satellite imagery. A retraining loop overcomes the difficulties of creating a labeled training data set.

Computational analyses and developments are centered on a specific, but generalizable, geophysical problem: The classification of crevasse types that form during the surge of a glacier system. A surge is a glacial catastrophe, an acceleration of a glacier to typically 100-200 times its normal velocity, which for a marine-terminating glacier leads to sudden and substantial mass transfer from the cryosphere to the oceans, contributing significantly to sea-level-rise. The sudden and rapid acceleration characteristic of a surge results in formation of crevasses, whose spatial characteristics provide informants on the ice-dynamic processes that occur during the surge. GEOCLASS-image is applied to study the current (2016-2024) surge in the Negribreen Glacier System, Svalbard. The geophysical result is a description of the structural evolution and expansion of the surge, based on crevasse types that capture ice deformation in 6 simplified classes.

Combining “Deep Learning” and Physically Constrained Neural Networks to Derive Complex Glaciological Change Processes from Modern High-Resolution Satellite Imagery: Application of the GEOCLASS-image System to Create VarioCNN for Glacier Surges

Ute C. Herzfeld^{1,2,3}, Lawrence Hessburg^{1,2}, Thomas Trantow¹ and Adam Hayes^{1,2}

¹ Department of Electrical, Computer and Energy Engineering, Geomathematics, Remote Sensing, and Cryospheric Sciences Laboratory, University of Colorado Boulder, Boulder, CO, USA,

² Department of Computer Science, University of Colorado Boulder, Boulder, CO, USA,

³ Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO, USA

ABSTRACT. The objectives of this paper are to investigate the tradeoffs between a physically constrained neural network and a deep, convolutional neural network and to design a combined ML approach (“VarioCNN”). Our solution is provided in the framework of a cyberinfrastructure that includes a newly designed ML software, GEOCLASS-image, modern high-resolution satellite image data sets (Maxar WorldView data) and instructions/descriptions that may facilitate solving similar spatial classification problems. Combining the advantages of the physically-driven connectionist-geostatistical classification method with those of an efficient CNN, VarioCNN provides a means for rapid and efficient extraction of complex geophysical information from submeter resolution satellite imagery. A retraining loop overcomes the difficulties of creating a labeled training data set.

Computational analyses and developments are centered on a specific, but generalizable, geophysical problem: The classification of crevasse types that form during the surge of a glacier system. A surge is a glacial catastrophe, an acceleration of a glacier to typically 100-200 times its normal velocity, which for a marine-terminating glacier leads to sudden and substantial mass transfer from the cryosphere to the oceans, contributing significantly to sea-level-rise. The sudden and rapid acceleration characteristic of a surge results in formation of crevasses, whose spatial characteristics provide informants on the ice-dynamic processes that occur during the surge. GEOCLASS-image is applied to study the current (2016-2024) surge in the Negribreen Glacier System, Svalbard. The geophysical result is a description of the structural evolution and expansion of the surge, based on crevasse types that capture ice deformation in 6 simplified classes.

1 Introduction

The objective of this paper is to contribute to three challenges in different disciplines: (1) Earth observation and data analysis, (2) climatic and cryospheric change, and (3) machine learning (ML).

Challenge 1. Harnessing the data revolution in Earth observation from space. Observations of our rapidly changing Earth are largely carried out from space, and the collection of such Earth

observation data from satellites has rapidly advanced with increasingly large and detailed data sets becoming available for scientific investigations (Wagner, 2015). The data revolution has led to both new opportunities and challenges for science, as extraction of information on complex geophysical processes from large and high-resolution data sets is becoming increasingly difficult (a problem that has been summarized as “Harnessing the data revolution” by the U.S National Science Foundation (U.S. National Science Foundation, 2023)). In turn, this phenomenon has created a cyberinfrastructure problem in terms of a disconnect between the revolutionary increase in satellite image data on the one hand and the development of numerical Earth system models on the other hand, which are employed to aid in assessment of global climatic changes and their manifestations in warming and sea-level rise (SLR) (Masson-Delmotte et al., 2021; Greve, 1995; Greve and Blatter, 2009; Larour et al., 2012; Lipscomb et al., 2019; Sellevold et al., 2019; Hanna et al., 2020; Payne et al., 2021; Siahyaan et al., 2022; Gettelman et al., 2022). A bottleneck is created - and growing with the data revolution - as this new wealth of information revealed by the new satellites makes it hard to incorporate observations into physical-process models as the improved spatio-temporal scale introduced by the data sheds light onto subprocesses not easily incorporated into models.

In this paper, we will introduce an approach that integrates machine learning and physical knowledge into a physically-driven neural network, whose application will facilitate derivation of physical process understanding from high-resolution satellite data. Results include parameterized information in the form of thematic maps (time series of segmented satellite imagery) that can inform modeling as well as lend themselves to direct geophysical interpretation and discovery.

Challenge 2. Glacial acceleration and Sea-Level-Rise assessment. We address a climatic and cryospheric change problem, the phenomenon of glacial acceleration, that has been identified as one of two main sources of uncertainty SLR assessment, as identified by the Intergovernmental Panel on Climate Change (IPCC) in their 2013 Assessment Report 5 [the other source is atmospheric] (Stocker et al., 2013). The most recent IPCC AR 6, published in 2021, does not present a solution but rather elevates the urgency of understanding glacial acceleration by declaring it a “deep uncertainty” in SLR assessment (Masson-Delmotte et al., 2021). The different types of accelerating glaciers include surge-type glaciers, tidewater glaciers, fjord glaciers (isbræ) and ice streams (Clarke, 1987; Truffer and Echelmeyer, 2003; Mayer and Herzfeld, 2000; Jiskoot, 2011; Herzfeld et al., 2013; Straneo and Cenedese, 2015; Trantow and Herzfeld, 2018; Murray et al., 2003a; Robel et al., 2018; Hill et al., 2018; Nuth et al., 2019; Benn et al., 2019; Zheng et al., 2019; Alley et al., 2021; Frank et al., 2022; Grinsted et al., 2022; Ehrenfeucht et al., 2023). Acceleration frequency may be intrinsic to the glacier type, quasi-periodic, or single-time. Initialization of an acceleration may be due to internal dynamics of the glacier or externally forced, for instance, induced by warming ocean water at the front of the glacier or controlled by a combination of several factors (Straneo et al., 2012; Rignot et al., 2012; Herzfeld et al., 2014; Herzfeld and Wallin, 2014). Spatial acceleration may be due to subglacial (bed) topography (Herzfeld et al., 2014) or caused by a dynamic event. All types of acceleration typically lead to the formation of crevasse fields. Surging is the type of acceleration that has seen the least amount of research, and complexity of ice flow during surging defies many classic data analysis methods thus rendering most cyberinfrastructures incapable of modeling this geophysical process.

In this paper, we focus on an exemplary analysis of glacial acceleration during the surge of an Arctic glacier system, the Negribreen Glacier System (NGS), through classification of crevasse patterns as indicators of the drastic and rapid dynamic changes that occur during a surge. The surge has led to mass transfer from the glacier to the ocean on the order of 0.5-1% of global annual SLR in just a few months during the height of the surge (Herzfeld et al., 2021; Trantow and Herzfeld, 2023). The fact that a surge causes sudden mass transfer events from the cryosphere to the ocean leads to a catastrophic type of uncertainty in SLR estimation (the term “catastrophic” defined as continuous changes leading to sudden effects). If we are to reconcile SLR assessment, we need to understand surge processes.

Challenge 3. Integration of physically-constrained classification and modern “Deep Learning” approaches in satellite image classification. The surge is captured in time series of high-resolution satellite image data, which motivates a ML-based classification. While deep convolutional neural network (CNN) architectures have been considered to provide state of the art performance on standard image classification benchmarks such as the ImageNet dataset (Kampffmeyer et al., 2016; Rawat and Wang, 2017; Deng et al., 2009; Krizhevsky et al., 2012; He et al., 2016a), two problems exist: First, deeper networks only lead to increased performance up to a point, after which increased network depth results in increasingly worse performance due to the vanishing gradient problem (Hochreiter, 1998). Second, and more challenging for applications in the cryospheric sciences, is the fact that no published labeled training data sets exist for tasks of classification of ice-surface features, such as crevasses (see, Meyer and Pebesma (2021)). The role of crevasse types in identification of deformation types, which are directly related to glacial acceleration, will be described in section 3. A main task is thus the creation of such labeled data sets, required for training of a Neural Net (NN). For CNNs, the problem is exacerbated by the fact that very large numbers of training data (on the order of 100,000s) are needed.

We have previously developed a physically constrained ML approach, the connectionist-geostatistical classification method (Herzfeld and Zahner, 2001; Herzfeld, 2008; Herzfeld et al., 2013, 2023b). The connectionist-geostatistical method uses a two-tiered approach, in which the first step is a physically informed spatial statistical analysis, carried out in a discrete mathematics framework. The output of the geostatistical step provides the input for the NN, activating the neurons of the input layer. In order to carry out an actual classification, a connectionist approach is selected, which can utilize a multi-layer perceptron with backpropagation of errors (MLP-BP), or simply, a MLP. The MLP has proven to provide a robust and functional architecture for this type of classification and provided an efficient solution already 20 years ago (Herzfeld and Zahner, 2001). To train the connectionist-geostatistical classification, a small data set suffices, of a size that can reasonably be derived by an expert (Herzfeld and Zahner, 2001), on the order of several 100 labeled video-scenes or small subimages of a satellite image. However, advances in Earth observation, increasing data resolution and data set size, as well as advances in computer hardware and processing speed warrant investigation of modern “Deep Learning” architectures to facilitate fast and efficient processing.

The salient difference in the effectiveness of the two approaches lies less in the NN architecture (MLP versus CNN) than in the fact that the connectionist-geostatistical classification is a physically

informed approach (where the physical knowledge informs our approach to geostatistics), whereas in the case of the CNN the network’s much larger amount of degrees of freedom is what the determination of classes relies on. CNNs can be trained supervised or unsupervised (Camps-Valls et al., 2021b).

In this paper, we will investigate the trade-offs of a physically constrained NN and a CNN and introduce a first approach to leverage the advantages of both ML methods in an integrated image classification system. We propose a solution to natural science problems that takes an approach of combining and integrating physically constrained neural networks and modern ML methods. To this end, we will demonstrate that a physically constrained NN can be utilized to aid in creating a labeled training data set of sufficient size to train a CNN. We emphasize that physical knowledge needs to be leveraged in designing a ML approach that can be expected to provide solutions for the physical sciences and advance knowledge there.

2 Background on Neural Networks, Especially in the Geosciences and in (Satellite) Image Classification

In this section, we give a brief summary of the state of the art of ML in the geosciences, as well as ML applied to satellite image classification or analysis. Most works fall into one of several categories, addressed in the following subsections.

2.1 General References: Classic Papers, Review Papers and Books

While neural networks have seen a sudden rise in public attention in the most recent years, first research dates back to neural psychology at the end of the 19th century (Bain, 1873; James, 1890). Rosenblatt (Rosenblatt, 1958) introduced the perceptron, and the first deep learning perceptron came out in 1967 (Ivakhnenko and Lapa, 1967). The use of neural networks stalled in the early 1970s, mostly due to limitations computation (Minsky and Papert, 1969). Foundational research on CNNs and thus on Deep Learning dates back to the 1960s (Ivakhnenko and Lapa, 1967; Fukushima, 1980; LeCun et al., 1998; Krizhevsky et al., 2012). Important concepts that mark steps of development of NNs include backpropagation of errors and connectionism. Backpropagation of errors is an application of Leibniz’ chain rule (from 1673) to networks of differentiable nodes that has become a standard in optimizing MLPs (Werbos, 1990). Designing the connectionist-geostatistical classification approach, we applied an MLP with backpropagation of errors in the 1990s, using the Stuttgarter Neural Network software (Zell et al., 1994; Herzfeld and Zahner, 2001). The term connectionism refers to an approach to the study of human mental processes and cognition that utilizes mathematical models known as connectionist networks or artificial neural networks (Rumelhart et al., 1986; Sun, 2014).

A standard reference for Deep Learning is the book by Goodfellow and others (Goodfellow et al., 2016). A good general reference related to several topics of this paper is a book titled “Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences” (Camps-Valls et al., 2021b).

In their review of remote sensing image classification methods, Song et al. (2019) focus on applications of CNNs for extraction of semantic features in image data. Rawat and Wang (Rawat and Wang, 2017) present a review of deep convolutional networks for image classification, and Garcia-Garcia and others write a survey of deep learning techniques for image and video semantic segmentation (Garcia-Garcia et al., 2020). A review of ML methods for classification of remotely sensed imagery and applications to sea-ice classification is given in Herzfeld et al. (2016), and a review of hyperspectral image (HSI) classification using CNNs is presented by Li et al. (2019).

2.2 Classic Applications of NNs in the Geosciences

Prediction and assessment of sea-ice conditions in the Arctic, based on satellite remote sensing, has been an important tool for ocean navigation. Synthetic Aperture Radar (SAR) data lend themselves well for sea-ice observations, because the radar signal penetrates cloud layers and fog, which are frequent in Arctic atmospheric conditions and obscure optical satellite image data. A short review of ML methods for classification of remotely sensed imagery and applications to sea-ice classification is given in Herzfeld et al. (2016). Research on sea-ice classifications based on SAR data goes back to the 1990s (Kwok et al., 1992; Collins, 1992; Steffen and Heinrichs, 1994). These early methods typically allow distinguishing a small number of sea-ice types, such as 3 or 4. Most methods use multivariate statistics at pixel values in different channels, for example Ochilov and Clausi (2012); Wang et al. (2016); Dabboor and Geldsetzer (2014). The approach of Ochilov and Clausi (2012) is innovative in that it bases image segmentation on gradients in the original multivariate statistical parameters, using an edge-detection method. Early applications of NNs include Karvonen (2006); Zakhvatkina et al. (2013). Karvonen’s work (Karvonen, 2006) was a milestone in state-of-the-art statistical techniques in sea-ice classification with applications to the seasonal ice cover in the Baltic Sea, noting that understanding physical processes is an open problem. Zakhvatkina et al. (2013) introduce an interesting concept that combines a number of statistical parameters and a NN. Recent publications which utilize sea-ice classification include Roesel et al. (2012); Shen et al. (2017); Buckley et al. (2020, 2023).

Neural networks that address pattern recognition problems such as self-organized maps (Kohonen, 2012), a form of unsupervised classification, or “Learning vector quantification”, a supervised NN approach (Kohonen, 1990), achieved some popularity, but were found to be outperformed by MLPs with back propagation of errors (MLP-BP) for image analysis of repeated structural patterns (Herzfeld and Zahner, 2001). An overview of pattern recognition using NNs is given in Looney (1997).

2.3 Spectral Versus Spatial Classification

Most image classification methods are based on spectral or multi-spectral classification, i.e. they utilize the fact that an image is composed of several spectral bands (Song et al., 2019). The connectionist-geostatistical classification method that will be utilized in this paper is a form of spatial classification, which is based on the fact that repeating spatial structures of crevasse fields lead to characteristic types of vario functions (Herzfeld and Zahner, 2001; Herzfeld, 2008). In Herzfeld et al. (2016), we compare statistical and geostatistical classification methods to explore the potential of combined methods for sea-ice classification.

Vario functions are a formulation of the variogram in discrete mathematics (Herzfeld, 2002). Variograms are employed in satellite image characterization by Garrigues et al. (2007). Garrigues et al. (2007) explore first and second-order modeled histograms and variograms to characterize land-

scape spatial structures from remote-sensing imagery (SPOT-HRV NDVI data) and conclude that the method has potential to distinguish effects of anthropogenic landscape-forming processes from those of environmental and ecological processes, however they note that the method can be improved. In contrast, the connectionist-geostatistical method in the form used in this paper employs experimental vario functions directly to initialize the input neurons in a NN.

Most applications of variograms in satellite image analysis are estimations (kriging) or spatial or temporal analyses, rather than classifications, of satellite data, for example [Qing et al. \(2003\)](#); [Schwartz and Pinel-Puysegur \(2018\)](#), specifically, of Synthetic Aperture Radar data (SAR data). The differences between geostatistical estimation/ interpolation and extrapolation, characterization and classification are explained in [Herzfeld \(2008\)](#).

2.4 Computer Science Developments of ML Methods for Image Processing and Classification. CNNs.

Recent advances in NN research, especially for applications to image analysis/processing/classification, have been led by computer scientists. In the last approximately 10-15 years, Deep Learning methods have dominated. Within the field of Deep Learning approaches, CNNs are preeminent ([Camps-Valls et al., 2021b](#)). Deep Learning summarizes ML approaches that involve Neural Nets with large numbers of internal layers (for example, ResNet-1001 has 1001 layers ([He et al., 2016c, 2021](#))). Overviews of these methods are given in ([Camps-Valls et al., 2021b](#); [Song et al., 2019](#)). In contrast, the MLP used in the original (2001) connectionist-geostatistical classification has three layers: an input layer, an internal layer and output layer ([Herzfeld and Zahner, 2001](#)).

Types of CNNs that have been widely used include: (described largely following [Song et al. \(2019\)](#) with some updated references) (1) AlexNet, a CNN with five convolutional layers and two fully connected layers, first evaluated for ImageNet ([Deng et al., 2009](#); [Krizhevsky et al., 2012](#)), a prototype test data set. AlexNet won the so-called ImageNet challenge in 2012 ([He et al., 2016a](#)). (2) Network-in-Network (NiN) ([Lin et al., 2013](#)), where a MLP is added to each convolutional layer, replacing a simple linear convolutional layer, and an averaging method, called global average pooling, is applied to counteract overfitting, (3) VGG-Net ([Simonyan and Zisserman, 2014](#)), a 19-layer network with small (3x3) convolutional kernels, (4) GoogLeNet ([Tai et al., 2015](#)), a 22-layer network, (5) ResNet ([He et al., 2016a,c, 2021](#)), a family of so-called residual networks with depths of up to 1001 layers, (6) DenseNet ([Huang et al., 2017](#)), a NN type that uses cross-layer connections to improve network structure, (7) MS-CapsNet ([Xiang et al., 2018](#)), a multi-scale capsule network, ML methods, from multi-spectral statistical methods to CNNs, can be trained supervised or unsupervised ([Camps-Valls et al., 2021b](#)).

The work in this paper uses a form of ResNet, because ResNets have been found to excel at image classification problems. Hence ResNet principles and architectures will be described in more detail in section (8). Applications of CNNs in image classification are numerous (see, for example, [Krizhevsky et al. \(2012\)](#); [Kampffmeyer et al. \(2016\)](#); [Rawat and Wang \(2017\)](#)).

2.5 Identified Needs for Advancing Remote-Sensing-Data Classification Using ML Methods, in General and in the Geosciences

In this paper, we will address some of the shortcomings or challenges associated with applications of CNNs in image classifications, identified in [Song et al. \(2019\)](#): Lack of sufficient training data (see also [Virts et al., 2020](#); [Liu et al., 2020b](#); [Meyer and Pebesma, 2021](#)), need for remote-sensing-specific CNN architectures, time-efficiency of training CNNs for image classification, and a need for high-level CNN-based applications in remote-sensing image classification. The first three challenges concern technical aspects of NN developments, and our work will address all three. Most interesting to us is the observation (made by [Song et al. \(2019\)](#)) that most current remote-sensing image ML applications resemble those in computer vision, whereas identification of semantically complex information is largely missing in state-of-the-art research. This resonates with the authors’ observations that many modern CNNs are constructed for the same type of simple applications that were tackled with image processing methods several decades ago. For example, the hyper-deep ResNet-1001 ([He et al., 2016c, 2021](#)) is derived for multiframe video satellite image super-resolution processing, but then applied to a problem of differencing aircraft-presence/aircraft-absence already analyzed decades earlier. Another application to moving object detection is described in [Ou et al. \(2019\)](#). Note that the ResNets use very small convolutional kernels, which is a match to the fact that many image denoising or sharpening techniques of the 1900s used 3x3 or 5x5 or 7x7 kernels ([Herzfeld and Zahner, 2001](#)). It appears that the modern ML methods often perform similar applications as old methods, only faster, at higher resolution, or for more modern observations, e.g. satellite videography. In our paper, we aim to create an approach that allows to understand a certain, complex geophysical (cryospheric) phenomenon.

In part, the lack of actual conceptual advances or physical process understanding in the Earth Science from ML applications to image classification is tied to the fact that ML research is based on a relatively small number of labeled training data sets (an example is ImageNet ([Krizhevsky et al., 2012](#))).

Physically-driven NNs fall in the category that is termed “high-level (C)NN-based applications” (by [Song et al. \(2019\)](#)) or classification of geophysically complex information, such as crevasse classification for the surge problem in this paper. Identification and classification of complex information in imagery requires large sub-images, or large moving windows (not the same) ([Herzfeld and Zahner, 2001](#)), and last not least the creation of labeled training data for cryospheric applications. Along similar lines, [Reichstein et al. \(2019\)](#) in their review of Earth science applications highlight NN structures that include modules of data analysis from other than ML fields (see, subsections (2.5) and (2.6)), however, there are only a small number of such approaches listed - and none in cryospheric sciences. Our work falls in this category.

2.6 Recent Applications of NNs in Geosciences

Dominant application fields include land-cover/land-use (urban areas, farmlands, roads, water bodies), biogeosciences, and military applications, there especially change detection of airplanes present/absent at terminals (see, for example, [Aguilar et al. \(2013\)](#); [Waser et al. \(2014\)](#); [Kampffmeyer et al. \(2016\)](#); [Song et al. \(2019\)](#); [Sen et al. \(2020\)](#); [Virts et al. \(2020\)](#); [Vigneshwaran and Vasantha Kumar \(2021\)](#)). Neural Nets and other ML methods are increasingly finding applications in the geosciences. Reviews are found in [\(Luan and Tian, 2022\)](#); [\(Herzfeld et al., 2016\)](#).

Examples of papers where ML structures are applied include the following: Neural networks have been utilized in studies of vegetation canopy height, using ICESat-2 and Landsat data [\(Malambo and Popescu, 2023\)](#); [\(Narine et al., 2019\)](#). In a case study of a forest in Texas, [\(Narine et al., 2019\)](#) investigate the potential of using a Deep Neural Net (DNN) or a Random Forest (RF) model for above-ground biomass assessment based on ICESat-2 and Landsat data, finding similar performance values for the DNN and the RF. [\(Palm et al., 2021\)](#) explore applications of NNs for analysis of atmospheric data from ICESat-2, treated like image data. Common to these studies is that they are case studies, which investigate the applicability of previously published types of NNs to satellite data analysis. Other applications include forest canopy height determination from ICESat-2 and Landsat data [\(Li et al., 2020\)](#), disaster detection and monitoring (flood detection) using Random Forests [\(Gaffinet et al., 2023\)](#), and geological image classification using CNNs [\(de Lima et al., 2019\)](#).

In summary, recent applications of ML in the geosciences fall into two categories, (1) Computer scientists taking summative approaches to geoscience data classification (different formulation) and (2) geoscientists exploring applications of existing, previously published ML approaches to image analysis. Notable exceptions include feature augmented neural nets for satellite image classification (an approach that augments data sets with handcrafted feature data sets, see, for example, [\(Liu et al., 2020a\)](#)) and a new strand of methods that aim to integrate ML and physics (see, next section).

2.7 Approaches Aimed at Integrating Physical Sciences and ML

Most relevant for the work in this paper is a class of approaches that are aimed at integrating physical sciences and ML, by either using physical knowledge in ML or by using ML to improve physical models.

Exemplary approaches that include physics in ML have been termed “physics-aware ML” [\(Camp-Valls et al., 2021a\)](#), based on the concept that the elementary laws of physics ought to be respected by ML approaches in the geosciences. Under this label, challenges, more so than solutions, in the interplay of physics and ML have been identified that may help advance Earth system knowledge (encoding differential equations from data, constraining data-driven models with physics-priors and dependence constraints, improving parameterizations, emulating physical models, and blending

data-driven and process-based models). [Ge et al. \(2022\)](#) propose an approach termed “Geoscience-aware deep learning” (GADL), which will include geoscience features into Deep Learning models. This is similar to the concept of including handcrafted features in CNN-based satellite image classification suggested by [Liu et al. \(2020a\)](#). Other authors recognize the need for collaborative efforts in the field of geoscience and ML (e.g. [Karpatne et al. 2018](#)).

Physically-guided neural networks (PGNNs) leverage scientific knowledge, physical models and observational data in a neural network in order to make better predictions ([Daw et al. 2022](#)). The idea of physical consistency is used as a learning objective to allow generalization of the learned network. PGNN’s have been used to model complex physical systems that either lack required data constraints or incur large computational costs, such as those found in fluid dynamics problems [Muralidhar et al. \(2020\)](#) or power flow analysis [Hu et al. \(2020\)](#). These include applications of ML methods in the determination of numerical modeling parameters.

In a recent overview of ML in the Earth sciences or physical sciences in general, [Reichstein et al. \(2019\)](#) emphasize that advance of knowledge in the sciences, facilitated with the help of ML methods, requires development of novel NN approaches. Examples of methods that include non-ML physical data analysis modules in the NN operations flow/ architecture stem from oceanography (sea-surface temperature patterns, [De Bézenac et al. \(2019\)](#)) and biological applications ([Willis and von Stosch, 2017](#)).

3 Glaciology Background

3.1 Importance of Surging

Glacier surging is an important type of glacial acceleration, with surge-type glaciers found around the world in many but not all geographic regions, however the phenomenon remains poorly understood due to a relative paucity of comprehensive observational data and a lack of model application to actual, complex ice systems (Clarke, 1987; Truffer and Echelmeyer, 2003; Mayer and Herzfeld, 2000; Jiskoot, 2011; Herzfeld et al., 2013). A surge-type glacier experiences a quasi-periodic cycle between a long quiescent phase of normal flow and gradual retreat, and a short surge phase when the glacier accelerates to typically 10-200 times its quiescent speeds with heavy crevassing occurring throughout the ice system.

The recent surge of the Negribreen Glacier System (NGS), an Arctic glacier system located in eastern Spitsbergen, Svalbard, provides a rare opportunity to study a surge in a large and complex system (Lefauconnier and Hagen, 1991; Strozzi et al., 2017; Herzfeld et al., 2021). Beginning in 2016, the NGS began to surge with acceleration and heavy crevassing within 10 km of the terminus (Strozzi et al., 2017; Haga et al., 2020; Herzfeld et al., 2021; Trantow and Herzfeld, 2023). Largest surge speeds of around 22 m/day, equivalent to 200 times the glaciers quiescent flow velocity, occurred during the height of the acceleration phase in July 2017 (Herzfeld et al., 2022).

Negribreen last surged in 1935/36 (Lefauconnier and Hagen, 1991), which indicates that the quasi-cycle of the surge in the NGS is approximately 80 years. From a methodological point of view, it is worth noting that there has been no opportunity for modern data analysis and study of the Negribreen surge process prior to the current surge – this example indicates how the relative paucity of surges limit our ability for their study, but also that the Negribreen surge has provided a unique opportunity to advance several branches of science, mathematics and engineering (Herzfeld et al., 2020, 2022). Relevant to the study in this paper, the NGS has provided a unique collection of ice surface structures and crevasses types in close proximity, for an Arctic glacier system, and thus enabled the ML work reported here.

3.2 The Surge in the NGS

Negribreen is located on Spitsbergen in Svalbard, Norway, with the calving front at approximately (78.57°N, 19.083°E) in 2019, approximately 1000 km south of the North Pole. Negribreen receives most of its inflowing ice from the accumulation zone above the glacier to the west called Filchnerfonna and its northern part, the Lomonosovfonna, through Transparentbreen, Opalbreen and the Negribreen ice falls. The NGS, as defined by the blue contour in Figure 1(a), has an ice extent of approximately 500 km². The main glacial trunk, referred to simply as Negribreen, is fed by several major tributaries: Rembebreen to the south, and to the north, Akademikarbreen and Ordonnansbreen. Rembebreen and Petermannbreen (southwest of Negribreen) flow out of a southern

part of the Filcherfønna. Ordonnansbreen does not flow out of an icecap, but its source areas are mountain cirques. The area of the NGS is classically referenced as 1180 km² (Lefauconnier and Hagen, 1991), based on the extent of the glacier system at a time when Petermannbreen and Gardebreen (east of Ordonnansbreen) and their tributaries were still connected to Negribreen.

The NGS is a polythermal glacier, consisting of ice at and below the pressure melting point, and a marine-terminating (tidewater) glacier with ice calving into Storfjorden and the Arctic Ocean. Like other tidewater glacier surges in Svalbard (e.g., Strozzi et al. (2017); Sevestre et al. (2018); Nuth et al. (2019)), Negribreen began accelerating near the terminus after a collapse near the glacier front (Trantow and Herzfeld, 2023). Surge effects, such as heavy crevassing and elevated velocities, proceeded to propagate upglacier through the end of 2020 when it reached the NGS boundary with Filchnerfønna 30 km upglacier from the terminus (Trantow and Herzfeld, 2023). Mean ice-speeds remain significantly elevated in 2023 relative to quiescent speeds, with a maximum of 4m/day near the calving front, though ice-speeds have been decreasing steadily since the peak in 2017 (see Fig. 1(c)). High velocities, large-scale crevassing and enhanced calving during the surge has led to rapid disintegration of the system and large mass loss (Trantow and Herzfeld, 2023), thus contributing a significant amount to annual sea level rise during the surge years. Examples of surge crevasses are shown in aerial photographs in Herzfeld et al. (2022); Trantow and Herzfeld (2023).

3.3 The Crevasse-Centered Approach

Because analysis of crevasse patterns takes a central role in the physical part of our ML approach, we give a brief background summary on the role of crevassing in glacial acceleration and to the utilization of the crevasse concept in data analysis and modeling. The central idea of the crevasse-centered approach is that dynamic signatures of fast-moving ice and glacial acceleration are imprinted in ice in the form of crevasses and consequently the deformation history of a glacier can be reconstructed through analysis of crevasse patterns. Structural geologic principles provide links between dynamics, kinematics and deformation, which can be physically formalized and quantified using continuum mechanics, and simulated in numerical models (Means, 1976; Suppe, 1985; Twiss and Moore, 1992; Ramsay and Lisle, 2000; Liu, 2002; Greve, 2003; Trantow and Herzfeld, 2018).

Crevasses can be characterized using generalized spatial surface roughness, which is a mathematical approach that utilizes parameters derived from spatial statistical functions to capture spatial properties of a surface (Herzfeld, 2008). Roughness-based characterization applies to both crevassed and non-crevassed ice surfaces and thus allows to map an entire glacier. The approach of combining structural geology and mathematical roughness analysis to derive deformation characteristics in fast-moving glaciers is described in theory in Herzfeld et al. (2004) and has been applied to map deformation provinces in surging and continuously fast-moving glaciers throughout the cryosphere (Herzfeld and Mayer (1997); Herzfeld (1998, 2008); Herzfeld et al. (2000); Mayer and Herzfeld (2000, 2001); Herzfeld and Zahner (2001)). Applications of other approaches to structural glaciology have been reported by (Vornberger and Whillans, 1990; Marmo and Wilson, 1998; Rist et al., 1999). These studies have shown that observations of crevasse patterns and surface roughness can be used

as a source of geophysical or glaciological information.

Furthermore, the crevasse-based geophysical information obtained from remote sensing observations, such as satellite imagery, can be utilized in numerical models. [Trantow and Herzfeld \(2018\)](#) use Landsat-7 imagery of Bering Glacier, Alaska, during its peak surge phase in 2011 to derive crevasse locations, based on roughness characterization, and crevasse orientations, which are also modeled by simulating the stress regime in a 3D, full-Stokes finite element model. Differences in crevasse characteristics are minimized by optimizing important surge-model parameters such as the basal friction coefficient. This method is extended in [Trantow \(2020\)](#) to include other sources of model-data comparisons, such as surface velocity, which allows the optimization of additional model parameters such as those related to ice rheology.

Ice velocity observations are popularly used to constrain unknown model parameters (e.g. [Larour et al. \(2014\)](#)), however, during a surge the large-scale non-linear dynamics complicate velocity determination ([Fatland and Lingle, 1998](#); [Trantow and Herzfeld, 2018](#)), particularly on short time scales relevant to a surge. Therefore, crevasse observations are our most reliable source of dynamical information during peak surge activity and can be used to derive and optimize basal sliding laws for modeling a surge phase (?).

Crevasse classes, like those derived in the present paper, offer a more sophisticated picture of a glacier’s dynamic and structural state compared to simple crevasse-location and crevasse-orientation characterization. With more detailed geophysical information from crevasse classification, we expect to provide better constraints for a numerical model allowing more optimal parameterization, better error correction for input data sets such as bed topography, and ultimately more realistic simulation of glacial acceleration and its resulting effects on SLR and the evolving cryosphere.

4 Summary of the Approach

4.1 Objectives, Summary of Approach, Classification and Analysis Steps

The main objective of this paper is the exploration of the trade-offs between a physically constrained NN, a CNN (“Deep Learning”) for a specific, but generalizable, problem in the geosciences: The classification of crevasse types that form during the surge in an Arctic glacier system, the Negribreen Glacier System, Svalbard, to derive objective information about the evolution of the surge. To achieve this objective, we create a software system, termed *GEOCLASS-image* that facilitates classification of surface features from high-resolution satellite imagery and other imagery, perform testing and quality assessment (Q/A) of the software system, and release it as core of an associated Cyberinfrastructure.

Based on the results of the two trade-offs studies, we derive an example of a ML approach that combines the advantages of a physically constrained, classic NN with those of a CNN, thereby creating a physically constrained NN with a combined architecture, that will be termed VarioCNN. The final VarioCNN is applied to a time series of WorldView images, to derive information on the evolution of the surge in an Arctic Glacier System, the Negribreen Glacier System.

The combined NN, VarioCNN, will be applied to a time series of WorldView-1 and WorldView-2 images, collected in 2016 – 2018 during the acceleration stage and mature stage of the the surge in the NGS. Each image will be analyzed individually and provide an element in a time series of thematic maps of crevasse provinces. The goal is to derive geophysical information on the evolution of the surge during these core stages. Specifically, we aim to create a classification of crevasse patterns, as they relate to deformation types that occur as a result of ice-dynamic processes. Crevasses are manifestations of the local strain state of the ice. Occurrence of fresh crevassing indicates the expansion of the surge, and as the surge progresses, new types of crevasse patterns form. The time series of crevasse maps will be interpreted geophysically. Lastly, we provide a description of the GEOCLASS-image software system.

In summary, the work in this paper builds on the following three ideas:

- (1) **Employ geostatistical parameters as mathematical formulation for physically informed extraction of complex information from imagery**
- (2) **Utilize different NN types as connectionist association structures: MLPs and CNNs**
- (3) **Compare and then combine the NNs into a three-tiered approach: Geostatistical-connectionist with MLP and CNN**

4.2 Approach Steps

Objectives of the work in this paper are the following:

- (1) Create a software that
 - (1.1) encompasses the main principles of the connectionist-geostatistical classification method,
 - (1.2) is sufficiently tested/ robust/ quality-assessed to form the center-piece of a community software for image classification in the geosciences and beyond,
 - (1.3) has a user-friendly GUI for image manipulation, selection of training data, through classification,
 - (1.4) facilitates training and classification of several crevasse types
 - (1.5) allows analysis of different types of satellite imagery
 - (1.6) includes utility tools for cartographic projections and other image manipulations,
 - (1.7) includes several Neural Network Types, including Multi-Layer Perceptrons, Convolutional Neural Networks, and
 - (1.8) is open to generalization to more architecture types
- (2) Explore the trade-offs between a physically constrained NN and a CNN for a specific, but generalizable, problem in the geosciences: the classification of crevasse types that form during a glacier surge
- (3) Create an example of a ML approach that combines the advantages of a physically constrained, classic NN with those of a CNN, thereby creating a physically constrained NN with a combined architecture, and
- (4) Apply the resultant NN to a time series of WorldView images, to derive information on the evolution of the surge in an Arctic Glacier System, the Negribreen Glacier System.

4.3 Terminology

We use the following terms to distinguish ML approaches and NNs in this paper; further explained in sections (6), (7), (8), (9) and in the Appendix A.

- (1) The **connectionist-geostatistical classification method** (Herzfeld and Zahner, 2001) is the original approach that combines a physically driven geostatistical analysis of an input data set and a neural network into a ML approach. As described in Herzfeld (2008), the geostatistical analysis or characterization can take several different forms, in any case, the output of the geostatistical analysis is used as input for the neural network. Examples of geostatistical analysis include (a) the experimental variogram, a discrete function, and (b) results of

geostatistical characterization parameters. The neural network type applied in most of our studies is generally a form of a multi-layer perceptron (MLP) with back-propagation of errors (Herzfeld and Zahner, 2001; Herzfeld et al., 2016; Herzfeld, 2008; Herzfeld et al., 2023b) (see section (6)).

- (2) The acronym **VarioMLP** is used for connectionist-geostatistical NN type that is applied in this paper; it employs an four-directional experimental vario function to activate the input neuron of a MLP with back-propagation of errors (see section (6)).
- (3) The term **Convolutional Neural Network (CNN)** stands for a specific class of neural networks that realize the concept of “deep learning” (Goodfellow et al., 2016; Camps-Valls et al., 2021b).
- (4) **ResNet-18** is the acronym for the specific CNN used in this paper (He et al., 2016a,c) (see section (8)).
- (5) The acronym **VarioCNN** will be used for the combined new method that integrates VarioMLP and **ResNet-18** into a unique, physically constrained ML approach (see section (9)).
- (6) Specific architectures of a NN are identified by adding information in square brackets, for example, VarioMLP[18, 4,(5,2)] identifies a VarioMLP, where 18 is the number of steps in the vario function (for each direction), 4 the number of directions of vario-fcuntion calculations, yielding 72 nodes in the input layer, and (5,2) the factor in the number of nodes of hidden layers; here a MLP with two hidden layer is used, where the first layer includes 72 times 5 nodes and the second layer 72 times 2 nodes (see section (7)).

More generally, $VarioMLP[n_{steps}, n_{dir}, (m_1, \dots, m_{nl})]$ identifies a VarioMLP, where n_{steps} is the number of steps in the vario function (for each of n_{dir} directions) and $(m_1, \dots, m_{nl})$ with $nl \in \mathcal{N}$ the factor in the number of nodes in nl hidden layers; here a MLP with nl hidden layers is used, where layer i has m_i n_{steps} nodes for $i = 1, \dots, nl$ (see section (7)).

- (7) **GEOCLASS-image** is the software system utilized to create the neural networks and labeled data sets referred to in this paper and carry out the classifications of crevasse types during the surge of the NGS, Svalbard (Herzfeld et al., 2023a) (see Appendix A)).

5 Approach Component: Image Classification and Data Sources

5.1 Image Classification Challenges and Approaches

The data analysis challenge is a type of image classification, more specifically, image segmentation. Different types of image classification are the following: (1) Each image is associated to a class, (2) features are extracted from images (an often-analyzed example is the detection of moving features between consecutive images; e.g. planes at terminals (He et al., 2021)). (3) application of the image classification to videography, i.e. time series of images (e.g. Herzfeld and Zahner (2001)) or satellite videography (He et al., 2021), (4) Segmentation of a single image into areas of several different classes, resulting in thematic maps. Early applications in the geosciences, e.g. sea-ice classification, land-cover classification, fall in this category. Typically, the classification is applied as a moving-window operator (i.e. to subimages, which can overlap). From a classification standpoint, the types of image arrangement can all be treated the same way, with different data handling utilities. Challenges in this context lie in the specifics of the observational data, which may include remotely sensed imagery from any tier of observation (satellite, airborne, subaerial, ground), in the specific spatial and spectral resolution of the sensor, signal-background separation, and other characteristics of image. The problem treated in this study is a combination of (4) and (1), applied to a time series of satellite image data of the glacier surface. The classification will be applied to each image individually (i.e. without providing information on the previous image).

Because a surge in an Arctic glacier extends over several years, typically 7-10 (Murray et al., 2003b), data from several different satellite sources need to be integrated in an analysis. Here, we utilize Maxar (formerly DigitalGlobe) data from the WorldView-1 and WorldView-2 satellites. Both satellites carry high-resolution multispectral optical image sensors, but with different resolutions and spectral channels (see, Table 1). Thus, a specific challenge lies in the identification of subimages for training that work for both satellite data types.

In order to facilitate application of our classification system GEOCLASS-image to data from different sources, a large range of data handling utility modules is included (see, software description). The software is designed to be generalizable to several data types, both (a) for different studies, using a single data type, and (b) to integrate data from several sources into one classification.

The classification will be trained using a set of labeled images. A challenge in a spatially based classification, but also in any image classification that uses subimages (or: splitimages), is the selection of a subimage size that is large enough to include several repetitions of the crevasse pattern, but also small enough to be sufficiently homogenous to be assigned to a single class.

5.2 Data Sources and Processing

The analysis in this paper utilizes Maxar WorldView-1 and WorldView-2 optical satellite image data. WorldView-1, WorldView-2 and WorldView-3 data are a widely used type of commercial satellite imagery (Neigh et al., 2013). Hence the classification approach described in this paper is relevant to large parts of the Earth science community. For example, an Arctic-wide Mosaic and DEM has been created from WorldView data (Porter et al., 2022). Immitzer et al. (2012) use a random forest classification applied to Worldview-2 imagery to identify tree species in the forests of Austria at high resolution. WorldView data is used heavily as the data source for classifications by the Land-Cover/Land-Use and the Vegetation Ecology Communities (e.g., Elsharkawy et al. (2012); Immitzer et al. (2012); Aguilar et al. (2013); Koc-San (2013); Ghosh and Joshi (2014); Waser et al. (2014); Li et al. (2015); Gaertner et al. (2017); Melville et al. (2018); Vigneshwaran and Vasantha Kumar (2021); Caraballo-Vega et al. (2023)).

The Worldview-1,2,3 satellites, owned and operated by Maxar (formerly DigitalGlobe), provide submeter optical imagery of much of the cryosphere (Porter et al., 2022), including all of Negribreen. WorldView-1 carries a single high-resolution optical imager called the WorldView 60 camera, which has a single panchromatic channel with a spectral range of $0.45\ \mu\text{m}$ - $0.90\ \mu\text{m}$. WorldView 60 is a pushbroom sensor operating in a swath of 17.9 km with 0.5 m resolution at nadir down to 0.55 m resolution 20° off nadir. (see Table 1). WorldView-2 also carries a single high-resolution optical imager called the WorldView 110 camera, but has two operational channels. The first is panchromatic channel with a spectral range of $0.45\ \mu\text{m}$ - $0.80\ \mu\text{m}$, while the second is an 8-band multispectral channel ranging from $0.4\ \mu\text{m}$ - $1.05\ \mu\text{m}$. WorldView 110 is also a pushbroom sensor with a swath-width of 16.4 km with 0.46 m resolution at nadir down to 0.52 m resolution off nadir. A full comparison of the WorldView-1 and WorldView-2 specs is given in Table 1.

In this analysis, we utilize panchromatic imagery from WorldView-1 (launched 18 September 2007, decommissioned September 2023, Earth Observation Portal (EOPortal) (2023a)) and WorldView-2 (launched 8 October 2009, remains operational in 2024, Earth Observation Portal (EOPortal) (2023b)) to analyze the NGS surge from its start in 2016 through 2019. Data from the panchromatic channel will be utilized for the classification approaches in this paper, because it has the highest spatial resolution for each satellite (0.45m pixel size for WorldView-1 data and 0.42m for WorldView-2 data) and thus retains the most information on spatial properties of the ice surface. While we will not employ data from the other spectral channels, we have described statistical and geostatistical image classification approaches for multispectral data elsewhere (Herzfeld et al., 2016).

The VarioMLP has also been applied to classify Negribreen crevasse provinces based on Planet SkySat data (Herzfeld et al., 2021, 2022) (for data description, see (Fiuczynski, 2006; Peterson et al., 2006)). Other commonly used satellite imagers include Landsat (Wulder et al., 2019) or Sentinel-2 (Drusch et al., 2012).

Processing. Images were selected w.r.t spatial coverage, temporal coverage and lack of obfuscation.

The GEOCLASS-image system includes a tool for evaluation of the area of overlap of a given WorldView image with the area of interest, as outlined by the polygon encompassing the NGS region (Fig. 1(a)). Only images with a 50% or more overlap with the NGS area of interest were used for analysis. To avoid obfuscation, (1) images with a high percentage cloud cover over the area were avoided, as were images with a deep snow cover on the glacier, which would obliterate the crevasse patterns. Thus winter images were rejected. Applying these criteria, 11 high-quality images from spring and summer 2016-2018 were selected from several hundred WorldView data sets (Table 2). All 11 images were used for creation of the labeled training data sets of split images, whereas 7 are used in the time-series analysis (leaving out images that are too close in time to other images already in the time series).

The pixel intensity of the source geotiff images was normalized to the 95th percentile and compressed to an 8-bit range. This was done because although the Worldview-1 and Worldview-2 sensors allow for 11-bit depth, only a small number of pixels in each image actually reached those intensity values, the rest being several orders of magnitude below. By normalizing to an 8-bit range data processing during training and testing was made much faster and more efficient, and by thresholding the highest-intensity pixels, the resulting geotiff images and the split images extracted from them were much easier to view for the human eye, a crucial component for a labeling process.

Custom software was written to extract the coordinates in both pixel space and UTM space for each split image within a given set of geotiff images which falls 100% within the NGS area of interest. These coordinates, along with fields for a class label, class prediction, confidence, and enumerator to reference the source geotiff images were stored in a large table in addition to metadata containing the filepaths, affine transforms to convert between pixel and UTM space, and class enumerations. Thus, this pipeline produces a standardized and efficient split image dataset format which can be utilized by the classification model, visualization tools, labeling tools and utility tools to reproducibly extract the split images from their corresponding source images at runtime. For the 11 source geotiff images selected (Table 2), a total of 108,623 split images were extracted using this pipeline. A breakdown of the total number of split images from each source geotiff is also given in Table 2.

6 The Connectionist-Geostatistical Classification Method

The connectionist-geostatistical classification method (Herzfeld and Zahner, 2001; Herzfeld, 2008) integrates and interleaves physical knowledge, spatial statistical analysis and computational components at several levels. The approach includes the following concepts:

- (1) The idea of using spatial classification to extract features from image data
- (2) The idea of using geostatistical parameters to pre-process the imagery
- (3) The vario function and residual vario function
- (4) Creation of input data to activate the input-layer neurons of the NN
- (5) The feed-forward multi-layer perceptron with back-propagation of errors

The idea of the connectionist-geostatistical classification method is to utilize geostatistical parameters to pre-process the input image data, thereby reducing the complexity of a NN required to identify spatial structures that are surface signatures resultant from cryospheric processes. The relationship between glacial acceleration, crevassing and resultant spatial structures reflected in imagery has been explained in section (3.3) and in more detail in (Herzfeld, 2008). In this section we describe the mathematical and computational ingredients of this approach, in the form that is employed in VarioMLP, the type of connectionist-geostatistical classification utilized as a component of VarioCNN.

6.1 Geostatistical Processing of the Input Image Data

6.1.1 Spatial Homogeneity

Depending on the type of image classification problem at hand, an input image can be a video frame, a photograph, a subset of a video frame or photograph, or a subimage of a large image such as a satellite image (termed split-image here). Split-images are created from satellite images by a moving window process. The goal is to associate each image to a surface class, here a crevasse class, using the classification method. Considering the entire classification a moving-window operation applied to a satellite image, a segmentation of the area of the satellite image into crevasse classes will be obtained, in other terms, a thematic map of structural glaciologic provinces. Similarly, a time-dependent segmentation of a video stream of a glacier will result in a mapping of surface classes.

To allow for characterization and classification of the spatial structures captured, the optimal size of a subimage (split-image) is determined as follows: A feature type needs to repeat approximately

three times in the split-image, and the split image needs to be spatially homogeneous with respect to surface structure, here, crevasse type. These two criteria will not be met exactly across an entire glacier region, thus a split image size needs to be selected that meets the criteria sufficiently often to make the classification operational. For experiments with only VarioMLP, split-images of sizes 201 pixels by 268 pixels were used, which follow the (3-4-5) rectangle convention and are approximately 123 m by 92 m for WorldView-1 data. An additional constraint is that the structure requires input imagery of 224 by 224 pixel sizes. For WorldView data and Negribreen surge crevasses, this requirement can be met, however, it limits generalizability of the approach. The entire training was rerun for the combined architecture of VarioCNN using square images.

6.1.2 Vario Functions

In order to characterize the spatial surface structure in a given area, recorded in an image or subimage, we calculate vario functions, defined as follows:

$$v_1(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_i + h)]^2 \quad (1)$$

for pairs of points $(x_i, z(x_i)), (x_i + h, z(x_i + h)) \in \mathcal{D}$, where $\mathcal{D}$ is a region in $\mathcal{R}^2$ (case of profile data) or $\mathcal{R}^3$ (case of image data) and n is the number of pairs separated by h ; the distance value h is also termed “lag”. The function $v_1(h)$ is called the *first-order vario function*. This function exists always and has a finite value.

The residual vario function is often more useful to analyze roughness in situations where a regional trend or a local drift underlies the data. Using

$$m(h) = \frac{1}{n} \sum_{i=1}^n [z(x_i) - z(x_i + h)], \quad (2)$$

the residual vario function is defined as:

$$res_1(h) = v_1(h) - \frac{1}{2}m(h)^2. \quad (3)$$

First-order vario functions are formally equivalent to variograms of geostatistics, but introduced in a discrete mathematics framework that facilitates easy numerical implementation as well as generalization to higher order (Herzfeld, 2008).

The variogram is defined for a data set that may be considered a realization of a spatial random function satisfying the intrinsic hypothesis (see Matheron (1963, 1973)), for which generalization to higher order is difficult because of the statistical assumptions that need to be met. Equation (1) corresponds to the statistical second-order moment and equation (2) to the first-order moment. Residual vario functions work best for data that underly a trend. The second-order vario function

and residual vario function can also be used (see, Table 3). Numerical outputs of the first-order vario function have been used in the original connectionist-geostatistical classification in (Herzfeld and Zahner, 2001), they correspond to the experimental variogram values.

In VarioMLP, first-order vario functions are calculated sampling along four directions of each image, paralleling each side and the two diagonals (see, Figure 2). An efficient sampling algorithm makes use of the matrix structure of the 2D image. The sampling algorithm in (Herzfeld and Zahner, 2001) utilizes images of relative sizes (3-4-5), where 3 and 4 are the relative lengths of the split image sides and 5 the diagonal. However, this cannot be transferred to training, which requires square images.

The discretization of the vario function is determined by the lag value h in pixels. In the final implementation of the algorithm, the lag value is determined such that 18 lag steps exhaust 80% of the image size (244). The values of

$$(v1(1), \dots, v1(m)) \quad (4)$$

become the activation values for the input layer of the MLP for any value of $m \in \mathcal{N}$, in our final structure it will be $m = 18$. Accounting for the 4 directions of directional vario functions calculated for each split image, we have a matrix of input values

$$V = (v1_{(i,j)})_{j=1, \dots, n_{dir}}^{i=1, \dots, m} \quad (5)$$

With $n_{dir} = 4$ for the number of directions, the number of input values is $m_{in} = n_{dir}m = 72$.

6.2 NN Architecture: Multi-layer Perceptron (MLP)

The NN structure of the connectionist-geostatistical classification is a multi-layer perceptron with back-propagation of errors (MLP-BP or simply MLP). The MLP has an input node per vario-function value, in the final VarioMLP structure $m_{in} = 4m = 72$, accounting for 18 lag steps and 4 directions. MLPs have been found to be useful NN types for the solution of this type of classification problem.

The number of nodes (neurons) in the output layer has to equal the number of surface classes, here crevasse classes. In our experiments, this number is $m_{out} = 6$. Larger numbers of crevasse classes have been used, ranging up to 18. In (Herzfeld et al., 2023b), we describe a classification with up to 13 crevasse classes.

This leaves the number and size of internal, hidden layers as variables of the NN architecture that

will be determined experimentally (see section (7.7.2)). The original work in [Herzfeld and Zahner \(2001\)](#) uses a single hidden layer, in fully connected or partly connected architectures. Here we experiment with two or three internal layers.

7 Image Labeling and Training Approach (for VarioMLP and ResNet-18)

7.1 Training Approach

The training approach reflects the goal of creating a physically constrained NN by combining knowledge of glaciological processes and Earth observation technology with ML methods at every step. In the last section, we already saw that the selection of sizes of training images is controlled by a requirement of spatial homogeneity, constraints associated with the spatial resolution of the satellite imagery, and the spacing of crevasses on the glacier surface, which results from the glacial movement and acceleration and we aim to analyze. Training is carried out as a form of supervised training; training as such is an optimization problem of the model’s internal parameters.

7.2 Crevasse Classes

Crevasse classes are selected by an expert, based on structural glaciology (section (3.3)). Because a main objective of this paper is the integration of a physically constrained NN and a CNN, we utilize (only) four basic crevasse classes: (a) one-directional crevasses, (b) multi-directional crevasses, (c) shear crevasses, and (d) chaos crevasses, or shear-chaos crevasses. The crevasse types associated with these classes are illustrated in Figure 3. Crevasse types (a), (b) and (c) are associated with basic deformation matrices (Herzfeld et al., 2004): The one-directional crevasse type results from an extension in one direction (Fig. 3a). The multi-directional, including two-directional, crevasse type results from a deformation with more than one stress axis (Fig. 3b). It can also result from two deformation processes that affect the material ice in sequence. The shear crevasse type results from shear, a deformation type that typically occurs when fast-moving ice borders slow-moving ice. In the case of a surge, the ice of one glacier (Negribreen) accelerates, while the ice of an adjacent glacier (e.g. Ordonnansbreen) continues to flow at normal, much slower speeds (Fig. 3c). Depending on the spatial and temporal velocity gradient, shear crevasses can take different appearances (Fig. 3c and 3d). Transportation, weathering and interaction of several deformation processes can lead to complex ice-surface and near-surface structures, in which the signatures of individual processes cannot be distinguished any more, thus they are summarized as “chaos” crevasse class (Fig. 3d). In some areas, the signature of shear deformation is still evident in the chaos crevasse fields (Fig. 3f), but separation in an image classification process may be too difficult, thus the class is summarized as chaos/shear-chaos. Two additional classes need to be added to each classification, one for undisturbed snow/ice and a rest class for “other” surfaces, which can include moraines, rock avalanches, subimages that include snow/ice and rock surfaces, and indiscernible images, to limit misclassification of the better defined 4 crevasse classes. A rendering of representative examples of split images, subselected from WorldView satellite imagery, is seen in Figure 4. The images have a size of $201(=3*67)$ pixels by $268(=4*67)$ pixels, i.e. the follow the (3-4-5) size rule.

7.3 Image Labeling

A second main objective of this paper is the derivation of a labeled training data set for the problem of crevasse classification from satellite imagery. With this objective, we address the problem that application of ML in the geosciences and specifically the cryospheric sciences has been hampered by the lack of labeled training data sets, as identified by authors working in the field (e.g. Virts et al. (2020); Liu et al. (2020b); Meyer and Pebesma (2021)) and described in more detail in section (2).

To initiate the training, sets of split-images for each class are identified and selected by the structural glaciologist. In our experiments, we found that several tens of example images per class are sufficient for an initial training run of VarioMLP.

Technically, image labeling is carried out using the Split Image Explorer Tool, visualized in Figure 5, described in more detail in Appendix A. Individual images can be selected from the WorldView image, optionally with a polygonal area of interest outlined that contains the glacier area, viewed enlarged at the top left, and associated to a class. The association can be (1) performed initially by the glaciologist, or (2) displayed as the result of the NN classification, or (3) overwritten (accepted or rejected) in a control pass in the training loop (see section (7.6)). A sliding bar in the left middle of the explorer toll allows application of confidence as a filter for visualization (only images classified with a confidence level exceeding the user-selected confidence threshold are displayed in color).

7.4 Data Handling and Feature Engineering

Feature engineering is the design of the input for the neural network. Of importance for robustness of the results is that identification of a crevasse type is independent of orientation and view angle of the satellite, relative to features on the ground. Directional bias is removed by calculating vario functions in several different directions for each split-image.

Prior to extraction of split-images, the satellite image needs to be oriented in a geographic or rectangular projection framework that facilitates output of the final classification in the form of a thematic map of crevasse provinces. Raw satellite imagery is typically collected along orbits and constrained by the view angle of the observatory, which is fixed for some imagers, but adjustable or sweeping for most (including WorldView). To accomplish mapping larger areas from a single or multiple satellite images, utility functions for image projection and mosaicking are implemented as part of the GEOCLASS-image system. To visualize, the reader may compare the different sizes and orientations of the input imagery shown in Figure (6).

Data from the panchromatic channel of WorldView-1 and WorldView-2 are utilized, because the classification principle is a spatial classification. In the more common form of multivariate statistical classification, data from several spectral channels are used. Our study combines imagery from

two different image systems, WorldView-1 and WorldView-2, which result in imagery of somewhat different pixel size and resolution (0.45 m for WorldView-1 and 0.42 m for WorldView-2, see section (5.2)). A utility function in GEOCLASS-image facilitates simultaneous analysis and classification of imagery from both satellite types.

Application of the vario function to a typical image from the classes of (1) undisturbed snow and ice surfaces and (2) one-dimensional crevasse types, seen in Figure 2, illustrates how the NN can separate these crevasse types based on the vario function values for different directions and distances. First, the maximum of the resultant vario function values is much lower for undisturbed surfaces than for crevassed surfaces (compare the v_1 axes in Fig. 2c and Fig. 2d). Second, an anisotropic behavior of the set of directional vario functions is typical for one-directional crevasses (Fig. 2b, 2d), where the direction that is near-parallel to the crevasse direction does not reach the sill of the vario function (green in Fig. 2d), whereas the other three directional vario functions exhibit a typical wavy pattern resultant from washed out cross-correlation, with spacing dependent on the relative angle of the crevasse orientation to the directional calculations.

7.5 Criteria for Evaluation of Training Success

We use the terminology of *intrinsic criteria* for quantitative, computational criteria (cross-entropy measure of training loss, confidence of classification result, co-occurrence matrix) and *extrinsic criteria* for glaciological criteria which are typically based on airborne field observations of the glacier system during surge and on additional expert knowledge on the evolution of crevasse types during a surge (Herzfeld and Mayer, 1997; Herzfeld, 1998; Mayer and Herzfeld, 2000; Herzfeld et al., 2013, [subm](#); Trantow and Herzfeld, 2018). The application of extrinsic criteria is best explained in an applied example of image labeling and in the geophysical interpretation (see sections (7.6) and (11)).

7.5.1 Softmax Function

A softmax function is used to convert the NN output layer to a probability distribution for the possible classes. Each output node is assigned a value between 0 and 1 (p_i), with all outputs summing to 1, so that they can be interpreted as probabilities. The class with the largest probability is selected as the NN's final classification of a given input and the confidence of the classification result is equal to that probability, i.e., the maximum of the softmax function. The loss function associated with the softmax function is given by the cross-entropy loss, which is used for training purposes (see, section 7.5.2). The softmax function is commonly used in many CNNs (Krizhevsky et al., 2012; He et al., 2016a) due to its simplicity and probabilistic interpretation.

7.5.2 Cross Entropy

Training an MLP is an optimization of the model’s internal parameters, carried out iteratively. At each iteration, VarioMLP predicts the class of each training example and uses the cross entropy loss function as a quantification of the difference between predicted values and training data. Entropy was first introduced in [Shannon \(1951\)](#) to quantify the level of uncertainty of a random variable X based on possible outcomes p_i according to

$$H(X) = - \sum_{i=1}^n p_i \log(p_i) \quad (6)$$

for $i = 1, \dots, n$ and $n \in \mathcal{N}$ is the number of classes. For VarioMLP, the outcomes are the crevasse classes and the probabilities are those which the model assigns to each output neuron. The DDA-MLP uses cross entropy loss as its loss criterion, calculated as

$$Loss_{ce} = - \sum_{i=1}^n t_i \log(p_i) \quad (7)$$

where n is the number of classes, t_i is the truth label for class i , and p_i is the model-predicted probability for class i as its loss criterion. The optimization problem is then for the model to learn an internal parameter set which minimizes this loss function, and to accomplish this the DDA-MLP employs stochastic gradient descent (SGD) via the Adam algorithm for first-order gradient-descent based optimization problems introduced by [Kingma and Ba \(2014\)](#). During training, backpropagation, as defined in [Hecht-Nielsen \(1992\)](#), involves computing the gradient of the loss layer by layer, starting from the output and moving backward towards the input layer. In this case, the Adam algorithm for SGD only computes the first-order gradient, and employs adaptive learning rates for parameters based on estimates of first and second order moments, and updates the parameters proportionally to the learning rate hyperparameter in the direction of steepest descent of the gradient ([Kingma and Ba, 2014](#)). Application of cross-entropy loss for training of deep NNs is described in [Zhang and Sabuncu \(2018\)](#).

Cross-entropy loss is utilized to identify functional training runs and reject training mistakes. For example, overfitting in a test-run of the model is illustrated in Figure 7.

7.5.3 Confidence

Classification confidence is a measurement of the probability that the association of an input image to a class is correct. Confidence approaches have been discussed in [Papadopoulos et al. \(2001\)](#). We

utilize confidence to accept or reject classified crevasse images into the training data set, applying a threshold of 90% confidence. The Split-Image Explorer Tool allows user-selected confidence values.

7.5.4 Other Training Hyperparameters

Overall, the training and feedback-loop experiments were repeated several times with different parameters and variations of the classification models. The split of the training data into actual training images and validation images was held constant at 80% (training) and 20% (evaluation) for all experiments training VarioMLP and ResNet-18. This means that however many labeled training images existed for a given run, 80% were randomly selected at runtime for the actual training process, and 20% were reserved to evaluate model performance after each epoch. It is important to separate the training and evaluation datasets, because if the model is not evaluated on images it did not see during training, it will simply memorize the training dataset if the model is sufficiently complex. Each training run was carried out with a maximum number of 50 epochs. For each epoch which resulted in a new best validation loss, a checkpoint of the classification model was saved for further evaluation. For all training experiments, cross entropy loss was used with the Adam optimizer as the method for gradient descent calculation (section 7.5.2).

7.6 Interleave of Split Image Labeling with the Training Process: The Feedback Loop

Following creation of an initial set of expert-labeled training data, a VarioMLP is trained. The resultant network architecture can be applied to simply classify an entire satellite image. However, in order to derive a large data set of labeled training images, an iterative approach to split-image labeling and VarioMLP training is taken. The goal is the creation of a data set that is large enough to train a CNN, which in turn can be expected facilitate rapid classification of many satellite images for similar problems, i.e. a higher level of generalization of the task of crevasse classification.

The iterative approach is implemented in as a feedback loop in VarioMLP, executed as a mix of computational criteria and expert interaction, interleaved in the training process of VarioMLP as follows (see, Figure 8). The initial data set is considered the first-order data set, used to train the NN. Validation loss and training loss are evaluated as quantified by the cross-entropy measure (see, section 7.5.2). A trained VarioMLP architecture results.

The VarioMLP, with first-approximation final structure, is then applied to classify the entire set of all split-images from a given satellite image (all split-images inside the polygon that outlines the NGS). Each split image is associated to a class and written out into a directory of that class. Next, only split images with a classification confidence at least 0.9 are retained in the crevasse class directories. Then, the glaciology expert quickly views all new images in each class (i.e. any images that were not part of the original labeled data set) and rejects images that are misclassified. This process is much faster, requiring a fraction of human expert time, than labeling thousands of split-images initially. The VarioMLP is then rerun, using the larger set of labeled data as training data.

By repeating the feedback loop, a labeled data set with 3933 images was obtained in a reasonable amount of time. The final labeled data set of 3993 split images includes between 522 and 953 images per crevasse class, with a distribution given in Table 4. This distribution is relatively even and not varied enough to cause a significant potential source of inaccuracy for the model training.

In this exemplary application, the expert that selected the initial data set was a glaciologist experienced in structural glaciology, especially observation of glacier surfaces during surges (the lead author of the paper), whereas in later iterations, the sorting of images was performed by a computer science student, indicating that the sorting procedure grows increasingly fast and simple, as the training goes through several iteration steps. To simplify the process, only a set of four main crevasse classes, plus undisturbed plus a rest class/chaos class were chosen for this study.

On the other hand, to ascertain general application of the labeled training data set to a range of previously unseen WorldView data sets from the NGS and other regions of surge glaciers, as well as to analysis and classification of data from WorldView-1 and WorldView-2, split-images were sourced from 11 different WorldView data sets collected over the NGS in 2016, 2017 and 2018. This resulted in a total of 108,623 split-images. The distribution of split images in the final 3933 data set per WorldView source files is given in (Table 2).

At this point, we have achieved two results: (1) The derivation of a labeled training data set, and (2) The VarioMLP together with the feedback loop as either a standalone NN or a component in a physically constrained CNN, the VarioCNN.

In the next sections, we will describe ResNet-18, the CNN component selected for VarioCNN, its training, comparison to VarioMLP, and finally design of the combined classification system, VarioCNN, and the classification software system, GEOCLASS-image. Experiments with VarioCNN, using GEOCLASS-image, are rounded off by geophysical application and interpretation of the evolution of crevasse provinces during the surge in the NGS.

7.7 Determination of VarioMLP Hyperparameters

The VarioMLP architecture includes hyperparameters which can be optimized to tune the model for testing performance and generalization. Both the Directional Variogram and Multi-Layer Perceptron steps of the VarioMLP architecture have hyperparameters which affect the training and testing in different ways. Input image size and resolution has already been discussed in section (6.1.1), as this is constrained by the observations technology, the surface signatures and the assumption of spatial homogeneity. To optimize the architecture of VarioMLP, experiments were carried out to determine the optimum number of lag steps in the vario function and the shape and number of the internal layers. In both series of experiments, cross entropy loss was used as the measure for assessment of training quality and network performance.

7.7.1 Number of Vario Function Steps

The optimal number of lag steps in the directional variogram becomes a function of the input image size in pixels. If the number of lag steps used is too small, the directional variogram may not be able to provide sufficiently different characteristics for a given set of surface types for reliable classification. If the number of lag steps is too high, the characteristics provided by the directional variogram can be polluted by noise and small-scale features that are present in multiple surface types. These characteristics may bury the salient features of each surface type needed for classification. During training, the lag step parameter was tested at values of 10, 12, 14, 16, 18 and 20 (Table 5). In this experiment, the hidden layer shape was fixed at [5,2] and the final validation data set included 786 images (20% of the final 3,933-image labeled data set.) The best performance was achieved with a lag step value of 18. It is interesting to note that performance is not correlated with the number of lag steps used in the Variogram phase. Rather, the model seems to perform relatively well with values of 12, 14 and 18, and relatively poorly with values of 10, 16 and 20.

7.7.2 Hidden Layer Structure in the MLP

The number of hidden layers in the MLP step of the VarioMLP architecture is a function of the size of the input layer, as well as the size of the training dataset. If the number of hidden layers is too large relative to the input layer size, then the model becomes unnecessarily complex and thus more susceptible to overfitting. Too few hidden layers produce the opposite problem - the model lacks the complexity necessary to capture the full variance of the dataset and suffers from underfitting. This is an example of what is commonly referred to in machine learning as the bias-variance tradeoff (Geman et al., 1992; Briscoe and Feldman, 2011; Belkin et al., 2019). Choosing a perfect model size and depth becomes increasingly difficult for problems where there is no existing reference dataset of labeled training examples, since as the size of the training dataset increases so too does the optimal fully-connected model size. However this relationship is nearly impossible to calculate, so trial-based estimation is necessary. To reduce the scope of this optimization during training, the shape of the hidden layers of MLP architecture were limited to being exact multiples of the input layer size. An MLP model denoted as [5, 10, 2] refers to a model with 3 fully-connected hidden layers, which contain 5, 10 and 2 times as many nodes as the input layer respectively. During training, model architectures of [2, 2], [5, 2], [5, 5, 2], [10, 5, 2], and [10, 10, 2] were tested (Table 6). For each test run, the lag steps for the variogram stage were fixed at 18. The best performing hidden layer shape was [5, 2]. For networks both wider and deeper than this, the performance significantly decreased. This is likely due to the fact that for the relatively small amount of information at the input layer (the concatenated output of the variogram stage), larger networks simply converge on memorizing the training dataset. This is another example of the bias-variance tradeoff at play, the network must not be overly complex for the scope of the input data.

8 ResNet-18

8.1 Description of the CNN ResNet-18

The term ResNet summarizes a family of convolutional neural networks with depths of up to 1001 layers (He et al., 2016a,c, 2021), based on residual learning. ResNet-18 is the deep learning network with the fewest internal layers that is commonly used today (e.g. Kochgaven et al. (2021); Ramzan et al. (2020); Jiang et al. (2023)) and the ResNet type that will be used for the studies in this paper. In medical sciences, labeled training data exist (Kochgaven et al., 2021). Notably, for image feature detection in medicine (brain tumors, alzheimers, to name a few) labeled data sets exist, which is not the case for image classification in geoscience studies.

8.1.1 Mathematical Principles of ResNets

The following mathematical description of deep residual networks (ResNets) is summarized from (He et al., 2016c), the basic paper that introduces ResNets. ResNets consist of many stacked so-called “Residual Units”. The defining equations of a ResNet are the following two:

$$y_l = h(x_l) + \mathcal{F}(x_l, \mathcal{W}_l) \quad (8)$$

$$x_{l+1} = f(y_l) \quad (9)$$

where x_l and x_{l+1} are input and output of the l -th unit, $\mathcal{F}$ is a residual function and $\mathcal{W}_l$ denotes the set of weights (and biases) associated with the l -th residual unit, which may consist of K number of layers itself ($K = 2$ or $K = 3$ are typical values). The residual units form the building blocks of the modularized architecture that characterizes a ResNet. In He et al. (2016a), only $h(x_l)$ is an identity mapping and f is a ReLU function, whereas in He et al. (2016c), both $h(x_l)$ and $f(y_l)$ are identity mappings. The work in He et al. (2016c) shows that in the case that both these are identity mappings, the signal can be directly propagated from one unit to any other unit; this finding leads to the definition of *skip connections*. The identity mapping $h(x_l) = x_l$ (already derived in their first paper, (He et al., 2016a)), achieves fastest error reduction and lowest training loss (among a number of model variants studied in (He et al., 2016c)). The use of the second identity mapping, $f(y_l) = y_l$, is a new interpretation of the activation functions (which can be, for example, ReLU, the function used in our ResNet-18 model) as so-called pre-activation of the weight layers, as opposed to a hitherto view of post-activation. This paper (He et al., 2016c) introduces a 1001-layer ResNet, which is easier to train and generalizes better than the original ResNet described in (He et al., 2016a).

The essential mapping of the ResNet is reduced to equation (3), by using the two identity equations and applying them to equations (1) and (2):

$$x_{l+1} = x_l + \mathcal{F}(x_l, \mathcal{W}_l) \quad (10)$$

and then, using definitions of the elements in a building block, the following recursive equation is obtained

$$x_L = x_l + \sum_{i=l}^{L-1} \mathcal{F}(x_i, \mathcal{W}_i) \quad (11)$$

that associates any deeper unit L with any shallower unit l . The summation term $\sum_{i=l}^{L-1} \mathcal{F}(x_i, \mathcal{W}_i)$ is the residual between units L and l . The innovative property of the ResNet is this residual function, which distinguishes it from previous designs of so-called plain networks, where a feature x_L is a series of matrix-vector products. The simplification used in the defining equations of a ResNet not only introduces the skip connections, but also results in the consequence that the gradient of a layer does not vanish, even when the weights are arbitrarily small. The signal can be propagated both forward and backward between layers L and l . This paper also states explicitly that deeper (plain) networks suffer from increased errors.

He and others (He et al., 2016c) compare $n = \{3, 5, 7, 9\}$, leading to 20, 32, 44, and 56-layer networks, and $n = 18$ that leads to a 110-layer ResNet. (He et al., 2016c) conclude that ResNets (including ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, whose main difference lies in the number of network layers) perform better in image classification aimed at feature extraction than other CNN models, evaluated for the ImageNet dataset. In our work, we will investigate the potential of ResNets for identification and classification of complex cryospheric spatial patterns, namely crevasse patterns.

8.1.2 Properties of ResNet-18

A few facts about ResNet-18 are the following:

(a) *Layer structure.* ResNet-18 consists of 16 convolutional layers, 2 downsampling layers and several fully connected layers. Convolution kernels are of size 7x7 for the first convolution layer and of size 3x3 for the following convolutional layers. A ResNet-18 can include shortcut connections.

(b) *Size of input images.* The input images need to have a size of 224 by 224 pixels, as a result of the requirement that the number of input neurons of the fully connected layer is fixed. This is a very limiting fact, because it cannot be assumed that patterns or objects can be identified in

a (sub)image of a specific size. After identifying optimal sizes for input images in the crevasse detection, the creation of labeled data sets was re-run, using images of 224x224 size, in order to allow training of a ResNet-18 model. The connectionist-geostatistical method and VarioMLP do not require a certain size of input imagery. Vario-function calculation is computationally most efficient when images of a size of 3-4-5 are used (Herzfeld and Zahner, 2001), however a rectangular input image of any size can be utilized. This type of flexibility is important, because it does not require a-priori assumptions about the relationship between sensor resolution, data fields and physical sizes of cryospheric or other patterns.

(c) *Class association* is carried out by an eigenvector composed of multiple probabilities, the class with the highest probability will be associated to the image to be classified.

(d) *Performance*. There is research that indicates (Ou et al., 2019) that ResNet-18 can actually not be expected to perform a complex class association, such as the crevasse classification, but only extraction of low-level features such as edge detection and texture. Ou et al. (2019) state that deeper ResNets (with more layers) would be needed for detection or classification of more complex features, such as spatial context, global semantic and local features of objects. As demonstrated in the work in this paper, the ResNet-18 model that was derived from a combination with the VarioMLP has resulted in classification of crevasse patterns. Of note, in classic satellite image processing and other image processing, edge-detection and texture analysis are obtained by convolution with small kernel images of sizes 3x3 to 7x7 (Herzfeld and Zahner, 2001), so the finding of limitations caused by small kernels in ResNets are an analogue of limitations in image processing known several decades ago.

(e) *Activation functions*: Pandey and Srivastava (2023) explore the effect of different activation functions on image classification results. They note that CNNs perform better than machine learning techniques because of their multi-layer hierarchical feature extraction which is controlled by variables such as number of hidden layers, activation functions, learning rates, initial weights, and decay functions, however, they attribute non-linearity of the network only to the activation function which motivates their comparative investigation, regarding under-researched problems including (a) vanishing and exploding gradients during back-propagation, (b) zero-mean and range of outputs, (c) compute complexity of activation functions and (d) predictive performance of the model. The activation function used in our ResNet-18 experiments is the Rectified Linear Unit (ReLU), which is commonly used elsewhere in the literature.

8.1.3 Applications of ResNets

Examples of studies that utilize ResNet-18 include applications to moving object detection in super-resolution videos (He et al., 2021) or complex scenes (Ou et al., 2019), an analysis of COVID presence (Kochgaven et al., 2021), an application to Alzheimer’s diagnosis (Ramzan et al., 2020), an engineering application to classification images of bearing faults (Jiang et al., 2023). In medical sciences, labeled training data exist (Kochgaven et al., 2021), other than in geosciences. In order

to process video satellite image sequences of relatively low resolution, collected by Chinese video surveillance satellites Jilin-1 and OVS-1, into data streams that have super-high spatial resolution while maintaining the high temporal resolution of the video data allowing to detect moving objects, He et al. (2021) develop a multiframe video super-resolution neural network (MVSRnet). The resultant MVSRnet is a ResNet with a 1001-layer depth, the largest depth published to date and first mentioned in He et al. (2016c), and it includes an attention mechanism to improve moving feature detection. While the complexity of the MVSRnet is impressive, it performs a relatively simple task, the detection of moving objects. In contrast, our research aims at extraction of complex information. In the next subsection, we cast some light on the expected and reported differences in capabilities and performance of comparatively shallow (such as ResNet-18) and deeper CNNs.

8.2 Selection of ResNet-18 as the CNN Structure for the Work in this Paper

ResNet-18 was chosen as the convolutional architecture. The ResNet architecture was first developed in 2015 in response to a growing set of problems with CNN image classification architectures at the time (He et al., 2016a,c). While deep convolutional architectures had been shown to provide state of the art performance on standard image classification benchmarks such as the ImageNet dataset (Deng et al., 2009; Krizhevsky et al., 2012; He et al., 2016a), it was quickly discovered that, in a somewhat counter-intuitive fashion, deeper networks only led to increased performance up to a point, after which increased network depth resulted in increasingly worse performance. This was due to what is known as the vanishing gradients problem (Pandey and Srivastava, 2023; Ruder, 2016). In essence, the deeper a network becomes, the smaller the derivative used to adjust model weights becomes during backpropagation. After a certain point during backpropagation, this value becomes so insignificantly small that the initial layers of the network are no longer trained at all. This results in significantly worse performance, and overall training inefficiency at scale. ResNet solves this problem by allowing skip or shortcut connections between convolutional layers, in which the output from a given layer is both fed directly to the next layer, and several layers later in the model. This effectively solves the vanishing gradient problem, as it provides a much shorter path for the gradient to adjust the initial layers of the model during backpropagation. Not only do ResNet models achieve similar or better performance to other state-of-the-art convolutional image classification models of its time such as VGG (Simonyan and Zisserman, 2014) or AlexNet (Deng et al., 2009; He et al., 2016a), they do so with far fewer trainable parameters. ResNet-18, the most shallow commonly used variation containing 18 layers has roughly 11 million trainable parameters, whereas VGG16 (a state of the art convolutional model contemporary to ResNet) contains around 138 million. This drastically reduced model complexity with comparable performance results in faster, more efficient and more generalizable training for ResNet models as opposed to traditional convolutional architectures (He et al., 2016c). Furthermore, deeper ResNet architectures have been shown to result strictly in increased performance models with even hundreds of ResNet layers perform better on benchmark datasets than shallower versions of the same architecture.

For this project, ResNet-18 was chosen because of its overall efficiency and demonstrably high

performance on similar image classification tasks. Because the problem domain requires building a custom dataset from scratch, deeper and wider convolutional network architectures which mainly show increased performance over ResNet-18 on benchmark datasets with millions of training examples are not necessary. In summary, ResNet-18 provides an efficient and lightweight reference for convolutional image classification architectures. The reference ResNet-18 architecture was left essentially unmodified for the purposes of this paper, in order to serve as a reasonable benchmark for convolutional models on the glacier surface type classification problem, and for ease of comparison with the VarioMLP model.

8.3 Determination of ResNet-18 Hyperparameters

Because the main goal of this paper is to evaluate a CNN compared to a physically constrained NN, and the ResNet-18 model was selected as explained above (section 8.2), the commonly used hyperparameters of the ResNet-18 model were not reevaluated. The purpose of the ResNet CNN architecture in this project is to provide a benchmark for CNN models on the surface classification problem to compare against the VarioMLP. The goal is to provide a methodology which inherits the strengths of both the “shallow” fully-connected VarioMLP architecture with the deep ResNet model.

For ResNet-18, the one parameter tested is the batch size (Table 7). Batch size refers to the number of input examples which are fed forward through the MLP before backpropagation is performed during training. The resulting loss and gradient for backpropagation is then an average of the losses from each input in the “batch”. With a larger batch size, the model sees a greater variety of examples for which to tune its weights and biases, and this can improve the model’s generalizability. Too large of a batch size relative to the amount of training data can result in a poorly directed gradient during backpropagation. For the Resnet-18 model, batch sizes of 1, 2, 4 and 8 were tested, with 2 achieving the lowest loss of any tested model or training hyperparameter set (Table 7). It is interesting to note that after a batch size of 2, performance degrades significantly. This is likely a factor of the size of the training dataset. The larger the ratio of batch size to overall training set size, the less directed the gradient will be during backpropagation.

9 Comparison and Integration of VarioMLP and ResNet-18 in a Combined NN Model: VarioCNN

9.1 Comparison of Capabilities and Performance of VarioMLP and ResNet-18

The primary advantage of VarioMLP over the CNN (ResNet-18) is that VarioMLP can be trained with a relatively small set of labeled training data, of a number of input images that can feasibly be labeled by an expert in the field. This is simply not possible for a CNN, not even a relatively shallow CNN such as ResNet-18, because the number of training images ranges in the 100,000s (and is orders of magnitudes higher for other types of CNNs, as reported in section (2)). Data sets of several hundred initially labeled images of crevasse types suffice to train VarioMLP, and thereafter an efficient training feed-back loop, developed here, can be applied to increase the size of the training data set, until a number is achieved that facilitates training of ResNet-18.

In comparisons of the efficiency and performance of VarioMLP and ResNet-18, the VarioMLP outperformed the CNN for training data sets for up to several 1000 (3500 or 5000) labeled images. An explanation for this lies in the fact that we utilize a physically informed spatial analysis/geostatistical approach to NN neuron activation is more important than the NN model architecture. In other words, the performance-determining factor is not the selection of a shallow, potentially outdated model structure, here, the perceptron, but rather the pre-training effect of the geostatistical analysis, combined with expert understanding of the relationship between surface signatures of crevasses on the ice surface and resultant outputs of experimental vario-functions. The MLP turns out to remain an efficient model architecture for this first classification task.

For classification tasks of many more images, ResNet-18 is faster and trained more efficiently. For this study, we have only carried out a limited number of experiments with CNNs.

9.2 Derivation and Application of a Three-Tiered VarioCNN

The main idea is to use the physically constrained Vario-MLP to drive the CNN. Thereby we employ the main advantage of the VarioMLP, to derive labeled data sets of 4000 (3933) training images, starting with approximately 300 labeled images and using the iterative active learning approach.

The combined architecture, illustrated in Figure 8, consists of three tiers: (1) the vario-function calculation for each input image data set, which is used to activate the nodes in the VarioMLP, using the connectionist geostatistical method, (2) the MLP component of Vario MLP, which employs error backpropagation through the layers as a means for optimization of weights, and (3) ResNet-18, a CNN that can take input from satellite split images of a size of 244 by 244 pixels, which are piped through the convolutional structure and associated with 6 output classes, here, 6 crevasse classes. A retraining loop around VarioMLP serves to grow a training data set from a size of several hundred images, selected by an expert, to a size that sufficiently large to train the CNN (here 4000 images

approximately). The ResNet-18 has a stable and robust outcome for classification. Based on the research in this paper, we can formulate the hypothesis that a CNN can be used as a component in a physically constrained NN. We have tested this hypothesis for ResNet-18.

In the next section, we will use the trained ResNet-18 to perform a classification of a time series of WorldView-1 and WorldView-2 data sets to analyze the evolution of the surge in the NGS, based on insights from formation and expansion of 6 basic crevasse classes (4 classes and two rest classes). Experiments using VarioMLP alone for classification of crevasse types of the Negribreen surge from WorldView imagery and Planet SkySat imagery are described and analysed in [Herzfeld et al. \(2023b\)](#).

10 Experiments with VarioCNN: Application to Classification of Crevasse Types from a Time Series of WorldView Satellite Imagery

Following training of VarioCNN using the 3933 set of labeled training images, a final architecture of VarioCNN was derived. The final, trained VarioCNN was then applied to a time series of 7 WorldView-1 and WorldView-2 data sets (Table 2). From a large catalog of WorldView images, 11 images were found to be suitable w.r.t. cloud cover and area coverage, of those, 7 images were selected to represent the time interval between May 2016 to May 2018. A disadvantage of any analysis that utilizes WorldView imagery is the large delay between the time of data collection and the time when imagery is first made available to the glaciological research community. All useful images are WorldView-1 or WorldView-2 data.

As described in section (3), crevasse types are the results of ice-dynamic processes that occur during the surge. The spatial patterns recorded in the satellite image provide a snap-shot of the local result of the dynamic state of the material ice, which is the kinematic force/ state associated with the deformation that results in the crevasse type.

At the beginning of the classification work for this paper, 22 crevasse classes (including ancillary classes), were created. To facilitate efficient implementation and application of the software, crevasse classes were combined into four larger classes: The current selection of classes ((1) one-directional, (2) multi-directional, (3) shear and (4) shear/chaos) provides relatively simple descriptors of deformation kinematics, but allows to capture the formation of main crevasse provinces, as the following analysis will demonstrate.

The resultant time series of thematic maps of the 6 surface classes, which include four crevasse types, undisturbed surface and a rest class, is shown in Figure 9. A criterion for the consistency and geophysical interpretability of the results is the fact that the areas of each crevasse class consist of one or several simply connected regions, without being post-processed, such as smoothed. The region of crevassed ice expands up-glacier, as time evolves and the surge progresses. Therefore, interpretation of our results from the physically constrained CNN, VarioCNN, is warranted and will be presented in the next section.

11 Geophysical Application: Evolution of the Surge in the NGS

The following geophysical analysis is based on the time series of thematic maps of crevasse classes, derived from WorldView imagery using VarioCNN (Figure 9).

11.1 Evolution of Crevasse Classes in 2016

The first image, collected in 2016-06-20 (Figs. 6(a) and 9(a)), corresponds to the time when the start of a surge was first detected in Sentinel imagery (Strozzi et al., 2017). The area of crevassing at this time coincides with an area of fast movement (Trantow and Herzfeld, 2023; Herzfeld et al., 2023b). At this time, one-directional crevasses form the center of the fast-moving region, which is flanked by shear crevassing on both its northern and southern edge. This classification is correct, assessed by visual interpretation and field observations that (1) acceleration occurs in an along-flow direction near the calving front, and the basic physical notion that (2) shear crevasses form between fast-moving and slow-moving ice. Only five days later, on 2016-06-25, the crevassed area had already expanded in both upglacier and across-flow directions (Fig. 9(b)), as indicated by the increased region of ice classified as crevassed. This rapid change matches existing knowledge that a surge is a catastrophic event, which expands rapidly in the acceleration phase (Raymond, 1987; Lingle et al., 1993; Molnia and Post, 1995; Herzfeld, 1998; Mayer et al., 2002; Molnia, 2008; Molnia and Post, 2010; Herzfeld et al., 2013; Trantow and Herzfeld, 2018; ?). At this point, two new crevasse classes occur, (2) multi-directional and (4) shear/chaos. Multi-directional crevasses form, as the next acceleration wave affects existing one-directional crevasses (so these are multi-generational crevasses) near the calving front. Further progression of deformation results in crevasses that are summatively termed “chaos”, as the ice is too fractured to allow identification of crevasse phases (Fig. 3(d, f)). Notably, the new crevasse types occur first in the oldest regions of crevassed ice. Thus, the analysis of crevasse formation allows reconstruction of the surge evolution (beyond velocity). By 2016-July-08 (Fig. 9(c)), the surge has expanded further upglacier, with its leading edge reaching almost as far upglacier as the Negribreen-Ordonnansbreen junction. Fields of one-directional crevasses directional crevasses are always on the upglacier edge of the surge expansion (see also Figs. 9(d),(e),(f), and (g)).

11.2 Evolution of Crevasse Classes in 2017

In 2017, the surge in the NGS reached maximal velocities of 22m/day in July, which marked the height of the acceleration phase and the most intensive phase of new crevassing. This result was obtained from velocity analysis of Sentinel-1 SAR data and from field observations of the authors (Trantow and Herzfeld, 2023) (see also Fig. 3). Leading up to this, the surge expanded far upglacier, as seen in analysis of WorldView imagery from 2017-04-25 (Figs. 6(d)) and 9(d)) and 2017-05-30 (Figs. 6 (e) and 9(e)). The image collected 2017-04-25 only covers the lower glacier region, but the image from 2017-05-30 shows that the surge has expanded as far as the Negribreen-Akademikarbreen junction, indicated by isolated fields of one-directional crevasses, separated by undisturbed ice. Notably, the marginal area of Negribreen is not part-taking in the surge acceleration at this point in time (summer 2017). In both April and May 2017, a large area of multi-directional crevasses is mapped, that extends upglacier from the calving front to the Negribreen-Ord-breen junction, which is as far upglacier as the crevasse provinces of one-dimensional crevassing extended in 2016. In consequence, we conclude that a new surge wave, or acceleration phase, has affected the area

of the 2016 surge in summer 2017. Based on these results, it is likely that surge speeds decreased in winter 2016 and increased again in summer 2017, reaching the all-time maximum. However, crevasse classification maps provide information on the deformation type at a high resolution, whereas velocity maps yield only a single parameter (ice-flow speed). The classification map from 2017-05-30 shows an expansion of the shear zone (class 3: shear) along the southern margin of Negribreen, upglacier to the Negribreen-Rembeebreen junction, bordering the most extensive region of on-dimensional crevassing. In comparison, the shear zone (class (3)) in the northern margin only extends to about 5 km or 10 km upglacier of the Negribreen-Ordonnansbreen Junction (Fig. 1a). Hence, there is a clear asymmetry in progression of the surge through the NGS, with a far more extensive surge region in the longitudinally southern part of the glacier. In the 2017 crevasse maps, we see two regions of class 4: shear/chaos. The northern region follows the northern edge of class (3) “shear”, with slow-flowing ice of non-surgeing Ordonnansbreen running along the northern edge of the class-(3) province, and thus this region is classified correctly. The new and stronger acceleration in 2017 induced a stronger shear pattern, identified in training images for this class. The strong velocity gradient leads to so-called shear holes near the folded moraines (Fig. 3(c),(e)). However, the region identified as class (4) includes a wider part, which actually coincides with the region of the “retreating bay” (Herzfeld et al., [subm](#)), where the ice has retreated along the previous Negribreen-Ordonnansbreen medial moraine, leaving an area where open water is covered with ice chunks of various sizes, rendering “chaos class” (class 4) surface types. The second area, where class (4) is identified, covers much of the region of multi-directional crevasses (class (3)) in 2016, and field observations and imagery show that this is a region of “chaos”, as described above (Fig. 3(d), (f)). The significant differences between strong shear and chaos crevasse types show the limitations of a classification that is based on insufficient numbers of characteristic crevasse classes. Furthermore, the 2017 classifications indicate that Ordonnansbreen has not been affected by the surge in Negribreen.

11.3 Evolution of Crevasse Classes in 2018

Classification results from two WorldView images from 2018 allows analysis of the surge progression in that summer. The images were collected on 2018-04-29 and 2018-05-26. In 2018, the marginal area of Negribreen is also affected by surge crevassing, and crevassing is overall more pervasive (see, Figs. 9(f) and (g)). Thus, we conclude that in 2018 the surge is expanding into areas of shallower bed topography, which coincide generally with marginal glacier areas. Comparison of the width of the crevassed regions in the maps in Figures 9(f) (2018-04-29) and 9(g) (2018-05-26) suggests that this process of across-glacier expansion happens in summer 2018. However, the longitudinally southern part of the glacier continues to be more crevassed than the northern part, as was already noted for summer 2017. Crevasse patterns may still be evolving, as the map for Figures 9(f) (2018-04-29) shows. The northernmost shear zone that parallels the glacier edge between the Negribreen-Ordonnansbreen junction and the Negribreen-Akademikarbreen junction is misclassified as “other”, but shown on the map for (2018-05-30; Fig. 9(g)) correctly as a shear margin (class 3). Similar to progression of crevasse types observed for 2017 compared to 2016, the new surge wave leads to changes of crevasse type, with type (2) “multi-directional” following areas where

type (1) “one-dimensional” formed in 2017, especially in central Negribreen. The regions of class 4 “shear/chaos” increased, expanding the regions of complex shear as well as the regions of chaos crevasses, compared to regions where these types occurred in 2017, with the problem of lack of separation of “complex shear” and “chaos” persisting.

Of note is the mapping of crevasse types in upper Negribreen, south of the Negribreen-Akademikar-breen junction and north of the Rembeebreen, and downstream of the Filchnerfonna ice falls. The Negribreen-Akademikarbreen medial moraine is classified correctly as class (5) “other” in its upstream (eastern) part, but as class 2 “shear” in its downstream half, which is incorrect, as moraine material covers this part of the glacier (seen in airborne imagery from summer 2018, cf. Figure 1 in (Herzfeld et al., 2022)).

In this area, crevasse fields overflowed by our ICESat-2 validation campaign and analyzed in (Herzfeld et al., 2021, 2022) (RGT450-RGT594 areas) are clearly seen and correctly classified in Fig. 9(g) of 2018-05-26. The one-directional crevasses south of the Negribreen-Akademikarbreen medial moraine are the subtype of thin, parallel, freshly opened crevasses, measured by airborne altimetry and ICESat-2 satellite altimetry (Herzfeld et al., 2021). Furthermore, the classification of uncrevassed regions as class (1), demonstrates that the surge has not transgressed the medial moraines between Negribreen and Akademikarbreen, nor advanced into Rembeebreen in the south. In conclusion, as of summer 2018, the surge in the NGS has not or not yet affected side glaciers of Negribreen. However, it has been hypothesized in (Lefauconnier and Hagen, 1991) that some side glaciers may surge in a later part of the surge process.

11.4 Summary of Geophysical Findings

In summary, we conclude that crevasse classification using a physically constrained neural net yields a segmentation of a surging glacier into crevasse provinces, that allows geophysical interpretation. A time series of crevasse provinces, based on a time series of WV images, provides evidence of the complex deformation processes that occur during the evolution of the surge. Individual findings are, in summarized form, as follows:

- (1) More classes form, as the surge progresses.
- (2) Fields of one-directional crevasses directional crevasses are always on the upglacier, leading edge of the surge expansion. From airborne observations and numerical analyses, we know that these crevasses are of extensional crevasse type with direction of extension in the direction of largest strain (strain rate) (Herzfeld et al., 2013, 2022; Trantow and Herzfeld, 2023).
- (3) Fields of shear crevasse type form between areas of accelerating and fast-moving ice and areas of slow-moving ice that is not (or not yet) affected by the surge.
- (4) Multi-generational, multi-directional crevasse types form, as a new wave of the surge forces

affects regions with pre-existing crevasses. Multi-directional crevasses can also form as a result of a two-directional, expansive force field, as observed during the surge of the BBGS (Herzfeld, 1998), however, these types are not differentiated in the experiments in this paper.

(5) Lastly, continued deformation can render the crevassed area as a region “chaos class”, where individual deformation events cannot be traced back from the crevasse patterns any more.

(6) Combining complex shear and chaos into a single class limits the ability for geophysical interpretation. For simplicity, these two different processes are not differentiated in the experiments in this paper. In a paper in preparation, we discriminate up to 13 crevasse classes (Herzfeld et al., 2023b).

(7) Over time, the surge expands into marginal areas, in addition to expanding upglacier.

12 Summary, Discussion and Conclusions

The work in this paper has addressed three challenges, posed in the introduction: Challenge 1. Harnessing the data revolution in Earth observation from space; Challenge 2. Glacial acceleration and Sea-Level-Rise assessment; and Challenge 3. Integration of physically-constrained classification and modern “Deep Learning” approaches in satellite image classification.

Challenge 1. Harnessing the data revolution in Earth observation from space. Through the integration of physical knowledge and two different ML approaches into a physically-driven NN, the VarioCNN, we have provided a means for rapid and efficient extraction of complex information from submeter resolution satellite imagery (and other imagery). The new NN, VarioCNN, combines the advantages of a physically-driven, relatively easily trainable MLP, with those of an efficient CNN, and thus directly provides an answer to **Challenge 3. Integration of physically-constrained classification and modern “Deep Learning” approaches in satellite image classification.**

There are several key concepts which have been instrumental in the mathematical and computational formulation of a connection between physical understanding and physically constrained classification: (1) ice dynamics of glacial acceleration, especially surging, (2) deformation of the material ice during rapid acceleration, (3) resultant surface signatures: crevasse patterns, and their formation, transport and overprinting, (4) recording of ice-surface structures in optical satellite imagery (and other imagery), and (5) mathematical representation of crevasse patterns in multi-directional vario functions – all these components comprise the physical constraints of the VarioMLP. VarioMLP utilizes the connectionist-geostatistical classification method (Herzfeld and Zahner, 2001; Herzfeld, 2008; Herzfeld et al., 2013) to first process satellite imagery by calculation of directional vario functions, which are then used to activate the neurons of an input layer of a MLP.

While there has been an increasing acceptance of deep learning methods in the geosciences, the lack of adequate, problem-specific labeled training data has hampered derivation of new knowledge using said deep learning approaches, because CNNs require training data sets with on the order of 100,000s to millions of labeled data. Science applications of CNNs have been limited to areas where more training data exist, including (a) biology and medicine, (b) atmospheric sciences and weather forecasting, and (c) sea surface temperature (ocean remote sensing) (Reichstein et al., 2019).

In a comparison of VarioMLP and ResNet-18, the shallowest “deep” NN that is commonly used (He et al., 2016a,c), we find that the primary advantage of VarioMLP over the CNN is that VarioMLP can be trained with a relatively small set of labeled training data, of a number of input images that can feasibly be labeled by an expert in the field. Starting from a set of several hundred training data sets of crevassed surface images, associated to six classes by a structural glaciologist, a feedback loop of retraining and reinforcement, with a fast rejection/acceptance feature supported by a GUI in a combination of a confidence measure and expert-controlled decision, leads to creation of a labeled crevasse class data set of 4000 images.

We proceed to create a combined three tiered network, termed VarioCNN, which consists of VarioMLP, the feedback loop and a backend of a CNN (ResNet-18); this NN can be trained with the 4000-image labeled data set and has better training properties than VarioMLP alone. A flexible and versatile open-source software system, GEOCLASS-image (Herzfeld et al., 2023a), was designed and built for image classification. It performs all the tasks in this analysis and more; it is easily generalizable to other network structures and applications because of its modular design. GEOCLASS-image is user friendly, it includes a functional GUI that appeals to the expert and non-expert in glaciology or computer science alike (i.e. it does not require a lot of knowledge of ML, however, it has a pytorch framework). With GEOCLASS-image and VarioCNN, we have created an infrastructure that facilitates rapid analysis of submeter resolution commercial satellite image data, such as Maxar WorldView data, thus answering to Challenge 1. Furthermore, the work in this paper presents an approach for a path forward in harnessing the data revolution towards obtaining an advanced understanding of complex geophysical phenomena (here: glacial acceleration) in a climate-change science framework.

Challenge 2. Glacial acceleration and Sea-Level-Rise assessment. Our research in this paper presents an advance in the complexity of physics that can be extracted from satellite imagery (crevasse classification, deformation), in an area where such research has not been conducted yet. In the introduction, we have summarized the relationship between glacial acceleration and sea-level rise. In summary, glacial acceleration constitutes a deep uncertainty in SLR assessment, a term coined by the 6th Assessment Report of the IPCC (Masson-Delmotte et al., 2021). Surges are the least understood form of glacial acceleration. The work in this paper culminates in an application of VarioCNN to study the evolution of crevasse provinces during the current (2016-2024) surge of an Arctic glacier system, the Negribreen Glacier system, Svalbard, based on classification of crevasse types in a time series of WorldView images for 2016-2018. This constitutes a novel approach, resulting in new results in glaciology. The classification is the first of its kind, carried out for an entire Arctic glacier system and for WorldView data. Negribreen last surged in 1935/36 (Lefauconnier and Hagen, 1991; Herzfeld et al., 2021, 2023b).

Using four principle crevasse types (one-directional, multi-directional, shear and chaos), plus a class for undisturbed snow/ice surfaces and a rest class, we have derived segmentations of a surging glacier into crevasse provinces that allow geophysical interpretation of the surge evolution in 2016-2018, which includes most of the acceleration phase of the surge. Some results are: More crevasses form, as the surge expands. Fields of one-directional crevasses always form on the upglacier, leading edge of the surge expansion. Fields of shear crevasse type form between areas of accelerating and fast-moving ice and areas of slow-moving ice that is not, or not yet, affected by the surge. Multi-generational, multi-directional crevasse types form, as a new wave of the surge forces affects regions with pre-existing crevasses. Lastly, continued deformation can render the crevassed area as a region of “chaos class”, where individual deformation events cannot be traced back to individual deformation events any more. Over time, the surge expands upglacier and into marginal areas. Links to modeling are outlined.

A limitation of the current analysis is the small number of crevasse classes, chosen to more easily

derive the first combined network that integrates the connectionist-geostatistical approach and a CNN. A classification that distinguishes up to 13 crevasse classes is in preparation (Herzfeld et al., 2023b). ResNet-18 requires square input images of 244 by 244 pixels, which VarioMLP does not require, however, this limitation is not severe, as split-images of any size can be created using the Split Image Explorer Tool in GEOCLASS-image (Herzfeld et al., 2023a) and capture the crevasse patterns.

More generally, the specific glaciological results obtained in this paper demonstrate that geoscience and computer science are equally important disciplines in the development of physically constrained NNs (i.e. glaciology is not merely “domain knowledge”), in light of the goal to utilize modern observation technology to advance geophysical understanding.

APPENDIX A: The GEOCLASS-image Software System

In this section, we describe the GEOCLASS-image Software System, as publicly shared under [Herzfeld et al. \(2023a\)](#), its functionalities, install and application. All components of the work in this paper were achieved by application of the GEOCLASS-image system.

A.1 The GEOCLASS Software System: Functionalities and Software Description

The GEOCLASS-image software includes subsystems that facilitate the following functionalities (see, Table 3). (A) Utility Scripts, (B) Geostatistical Methods: Functions, (C) Geostatistical Methods: Feature Vectors, (D) Machine Learning Models, (E) Machine Learning Training and Validation, and (F) Dataset Management.

Building Matlab and C classification software (lib-geoclass) developed in the group of the authors over the years, the GEOCLASS-image system includes features from the world of Geographical Information Systems (GISs), summarized under (A), (E) and (F). GEOCLASS-image has strengths in satellite image data input and handling, including different types of projections.

Types of satellite data that can be used include WorldView-1 and WorldView-2 data. The utility functions are implemented in a way that allows generalization to other types of satellite data relatively easily. The GUI has several features and capabilities which make it especially user friendly. Data handling, image projections and the selection of the split images as well as training features are integrated seamlessly and have been tested extensively (see, section (F) in Table 3). Dataset management includes a split-image labeling tool, which allows to select individual images or utilize a batch loader. The GUI and visualizations adapt dynamically to the creation of new classes. Multiple satellite images can be analyzed at the same time. To facilitate creation of labeled training data, classifications and labels can be output in NetCDF format.

The NN functionalities are embedded in the data management and utility-function0enabled environment. All functionalities can be performed within the GUI or using data handling scripts to meet different user preferences and working styles, as well as allow for efficient handling of large data sets. Because the lack of existence of labeled training sets for the cryospheric science problems and geoscience problems in general, particular attention was given to the Split Image Explorer Tool (included in modules under (F)) and its interaction with modules under (E) Machine Learning Training and Validation. Classifications can be visually assessed, which allows direct comparison with external information, such as visually analyzed satellite imagery or any other mapped data (field or airborne observations), or computationally evaluated using threshold values and confidence measures.

GEOCLASS-image has an open structure with respect to ML models (D) and geostatistical functions (B) and feature vectors (F). In the current version, the multi-layer perceptron and the model

are implemented and tested. Geostatistical functions include first- and second order vario functions and residual vario functions, with room for other functions. The analysis in this paper employs the values of the experimental vario function directly and this workflow has been realized in the computational infrastructure as well. The classification software lib.geoclass offers a number of additional functions, feature vectors and parameters that will be included in a future version of GEOCLASS-image.

A.2 Install and Use of the GEOCLASS-image Software System

In this section, we give a brief description of the components of the GEOCLASS-image software used for classifying crevasse provinces in satellite imagery. GEOCLASS-image is publicly available through GitHub in its first release version (v1, [Herzfeld et al. \(2023a\)](#)). The GitHub repository contains guides for installing and running the software, along with a more comprehensive documentation of its features and functionality. GEOCLASS-image is written mostly in python3 and is meant to be run on Unix-based systems, such as Mac OS and Ubuntu.

A.2.1 Dataset Generation

After installation and configuration, the user begins by creating a dataset of split imagery to be classified. First, one creates a contour outline of their region of interest, which can be imported or generated using the software. At present, the dataset generation script expects a GeoTiff-formatted image as input, which is a standard product for WorldView data and other satellite image products such as those from Landsat and Sentinel missions. The GeoTiff format allows geolocation of each pixel within the image and contains projection information that is utilized by the software. Given the input GeoTiff image, the software generates split images of ice regions within the contour which are to be classified.

A.2.2 Labeling Training Data using the Split Image Explorer Tool

A main feature of GEOCLASS-image is the graphical user interface (GUI) created to label, visualize and ultimately classify crevasses in an intuitive and visually appealing way. This interface is called the Split Image Explorer Tool, but here we refer to it simply as the GUI. After generating the split image data set from the desired GeoTiffs, the GUI is launched in order to start labeling split-images for training. This step is done only in the case when training data is not available and the user wishes to label their own crevasses classes for training a new neural network. The GUI displays a GeoTiff image with the contour plotted on top, and multiple full images can be loaded in at one time with easy switching between them. The user can then select any split image within the contour and visualize it at full resolution. Split images are then labeled based on the classes the user set up during the configuration step. The progress of the labeling process can be visualized at

anytime by selecting the “Visualize Labels” button within the GUI.

Many times, split images that are near each other in the source GeoTIFF image will be classified similarly. In this case, one can use the “batch labeling” option within the GUI to label a group of split images together at one time.

A.2.3 Training

After labelling training data with the GUI tool, the next step is to train the classification model, which is done by executing the training script via the terminal. The model parameters, specified in the configuration file, define the hyperparameters of the classification model. At present, GEOCLASS-image has the option to use either the VarioMLP model or the ResNet-18 model. With the structure well-defined (He et al., 2016b), only the VarioMLP requires specification of structure-based model parameters for the neural network. These include the number of lag values to be used in the directional variogram and the shape of the hidden layers of the VarioMLP network. Currently, only the Rectified Linear Unit (ReLU) can be chosen for an activation function for either model.

Training-based parameters are also defined in the configuration file. These include an option to utilize the power of GPU’s in processing via CUDA (Luebke, 2008), the maximum number of epochs to run the training loop, the initial learning rate for the optimizer, and the number of split images to be passed through network before each iteration of the backpropagation. At present, only the Adam optimization algorithm is available for training (Kingma and Ba, 2014).

Data augmentation options are also available in order to make classification more robust (Shorten and Khoshgoftaar, 2019). For example, the user can set data augmentation parameters in the configuration file that randomly rotate and shift variograms before being fed into a network.

Once the training script is executed, it displays the relevant training parameters in the terminal and provides estimates for training loss and validation loss for each epoch throughout the training process. Upon completion, loss data are stored in an output folder. The software also saves model checkpoints throughout the training process. By default, the training script saves a model checkpoint at the end of each epoch if and only if the validation loss is better than the previous best validation loss. The training script supports the ability to load and resume training from a previously defined model checkpoint.

A.2.4 Testing

After looking at the losses from the training run and selecting a desired model checkpoint, the testing script can be used to label all split images in the dataset using the trained classification

model. Upon completion, a file containing the classification labels for all split images in the dataset is generated which can be loaded into the GUI and visualized. Confidence measures are also provided and can be used as a filter within the GUI. Figure 5 shows the visualization of a simple crevasse classification result within the GUI.

A.3 Generalizations

While the above description largely lines up with the application described in this paper, the GEOCLASS-image system includes several utility functions that facilitate application to other types of satellite data sets, other geostatistical functions, different types of using the classification approach (Herzfeld, 2008), and a suite of ML training and validation tools, as well as dataset management tools (see, Table 3). Different combinations of these tools and features allow a user-specific development of ML architectures with geostatistical preprocessing modules.

Software/ Open Science Section

The data and code supporting the findings of this study are openly available in the GEOCLASS-image GitHub repository (Herzfeld et al., 2023a) found at

<https://github.com/Herzfeld-Lab/GEOCLASS-image/releases/tag/v1.0>.

The repository contains all the necessary datasets, scripts, and code used for analysis and reproduction of the results presented in the article. Researchers interested in accessing and utilizing the first release of this software may find them at the specified GitHub link.

Additionally, specific instructions, descriptions, and necessary dependencies to replicate and conduct similar experiments and analyses are documented within the repository's README file.

All materials in this repository are released under the MIT License. Please refer to the repository's LICENSE file for detailed information on the permissions and restrictions regarding the use, reproduction, and distribution of the data and code.

Author Contributions

UCH designed the research in this paper, developed the original Connectionist-Geostatistical Classification method and wrote most of this paper. The computational part of this paper is significantly based on unpublished work for a Masters Thesis, written by Lawrence “Jack” Hessburg, under advisory of U. Herzfeld. JH wrote the python geoclass software package, which has been utilized for all computational work in this paper, based on legacy software developed by UCH and members of the geomathematics laboratory, in various programming languages, including the GEOCLASS Library in C and GEOCLASS matlab software. TT contributed extensively to writing this paper. AH tested and used the new python GEOCLASS-image software and contributed to writing.

Acknowledgements

Thanks are due to Tasha Markley, Griffin Hayes, Lukas Goetz-Weiss, Alex Weltman, Alfredo de La Pena Gonzales, Connor Meyers and Chris Higginson, all Geomathematics Lab, University of Colorado Boulder, and to Oliver Zahner for previous work on the classification methods and the GEOCLASS software. Maxar WorldView satellite imagery from the surge of the NGS was acquired with help from the Polar Geospatial Center, University of Minnesota, here we are indebted to Paul Morin, Jonathan Pundsack, Cole Kelleher, Stephanie Linde and colleagues, and through the NASA Commercial Small Satellite Program (CSDAP).

The work in this paper was primarily funded by U.S. National Science Foundation (NSF) Office of Advanced Cyberinfrastructure award OAC-1835256. Research on the surge in the NGS and data collection were supported by the U.S. National Aeronautics and Space Administration (NASA) Earth Sciences Division under awards 80NSSC20K0975, 80NSSC18K1439 and NNX17AG75G and by the U.S. National Science Foundation (NSF) under awards OPP-1745705 and OPP-1942356 (Office of Polar Programs). Research on WorldView data analysis was also supported by NASA Earth Sciences under the CSDAP. Principal Investigator for all awards is Ute Herzfeld. Helicopter support was facilitated by the Norwegian Polar Center. Collection of airborne data in Svalbard was conducted with permission of the National Security Authority of Norway, the Civil Aviation Authority of Norway and the Governor of Svalbard, registered as Research in Svalbard Project RIS-10827 “NEGRIBREEN SURGE”. The data collection was also partly supported through a 2018 Access Pilot Project (2017.0010) of the Svalbard Integrated Observing System (SIOS). All this support is gratefully acknowledged.

References

- Aguilar, M., Saldaña, M., and Aguilar, F. (2013). GeoEye-1 and WorldView-2 pan-sharpened imagery for object-based classification in urban environments. *International Journal of Remote Sensing*, 34(7):2583–2606.
- Alley, K. E., Wild, C. T., Luckman, A., Scambos, T. A., Truffer, M., Pettit, E. C., Muto, A., Wallin, B., Klinger, M., Sutterley, T., et al. (2021). Two decades of dynamic change and progressive destabilization on the thwaites eastern ice shelf. *The Cryosphere*, 15(11):5187–5203.
- Bain, A. (1873). *Mind and body: The theories of their relation*, volume 4. D. Appleton.
- Belkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854.
- Benn, D., Fowler, A., Hewitt, I., and Sevestre, H. (2019). A general theory of glacier surges. *Journal of Glaciology*, 65(253):701–716.
- Briscoe, E. and Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118(1):2–16.
- Buckley, E., Farrell, S., Duncan, K., Connor, L., Kuhn, J., and Dominguez, R. (2020). Classification of sea ice summer melt features in high-resolution icebridge imagery. *Journal of Geophysical Research*, 125, DOI: [10.1029/2019JC015738](https://doi.org/10.1029/2019JC015738).
- Buckley, E. M., Farrell, S. L., Herzfeld, U. C., Trantow, T. M., Baney, O. N., Duncan, K., Han, H., Lawson, M., and Webster, M. (2023). Observing the evolution of summer melt on multiyear sea ice with ICESat-2 and Sentinel-2. *The Cryosphere*, 17(9):3695–3719, DOI: [10.5194/tc-17-3695-2023](https://doi.org/10.5194/tc-17-3695-2023). Published 2023/08/31.
- Camps-Valls, G., Svendsen, D. H., Cortés-Andrés, J., Mareno-Martínez, Á., Pérez-Suay, A., Adsuaara, J., Martín, I., Piles, M., Muñoz-Marí, J., and Martino, L. (2021a). Physics-aware machine learning for geosciences and remote sensing. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 2086–2089. IEEE.
- Camps-Valls, G., Tuia, D., Zhu, X. X., and Reichstein, M. (2021b). *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences*. John Wiley & Sons.
- Caraballo-Vega, J., Carroll, M., Neigh, C., Wooten, M., Lee, B., Weis, A., Aronne, M., Alemu, W., and Williams, Z. (2023). Optimizing worldview-2,-3 cloud masking using machine learning approaches. *Remote Sensing of Environment*, 284:113332.
- Clarke, G. (1987). Fast glacier flow: Ice streams, surging, and tidewater glaciers. *Journal of Geophysical Research*, 92:8835–8842.

- Collins, M. J. (1992). Information fusion in sea ice remote sensing. *Microwave remote sensing of sea ice*, pages 431–441.
- Dabboor, M. and Geldsetzer, T. (2014). Towards sea ice classification using simulated RADARSAT Constellation Mission compact polarimetric SAR imagery. *Remote Sensing of Environment*, 140:189–195.
- Daw, A., Karpatne, A., Watkins, W. D., Read, J. S., and Kumar, V. (2022). Physics-guided neural networks (pgnn): An application in lake temperature modeling. In *Knowledge Guided Machine Learning*, pages 353–372. Chapman and Hall/CRC.
- De Bézenac, E., Pajot, A., and Gallinari, P. (2019). Deep learning for physical processes: Incorporating prior scientific knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124009.
- de Lima, R. P., Bonar, A., Coronado, D. D., Marfurt, K., and Nicholson, C. (2019). Deep convolutional neural networks as a geological image classification tool. *The Sedimentary Record*, 17(2):4–9.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al. (2012). Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, 120:25–36.
- Earth Observation Portal (EOPortal) (2023a). Worldview-1. <https://www.eoportal.org/satellite-missions/worldview-1>.
- Earth Observation Portal (EOPortal) (2023b). Worldview-2. <https://www.eoportal.org/satellite-missions/worldview-2>.
- Ehrenfeucht, S., Morlighem, M., Rignot, E., Dow, C. F., and Mouginot, J. (2023). Seasonal acceleration of petermann glacier, greenland, from changes in subglacial hydrology. *Geophysical Research Letters*, 50(1):e2022GL098009.
- Elsharkawy, A., Elhabiby, M., and El-Sheimy, N. (2012). Improvement in the detection of land cover classes using the Worldview-2 imagery. In *ASPRS, International Scientific Conference*, pages 19–23.
- Fatland, D. R. and Lingle, C. S. (1998). Analysis of the 1993-95 Bering Glacier (Alaska) surge using differential SAR interferometry. *Journal of Glaciology*, 44(148):532–546.
- Fiuczynski, M. E. (2006). Planetlab: overview, history, and future directions. *ACM SIGOPS Operating Systems Review*, 40(1):6–10.
- Frank, T., Akesson, H., de Fleurian, B., Morlighem, M., and Nisancioglu, K. H. (2022). Geometric controls of tidewater glacier dynamics. *The Cryosphere*, 16(2):581–601.

- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202.
- Gaertner, J., Genovese, V. B., Potter, C., Sewake, K., and Manoukis, N. C. (2017). Vegetation classification of Coffea on Hawaii Island using WorldView-2 satellite imagery. *Journal of Applied Remote Sensing*, 11(4):046005–046005.
- Gaffinet, B., Hagensieker, R., Loi, L., and Schumann, G. (2023). Supervised machine learning for flood extent detection with optical satellite data. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 2084–2087. IEEE.
- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J. (2020). A review on deep learning techniques applied to semantic segmentation. arxiv 2017. *arXiv preprint arXiv:1704.06857*.
- Garrigues, S., Allard, D., and Baret, F. (2007). Using first- and second-order variograms for characterizing landscape spatial structures from remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1823–1834, DOI: [10.1109/TGRS.2007.894572](https://doi.org/10.1109/TGRS.2007.894572).
- Ge, Y., Zhang, X., Atkinson, P. M., Stein, A., and Li, L. (2022). Geoscience-aware deep learning: A new paradigm for remote sensing. *Science of Remote Sensing*, 5:100047.
- Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58.
- Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D. J., Stephens, G. L., van den Heever, S. C., Varble, A. C., and Zuidema, P. (2022). The future of earth system prediction: Advances in model-data fusion. *Science Advances*, 8(14):eabn3488.
- Ghosh, A. and Joshi, P. K. (2014). A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, 26:298–311.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Greve, R. (1995). *Thermomechanisches Verhalten polythermer Eisschilde – Theorie, Analytik, Numerik*. Doctoral thesis, Department of Mechanics, Darmstadt University of Technology, Germany, ISBN: [3-8265-0999-4](https://doi.org/10.1007/978-3-642-03414-5). Berichte aus der Geowissenschaft, Shaker Verlag, Aachen, Germany.
- Greve, R. (2003). *Kontinuumsmechanik*. Springer, Berlin, Germany etc., ISBN: [978354007609](https://doi.org/10.1007/978-3-642-03415-2).
- Greve, R. and Blatter, H. (2009). *Dynamics of Ice Sheets and Glaciers*. Springer, Berlin, Germany etc., ISBN: [978-3-642-03414-5](https://doi.org/10.1007/978-3-642-03414-5), DOI: [10.1007/978-3-642-03415-2](https://doi.org/10.1007/978-3-642-03415-2).
- Grinsted, A., Hvidberg, C. S., Lilien, D. A., Rathmann, N. M., Karlsson, N. B., Gerber, T., Kjær, H. A., Vallelonga, P., and Dahl-Jensen, D. (2022). Accelerating ice flow at the onset of the northeast greenland ice stream. *Nature Communications*, 13(1):5589.

- Haga, O. N., McNabb, R., Nuth, C., Altena, B., Schellenberger, T., and Kääb, A. (2020). From high friction zone to frontal collapse: dynamics of an ongoing tidewater glacier surge, negribreen, svalbard. *Journal of Glaciology*, 66(259):742–754.
- Hanna, E., Pattyn, F., Navarro, F., Favier, V., Goelzer, H., van den Broeke, M. R., Vizcaino, M., Whitehouse, P. L., Ritz, C., Bulthuis, K., and Smith, B. (2020). Mass balance of the ice sheets and glaciers—Progress since AR5 and challenges. *Earth-Science Reviews*, 201:102976.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016b). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016c). Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer.
- He, Z., Li, J., Liu, L., He, D., and Xiao, M. (2021). Multiframe video satellite image super-resolution via attention-based residual learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15.
- Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier.
- Herzfeld, U. (1998). *The 1993-1995 surge of Bering Glacier (Alaska) — a photographic documentation of crevasse patterns and environmental changes*, volume 17 of *Trierer Geograph. Studien*. Geograph. Gesellschaft Trier and Fachbereich VI – Geographie/Geowissenschaften, Universität Trier.
- Herzfeld, U., McDonald, B., Wallin, B., Krabill, W., Manizade, S., Sonntag, J., Mayer, H., Yearsley, W., Chen, P., and Weltman, A. (2014). Elevation changes and dynamic provinces of Jakobshavn Isbræ, Greenland, derived using generalized spatial surface roughness from ICESat GLAS and ATM data. *Journal of Glaciology*, 60(223):834–848, DOI: [10.3189/2014JoG13J129](https://doi.org/10.3189/2014JoG13J129).
- Herzfeld, U., McDonald, B., and Weltman, A. (2013). Bering Glacier and Bagley Ice Valley surge 2011: Crevasse classification as an approach to map deformation stages and surge progression. *Annals of Glaciology*, 54(63):279–286, DOI: [10/3189/2013AoG63A338](https://doi.org/10.3189/2013AoG63A338).
- Herzfeld, U., Trantow, T., and Bennetts, S. (subm). Surge-forced structural disintegration, enhanced calving and resultant rapid mass loss of a large Arctic fjord glacier (Negribreen, Svalbard). *GRL*.
- Herzfeld, U., Trantow, T., Lawson, M., Hans, J., and Medley, G. (2021). Surface heights and crevasse types of surging and fast-moving glaciers from ICESat-2 laser altimeter data — Application of the density-dimension algorithm (DDA-ice) and validation using airborne altimeter and Planet SkySat data. *Science of Remote Sensing*, 3:1–20, DOI: [10.1016/j.srs.2020.100013](https://doi.org/10.1016/j.srs.2020.100013).

- Herzfeld, U. and Wallin, B. (2014). Spatio-temporal analysis of surface elevation changes in Pine Island Glacier, Antarctica, from ICESat GLAS data and ERS-1 radar altimeter data. *Annals of Glaciology*, 55:248–258, DOI: [10.3189/2014AoG66A014](https://doi.org/10.3189/2014AoG66A014).
- Herzfeld, U., Williams, S., Heinrichs, J., Maslanik, J., and Sucht, S. (2016). Geostatistical and statistical classification of sea-ice properties and provinces from SAR data — methods and applications to ice environments near Point Barrow, Alaska. *Remote Sensing, MDPI*, 8(616):1–37, DOI: [10.3390/rs8080616](https://doi.org/10.3390/rs8080616).
- Herzfeld, U. C. (2002). Vario functions of higher order—definition and application to characterization of snow surface roughness. *Computers & Geosciences*, 28(5):641–660.
- Herzfeld, U. C. (2008). Master of the obscure — Automated geostatistical classification in presence of complex geophysical processes. *Mathematical Geosciences*, 40(5):587–618, DOI: [10.1007/s11004-008-9174-4](https://doi.org/10.1007/s11004-008-9174-4).
- Herzfeld, U. C., Clarke, G. K. C., Mayer, H., and Greve, R. (2004). Derivation of deformation characteristics in fast-moving glaciers. *Computers & Geosciences*, 30(3):291–302.
- Herzfeld, U. C., Hessburg, J., Hayes, A., and Trantow, T. (2023a). GEOCLASS-image (v1.0). DOI: [0.5281/zenodo.10232912](https://doi.org/10.5281/zenodo.10232912), <https://github.com/Herzfeld-Lab/GEOCLASS-image/releases/tag/v1.0>.
- Herzfeld, U. C., Lawson, M., Trantow, T., and Nylen, T. (2022). Airborne Validation of ICESat-2 ATLAS Data Over Crevassed Surfaces and Other Complex Glacial Environments: Results From Experiments of Laser Altimeter and Kinematic GPS Data Collection From a Helicopter Over a Surging Arctic Glacier (Negribreen, Svalbard). *Remote Sensing*, 14:1185–1224, DOI: [10.3390/rs14051185](https://doi.org/10.3390/rs14051185).
- Herzfeld, U. C., Markley, T., Hayes, A., de la Pena Gonzalez, A., Hale, G., and Han, H. (prep 2023b). Evolution of the surge in negribreen, svalbard, derived from automated connectionist-geostatistical classification of crevasse provinces in worldview and planet skysat satellite image data.
- Herzfeld, U. C. and Mayer, H. (1997). Surge of Bering Glacier and Bagley Ice Field, Alaska: an update to August 1995 and an interpretation of brittle-deformation patterns. *Journal of Glaciology*, 43(145):427–434.
- Herzfeld, U. C., Stauber, M., and Stahl, N. (2000). Geostatistical characterization of ice surfaces from ERS-1 and ERS-2 SAR data, jakobshavn isbræ, Greenland. *Annals of Glaciology*, 30(1):224–234.
- Herzfeld, U. C., Trantow, T., Hans, J., Lawson, M., and Medley, G. (2020). Surface heights and crevasse types of surging and fast-moving glaciers from ICESat-2 laser altimeter data — Application of the density-dimension algorithm (DDA-ice) and validation using airborne and PlanetLab SkySat data. *Science of Remote Sensing*. in review.

- Herzfeld, U. C. and Zahner, O. (2001). A connectionist-geostatistical approach to automated image classification, applied to the analysis of crevasse patterns in surging ice. *Computers & Geosciences*, 27(5):499–512, DOI: [10.1016/S0098-3004\(00\)00089-3](https://doi.org/10.1016/S0098-3004(00)00089-3).
- Hill, E. A., Carr, J. R., Stokes, C. R., and Gudmundsson, G. H. (2018). Dynamic changes in outlet glaciers in northern greenland from 1948 to 2015. *The Cryosphere*, 12(10):3243–3263.
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116.
- Hu, X., Hu, H., Verma, S., and Zhang, Z.-L. (2020). Physics-guided deep neural networks for power flow analysis. *IEEE Transactions on Power Systems*, 36(3):2082–2092.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Immitzer, M., Atzberger, C., and Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. *Remote sensing*, 4(9):2661–2693.
- Ivakhnenko, A. G. and Lapa, V. G. (1967). *Cybernetics and forecasting techniques*. Modern analytic and computational methods in science and mathematics. North-Holland, New York, NY, <https://cds.cern.ch/record/209675>. Trans. from the Russian, Kiev, Naukova Dumka, 1965.
- James, W. (1890). *The principles of psychology*. H. Holt and Company.
- Jiang, S., Hua, C., and Yuan, M. (2023). Image classification method of bearing fault based on boa optimization resnet-18. In *2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*, pages 510–514. IEEE.
- Jiskoot, H. (2011). Glacier surging. *Encyclopedia of Snow, Ice and Glaciers*, pages 415–428.
- Kampffmeyer, M., Salberg, A.-B., and Jenssen, R. (2016). Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1–9.
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554.
- Karvonen, J. (2006). *Compaction of C-band synthetic aperture radar based sea ice information for navigation in the Baltic Sea*. Helsinki University of Technology.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

- Koc-San, D. (2013). Evaluation of different classification techniques for the detection of glass and plastic greenhouses from WorldView-2 satellite imagery. *Journal of Applied Remote Sensing*, 7(1):073553–073553.
- Kochgaven, C., Mishra, P., and Shitole, S. (2021). Detecting presence of covid-19 with resnet-18 using pytorch. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*, pages 1–6. IEEE.
- Kohonen, T. (1990). Learning vector quantization for pattern recognition. *Technical Report TKK-F-A602*.
- Kohonen, T. (2012). *Self-organization and associative memory*, volume 8. Springer Science & Business Media.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Kwok, R., Cunningham, G., and Holt, B. (1992). An approach to identification of sea ice types from spaceborne sar data. *Microwave Remote Sensing of Sea Ice*, pages 355–360.
- Larour, E., Utke, J., Csatho, B., Schenk, A., Seroussi, H., Morlighem, M., Rignot, E., Schlegel, N., and Khazendar, A. (2014). Inferred basal friction and surface mass balance of the Northeast Greenland Ice Stream using data assimilation of ICESat (Ice Cloud and land Elevation Satellite) surface altimetry and ISSM (Ice Sheet System Model). *The Cryosphere*, 8(6):2335–2335.
- Larour, E. H., Seroussi, H., Morlighem, M., and Rignot, E. (2012). Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM). *J. Geophys. Res.*, 117 (F1):F01022.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lefauconnier, B. and Hagen, J. O. (1991). *Surging and calving glaciers in eastern Svalbard*.
- Li, D., Ke, Y., Gong, H., and Li, X. (2015). Object-based urban tree species classification using bi-temporal worldview-2 and worldview-3 images. *Remote Sensing*, 7(12):16917–16937.
- Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., and Benediktsson, J. A. (2019). Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709.
- Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., and Chen, H. (2020). High-resolution mapping of forest canopy height using machine learning by coupling icesat-2 lidar with sentinel-1, sentinel-2 and landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation*, 92:102163.
- Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.

- Lingle, C., Post, A., Herzfeld, U. C., Molnia, B. F., Krimmel, R., and Roush, J. (1993). Bering Glacier surge and iceberg-calving mechanism at Vitus Lake, Alaska, USA. *Journal of Glaciology*, 39(133):722–727.
- Lipscomb, W. H., Price, S. F., Hoffman, M. J., and A. R. Bennett, G. L., Bradley, S., Evans, K. J., Fyke, J. G., Kennedy, J. H., Perego, M., Ranken, D. M., Sacks, W. J., Salinger, A. G., Vargo, L., and Worley, P. H. (2019). Description and evaluation of the community ice sheet model (cism) v. 2.1. *Geoscientific Model Development*, 12:387–424, DOI: [10.5194/gmd-12-387-2019](https://doi.org/10.5194/gmd-12-387-2019).
- Liu, I. (2002). *Continuum Mechanics*. Springer, Berlin.
- Liu, Q., Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., and Nemani, R. (2020a). Deepsat v2: feature augmented convolutional neural nets for satellite image classification. *Remote Sensing Letters*, 11(2):156–165.
- Liu, X., Hu, Q., Cai, Y., and Cai, Z. (2020b). Extreme learning machine-based ensemble transfer learning for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3892–3902.
- Looney, C. (1997). *Pattern Recognition Using Neural Networks*. Oxford University Press, New York, NY, USA.
- Luan, Q. and Tian, Z. (2022). Application of machine learning methods in geoscience. In *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pages 563–567. IEEE.
- Luebke, D. (2008). Cuda: Scalable parallel programming for high-performance scientific computing. In *2008 5th IEEE international symposium on biomedical imaging: from nano to macro*, pages 836–838. IEEE.
- Malambo, L. and Popescu, S. (2023). Image to image deep learning for enhanced vegetation height modeling in Texas. *Remote Sensing*, 15(22):5391.
- Marmo, B. and Wilson, C. (1998). Strain localisation and incremental deformation within ice masses, Framnes Mountains, east Antarctica. *Journal of Structural Geology*, 20 (2-3):149162.
- Masson-Delmotte, V., Zhai, A., Pirani, S., Connors, C., Péan, S., Berger, N., Caud, Y., Chen, L., Goldfarb, M., Gomis, M., Huang, K., Leitzell, E., Lonnoy, J., Matthews, T., Maycock, T., Waterfield, O., Yelekci, R., Yu, X., and Zhou, B. e. (2021). *AR6: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, DOI: [doi:10.1017/9781009157896](https://doi.org/10.1017/9781009157896).
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8):1246.
- Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Probab.*, 5:439–468.

- Mayer, H. and Herzfeld, U. (2000). Structural glaciology of the fast-moving Jakobshavn Isbræ, Greenland, compared to the surging Bering Glacier, Alaska, USA. *Annals of Glaciology*, 30(1):243–249.
- Mayer, H. and Herzfeld, U. (2001). A structural segmentation, kinematic analysis and dynamic interpretation of Jakobshavns Isbræ, West Greenland. *Zeitschrift für Gletscherkunde und Glazialgeologie*, 37(2):107–124.
- Mayer, H., Herzfeld, U., and Clarke, G. (2002). Analysis of deformation types in fast-moving glaciers. *Terra Nostra*, 4:273–278.
- Means, W. (1976). *Stress and Strain: Basic Concepts of Continuum Mechanics for Geologists*. Springer, New York, NY, USA.
- Melville, B., Lucieer, A., and Aryal, J. (2018). Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *International journal of applied earth observation and geoinformation*, 66:46–55.
- Meyer, H. and Pebesma, E. (2021). Estimating the area of applicability of remote sensing-based machine learning models with limited training data. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 2028–2030. IEEE.
- Minsky, M. and Papert, S. (1969). An introduction to computational geometry. *Cambridge tiass., HIT*, 479(480):104.
- Molnia, B. and Post, A. (1995). Holocene history of Bering Glacier, Alaska: a prelude to the 1993-1994 surge. *Physical Geography*, 16(2):87–117.
- Molnia, B. F. (2008). *Alaska*. Satellite Image Atlas of Glaciers of the World. U.S. Geological Survey Professional Paper 1386-K, Washington, D.C.
- Molnia, B. F. and Post, A. (2010). Surges of the Bering Glacier. *Geological Society of America Special Paper*, (462):251–270.
- Muralidhar, N., Bu, J., Cao, Z., He, L., Ramakrishnan, N., Tafti, D., and Karpatne, A. (2020). Phynet: Physics guided neural networks for particle drag force prediction in assembly. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pages 559–567. SIAM.
- Murray, T., Luckman, A., Strozzi, T., and Nuttall, A. (2003a). The initiation of glacier surging at Fridtjovbreen, Svalbard. *Annals of Glaciology*, 36:110–116.
- Murray, T., Strozzi, T., Luckman, A., Jiskoot, H., and Christakos, P. (2003b). Is there a single surge mechanism? Contrasts in dynamics between glacier surges in Svalbard and other regions. *J. Geophys. Res.*, 108(B5):2237, DOI: [10.1029/2002JB001906](https://doi.org/10.1029/2002JB001906).
- Narine, L. L., Popescu, S. C., and Malambo, L. (2019). Synergy of ICESat-2 and Landsat for mapping forest aboveground biomass with deep learning. *Remote Sensing*, 11(12):1503.

- Neigh, C. S., Masek, J. G., and Nickeson, J. E. (2013). High-resolution satellite data open for government research. *Eos, Transactions American Geophysical Union*, 94(13):121–123.
- Nuth, C., Gilbert, A., Köhler, A., McNabb, R., Schellenberger, T., Sevestre, H., Weidle, C., Girod, L., Luckman, A., and Kääb, A. (2019). Dynamic vulnerability revealed in the collapse of an Arctic tidewater glacier. *Scientific reports*, 9(1):5541.
- Ochilov, S. and Clausi, D. (2012). Operational SAR sea-ice image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(11):4397–4408.
- Ou, X., Yan, P., Zhang, Y., Tu, B., Zhang, G., Wu, J., and Li, W. (2019). Moving object detection method via resnet-18 with encoder–decoder structure in complex scenes. *IEEE Access*, 7:108152–108160.
- Palm, S. P., Selmer, P., Yorks, J., Nicholls, S., and Nowottnick, E. (2021). Planetary boundary layer height estimates from ICESat-2 and CATS backscatter measurements. *Frontiers in Remote Sensing*, 2:716951.
- Pandey, G. K. and Srivastava, S. (2023). Resnet-18 comparative analysis of various activation functions for image classification. In *2023 International Conference on Inventive Computation Technologies (ICICT)*, pages 595–601. IEEE.
- Papadopoulos, G., Edwards, P. J., and Murray, A. F. (2001). Confidence estimation methods for neural networks: A practical comparison. *IEEE transactions on neural networks*, 12(6):1278–1287.
- Payne, A. J., Nowicki, S., Abe-Ouchi, A., Agosta, C., Alexander, P., Albrecht, T., Asay-Davis, X., Aschwanden, A., Barthel, A., Bracegirdle, T. J., Calov, R., Chambers, C., Choi, Y., Cullather, R., Cuzzone, J., Dumas, C., Edwards, T. L., Felikson, D., Fettweis, X., Galton-Fenzi, B. K., Goelzer, H., Gladstone, R., Golledge, N. R., Gregory, J. M., Greve, R., Hattermann, T., Hoffman, M. J., Humbert, A., Huybrechts, P., Jourdain, N. C., Kleiner, T., Munneke, P. K., Larour, E., Lec'h, S. L., Lee, V., Leguy, G., Lipscomb, W. H., Little, C. M., Lowry, D. P., Morlighem, M., Nias, I., Pattyn, F., Pelle, T., Price, S. F., Quiquet, A., Reese, R., Rückamp, M., Schlegel, N.-J., Seroussi, H., Shepherd, A., Simon, E., Slater, D., Smith, R. S., Straneo, F., Sun, S., Tarasov, L., Trusel, L. D., Breedam, J. V., van de Wal, R., van den Broeke, M., Winkelmann, R., Zhao, C., Zhang, T., and Zwinger, T. (2021). Future sea level change under coupled model intercomparison project phase 5 and phase 6 scenarios from the Greenland and Antarctic ice sheets. *Geophysical Research Letters*, 48(16):e2020GL091741.
- Peterson, L., Muir, S., Roscoe, T., and Klingaman, A. (2006). Planetlab architecture: An overview.
- Porter, C., Howat, I., Noh, M.-J., Husby, E., Khuvis, S., Danish, E., Tomko, K., Gardiner, J., Negrete, A., Yadav, B., Klassen, J., Kelleher, C., Cloutier, M., Bakker, J., Enos, J., Arnold, G., Bauer, G., and Morin, P. (2022). ArcticDEM - Mosaics, Version 4.1. DOI: [10.7910/DVN/C98DVS](https://doi.org/10.7910/DVN/C98DVS), <https://doi.org/10.7910/DVN/C98DVS>.
- Qing, D., Huadong, G., Yun, S., Zhen, L., and Changlin, W. (2003). Variograms: practical method to process polarimetric sar data. In *IGARSS 2003. 2003 IEEE International Geoscience and*

- Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, volume 2, pages 932–934 vol.2. DOI: [10.1109/IGARSS.2003.1293968](https://doi.org/10.1109/IGARSS.2003.1293968).
- Ramsay, J. and Lisle, R. (2000). *The Techniques of Modern Structural Geology, Vol. 3: Applications of Continuum Mechanics in Structural Geology*. Academic Press, San Diego, CA.
- Ramzan, F., Khan, M. U. G., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., and Mehmood, Z. (2020). A deep learning approach for automated diagnosis and multi-class classification of alzheimer’s disease stages using resting-state fmri and residual neural networks. *Journal of medical systems*, 44:1–16.
- Rawat, W. and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449.
- Raymond, C. (1987). How do glaciers surge? A review. *Journal of Geophysical Research*, 92(B9):9121–9134.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, f. (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204.
- Rignot, E., Fenty, I., Menemenlis, D., and Xu, Y. (2012). Spreading of warm ocean waters around Greenland as a possible cause for glacier acceleration. *Annals of Glaciology*, 53(60).
- Rist, M., Sammonds, P., Murrell, S., Meredith, P., Doake, C., Oerter, H., and Matsuki, K. (1999). Experimental and theoretical fracture mechanics applied to Antarctic ice feature and surface crevassing. *Journal of Geophysical Research*, 104 (B2):29732987.
- Robel, A. A., Roe, G. H., and Haseloff, M. (2018). Response of marine-terminating glaciers to forcing: time scales, sensitivities, instabilities, and stochastic dynamics. *Journal of Geophysical Research: Earth Surface*, 123(9):2205–2227.
- Roesel, A., Kaleschke, L., and Birnbaum, G. (2012). Melt ponds on arctic sea ice determined from modis satellite data using an artificial neural network. *The Cryosphere*, 6, DOI: [10.5194/tc-6-431-2012](https://doi.org/10.5194/tc-6-431-2012).
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- Rumelhart, D. E., McClelland, J. L., and PDP Research Group, C. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. MIT press.
- Schwartz, D. and Pinel-Puysegur, B. (2018). Spatial and temporal statistical analysis of stack of sar images: The contribution of the variogram. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 4205–4208. DOI: [10.1109/IGARSS.2018.8519396](https://doi.org/10.1109/IGARSS.2018.8519396).

- Sellevold, R., Van Kampenhout, L., Lenaerts, J. T., Noël, B., Lipscomb, W. H., and Vizcaino, M. (2019). Surface mass balance downscaling through elevation classes in an Earth system model: Application to the Greenland ice sheet. *The Cryosphere*, 13(12):3193–3208.
- Sen, A., Suleymanoglu, B., and Soycan, M. (2020). Unsupervised extraction of urban features from airborne lidar data by using self-organizing maps. *Survey Review*, 52(371):150–158.
- Sevestre, H., Benn, D. I., Luckman, A., Nuth, C., Kohler, J., Lindbäck, K., and Pettersson, R. (2018). Tidewater glacier surges initiated at the terminus. *Journal of Geophysical Research: Earth Surface*, 123(5):1035–1051.
- Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell system technical journal*, 30(1):50–64.
- Shen, X.-y., Zhang, J., Meng, J.-m., and Ke, C.-q. (2017). Sea ice type classification based on random forest machine learning with CryoSat-2 altimeter data. In *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*, pages 1–5. IEEE.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.
- Siahaan, A., Smith, R. S., Holland, P. R., Jenkins, A., Gregory, J. M., Lee, V., Mathiot, P., Payne, A. J., Ridley, J. K., and Jones, C. G. (2022). The antarctic contribution to 21st-century sea-level rise predicted by the uk earth system model with an interactive ice sheet. *The Cryosphere*, 16(10):4053–4086.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- Song, J., Gao, S., Zhu, Y., and Ma, C. (2019). A survey of remote sensing image classification based on CNNs. *Big earth data*, 3(3):232–254.
- Steffen, K. and Heinrichs, J. (1994). Feasibility of sea ice typing with synthetic aperture radar (SAR): Merging of Landsat thematic mapper and ERS 1 SAR satellite imagery. *Journal of Geophysical Research: Oceans*, 99(C11):22413–22424.
- Stocker, T. F., Qin, H., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. e. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Straneo, F. and Cenedese, C. (2015). The dynamics of greenland’s glacial fjords and their role in climate. *Annual review of marine science*, 7:89–112.
- Straneo, F., Sutherland, D., Holland, D., Gladish, C., Hamilton, G., Johnson, H., Rignot, E., Xu, Y., and Koppes, M. (2012). Submarine melting of Greenland’s glaciers by Atlantic waters. *Annals Glaciology*, 53(60).

- Strozzi, T., Paul, F., Wiesmann, A., Schellenberger, T., and Kääb, A. (2017). Circum-arctic changes in the flow of glaciers and ice caps from satellite sar data between the 1990s and 2017. *Remote Sensing*, 9(9):947.
- Sun, R. (2014). Connectionism and neural networks. *The Cambridge handbook of artificial intelligence*, pages 108–127.
- Suppe, J. (1985). *Principles of Structural Geology*. Academic Press, Englewood Cliffs, NJ.
- Tai, C., Xiao, T., Zhang, Y., Wang, X., et al. (2015). Convolutional neural networks with low-rank regularization. *arXiv preprint arXiv:1511.06067*.
- Trantow, T. (2020). *Surging in the Bering-Bagley Glacier System, Alaska – Understanding Glacial Acceleration through New Methods in Remote Sensing, Numerical Modeling and Model-Data Comparison*. PhD thesis, University of Colorado, [url=https://www.proquest.com/dissertations-theses/surging-bering-bagley-glacier-system-alaska/docview/2408533962/se-2](https://www.proquest.com/dissertations-theses/surging-bering-bagley-glacier-system-alaska/docview/2408533962/se-2).
- Trantow, T. and Herzfeld, U. C. (2018). Crevasses as indicators of surge dynamics in the Bering Bagley Glacier System, Alaska: Numerical experiments and comparison to image data analysis. *Journal of Geophysical Research: Earth Surface*, 123(8):1615–1637, DOI: [10.1029/2017JF004341](https://doi.org/10.1029/2017JF004341).
- Trantow, T. and Herzfeld, U. C. (2023). Progression of the surge in the Negribreen Glacier System from two years of ICESat-2 measurements. *Journal of Glaciology (in review)*, *Earth ArXiv (preprint: doi.org/10.31223/X5NT1Z)*, DOI: doi.org/10.31223/X5NT1Z.
- Truffer, M. and Echelmeyer, K. A. (2003). Of isbrae and ice streams. *Annals of Glaciology*, 36(1):66–72.
- Twiss, R. and Moore, E. (1992). *Structural Geology*. W.H. Freeman, New York, NY.
- U.S. National Science Foundation (2023). NSF’s Big 10 Ideas: Harnessing the Data Revolution. https://www.nsf.gov/news/special_reports/big_ideas/harnessing.jsp.
- Vigneshwaran, S. and Vasantha Kumar, S. (2021). Comparison of classification methods for urban green space extraction using very high resolution worldview-3 imagery. *Geocarto International*, 36(13):1429–1442.
- Virts, K., Shirey, A., Priftis, G., Ankur, K., Ramasubramanian, M., Muhammad, H., Acharya, A., and Ramachandran, R. (2020). A quantitative analysis on the use of supervised machine learning in earth science. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2252–2255. IEEE.
- Vornberger, P. and Whillans, I. (1990). Crevasse deformation and examples from Ice Stream B, Antarctica. *Journal of Glaciology*, 36 (122):3–10.
- Wagner, W. (2015). Big Data infrastructures for processing Sentinel data. In *Photogrammetric Week*, pages 93–104.

- Wang, L., Scott, K., and Clausi, D. (2016). Improved sea ice concentration estimation through fusing classified SAR imagery and AMSR-E data. *Canadian Journal of Remote Sensing*.
- Waser, L. T., Küchler, M., Jütte, K., and Stampfer, T. (2014). Evaluating the potential of worldview-2 data to classify tree species and different levels of ash mortality. *Remote Sensing*, 6(5):4515–4545.
- Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560.
- Willis, M. J. and von Stosch, M. (2017). Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models. *Computers & Chemical Engineering*, 104:366–376.
- Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., et al. (2019). Current status of landsat program, science, and applications. *Remote sensing of environment*, 225:127–147.
- Xiang, C., Zhang, L., Tang, Y., Zou, W., and Xu, C. (2018). MS-CapsNet: A novel multi-scale capsule network. *IEEE Signal Processing Letters*, 25(12):1850–1854.
- Zakhvatkina, N. Y., Alexandrov, V. Y., Johannessen, O. M., Sandven, S., and Frolov, I. Y. (2013). Classification of sea ice types in ENVISAT synthetic aperture radar images. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(5):2587–2600.
- Zell, A., Mamier, G., Vogt, M., Mache, N., Hubner, R., Doring, S., Herrmann, K., Soye, T., Schmalzl, M., Sommer, T., et al. (1994). Stuttgart neural network simulator user manual. *University of Stuttgart, Stuttgart*.
- Zhang, Z. and Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31.
- Zheng, W., Pritchard, M. E., Willis, M. J., and Stearns, L. A. (2019). The possible transition from glacial surge to ice stream on vavilov ice cap. *Geophysical Research Letters*, 46(23):13892–13902.

Figures and Tables:

Combining “Deep Learning” and Physically Constrained Neural Networks to Derive Complex Glaciological Change Processes from Modern High-Resolution Satellite Imagery: Application of the GEOCLASS-image System to Create VarioCNN for Glacier Surges

Ute C. Herzfeld, Lawrence Hessburg, Thomas Trantow and Adam Hayes

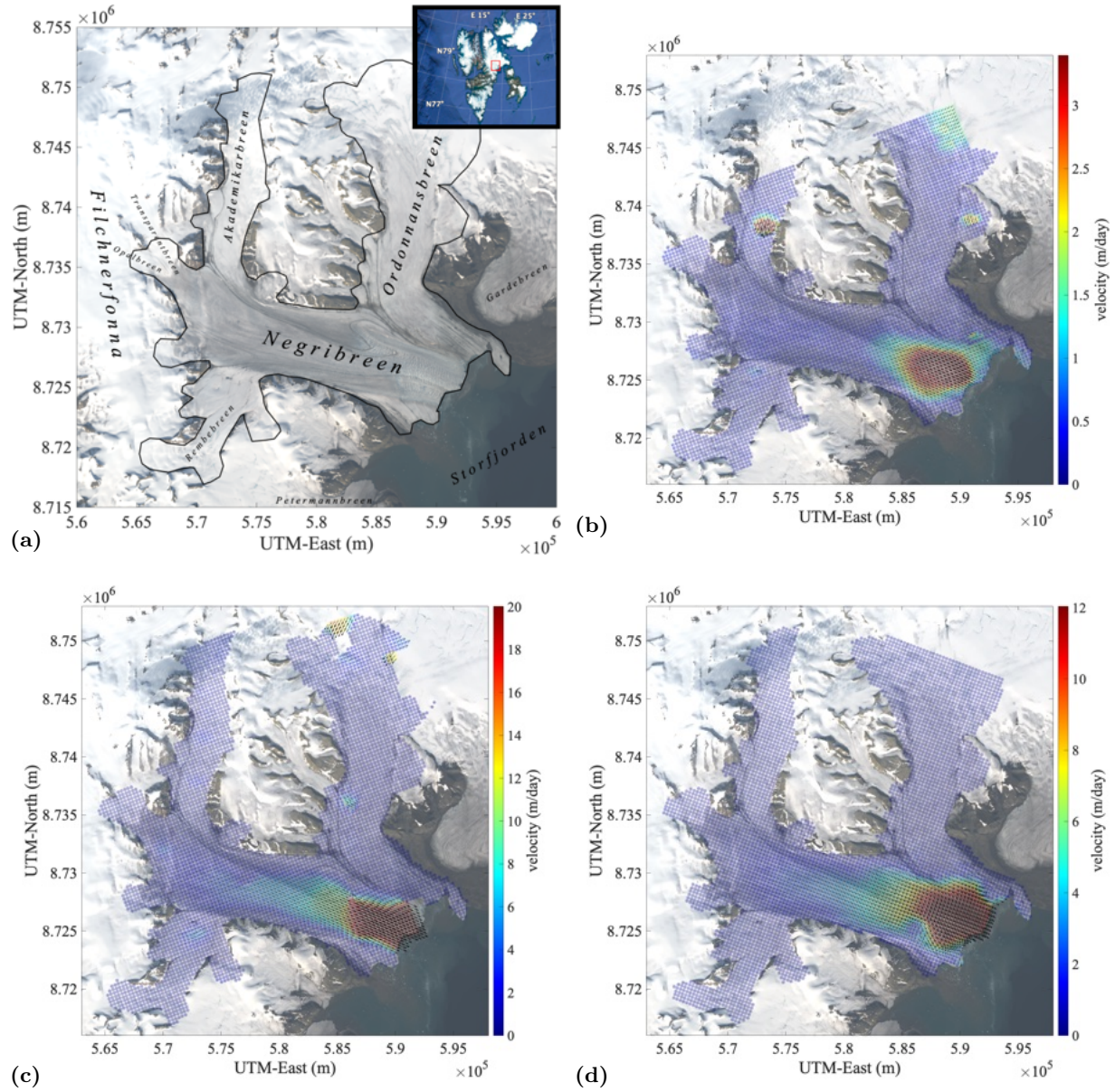


Figure (1). Location of study area and surface velocities from Sentinel-1 SAR data.

(a) The Negribreen Glacier System with important geographical features labeled. The location of the NGS within the Svalbard archipelago is indicated by the red box in the upper-right corner insert. (b) NGS mean surface velocity between 2016-07-03 and 2016-07-15 shortly after the surge began. (c) NGS mean surface velocity between 2017-07-10 and 2017-07-22 when peak surge speeds were reached (upwards of 22 m/day). (d) NGS mean surface velocity between 2018-05-10 and 2018-05-22. Each of the velocity maps in (b)-(d) are in m/day with black arrows indicating the magnitude and direction of mean surface velocity between the baseline dates. Background image for each subfigure: Landsat-8 RGB image acquired 2019-08-05.

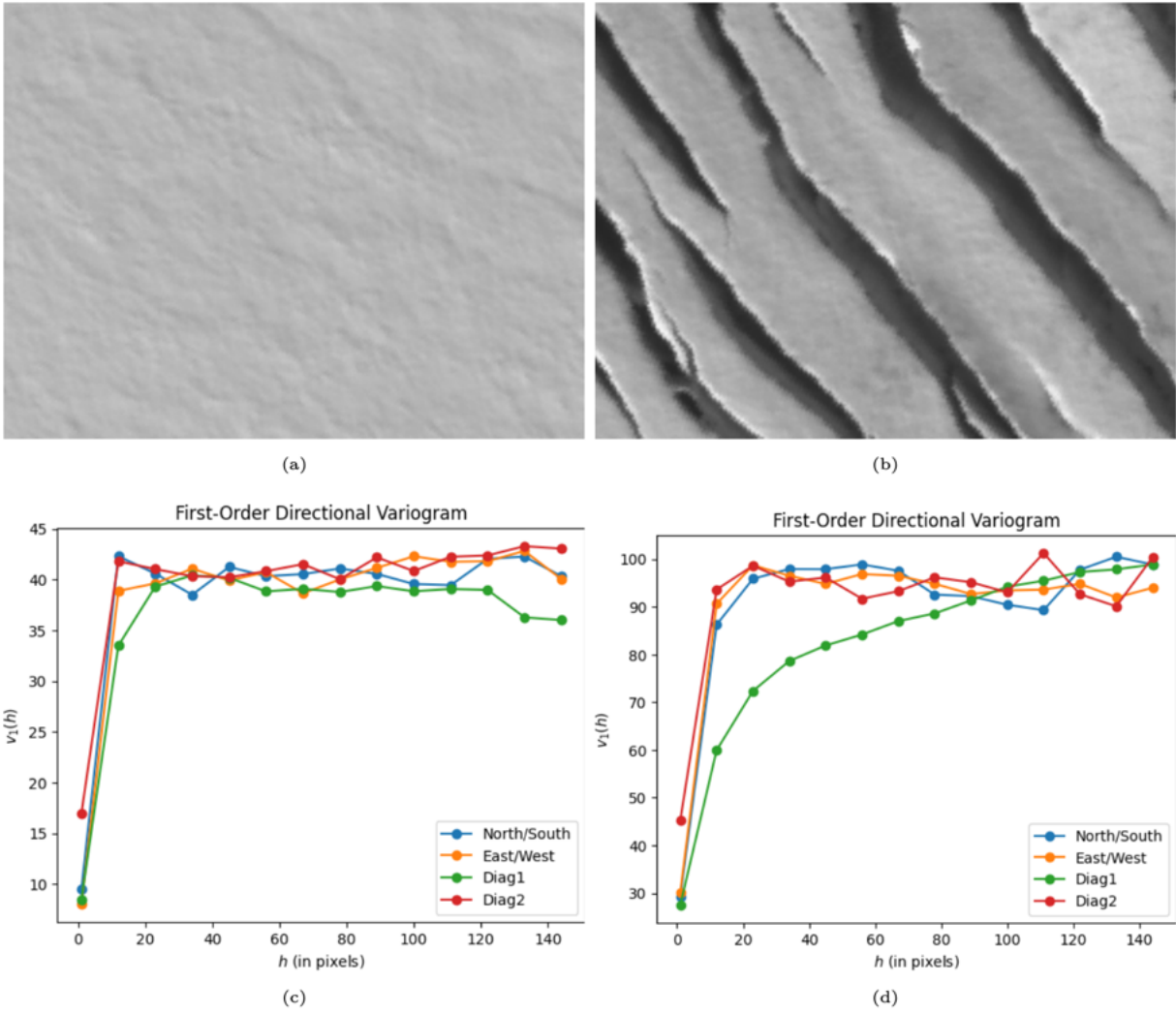


Figure (2). Typical images and associated directional variograms. (a) An example of an input image containing only undisturbed snow, which displays relatively uniform surface characteristics. (b) An example of an input image containing strong parallel crevasses, a prominent and repeating surface characteristic. (c) Output of the directional variogram $v_1(h)$ from 3.2a in 4 directions for 14 values of h . Note the relative uniformity of output values across all directions. (d) Output of the directional variogram $v_1(h)$ from (b) in 4 directions for 14 values of h . Note the much higher baseline values than in (c). Also note and the sharp contrast between the Diag1 direction (defined as diagonally from top left to bottom right) and the other directions. In this case, the Diag1 direction is nearly parallel to the direction of the crevasses and thus it does not reach the same sill as the other directions until a much higher lag value.



Figure (3). Airborne photographs of the four basic crevasse types.

Photographs acquired during the 2017 Negribreen campaign of the authors (Flight 2, 2017-07-15). (a) One-directional crevasses (DSC_0063). (b) Multi-directional crevasses (foreground, DSC_245). (c) Shear crevasses (DSC_0221). (d) Chaos or shear/chaos crevasses (DSC_0199). (e) Shear crevasses, co-called “shear holes” (DSC_0344). (f) Chaos or shear/chaos crevasses (DSC_0198). Example of chaos crevasses with a shear component.

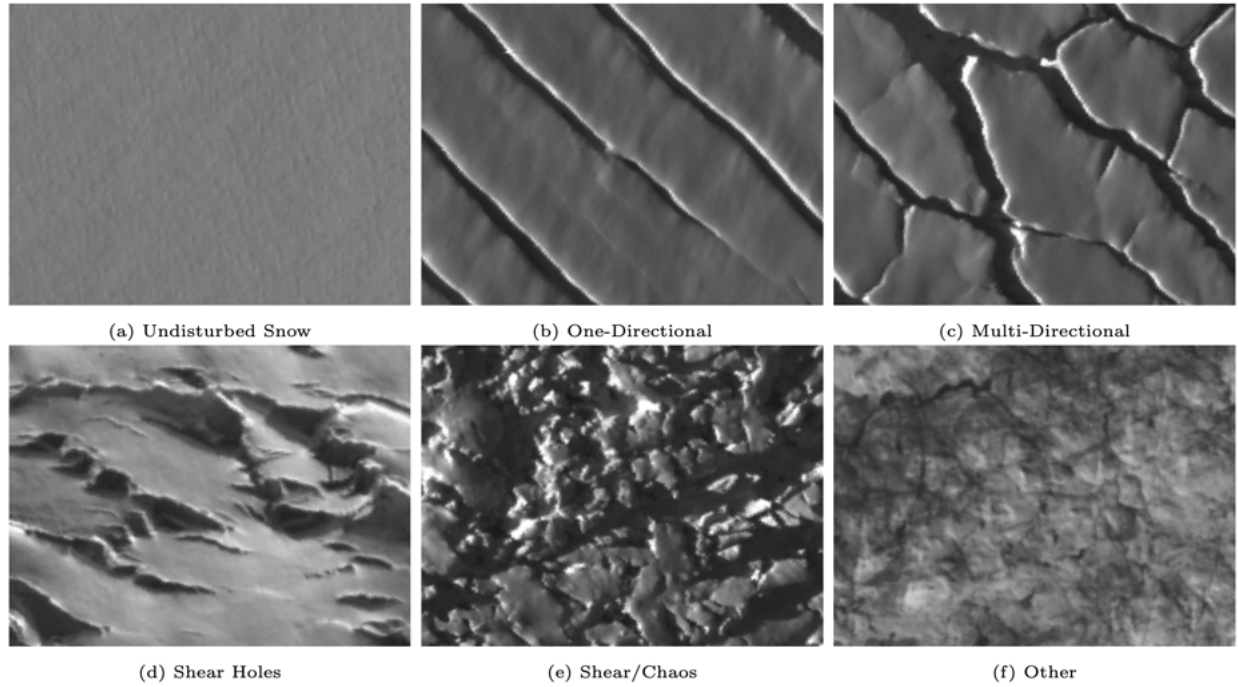


Figure (4). Examples of “split images” of the 6 surface classes, subselected from WorldView satellite imagery. (a) Undisturbed snow/ice, (b) One-directional crevasses, (c) Multi-directional crevasses, (d) Shear crevasses (shear holes), (e) Chaos crevasses (shear/chaos) and (f) Other (not classified). Same crevasse classes as illustrated in aerial photographs in Figure 2, with a class for undisturbed snow/ice and a rest class “other” added.

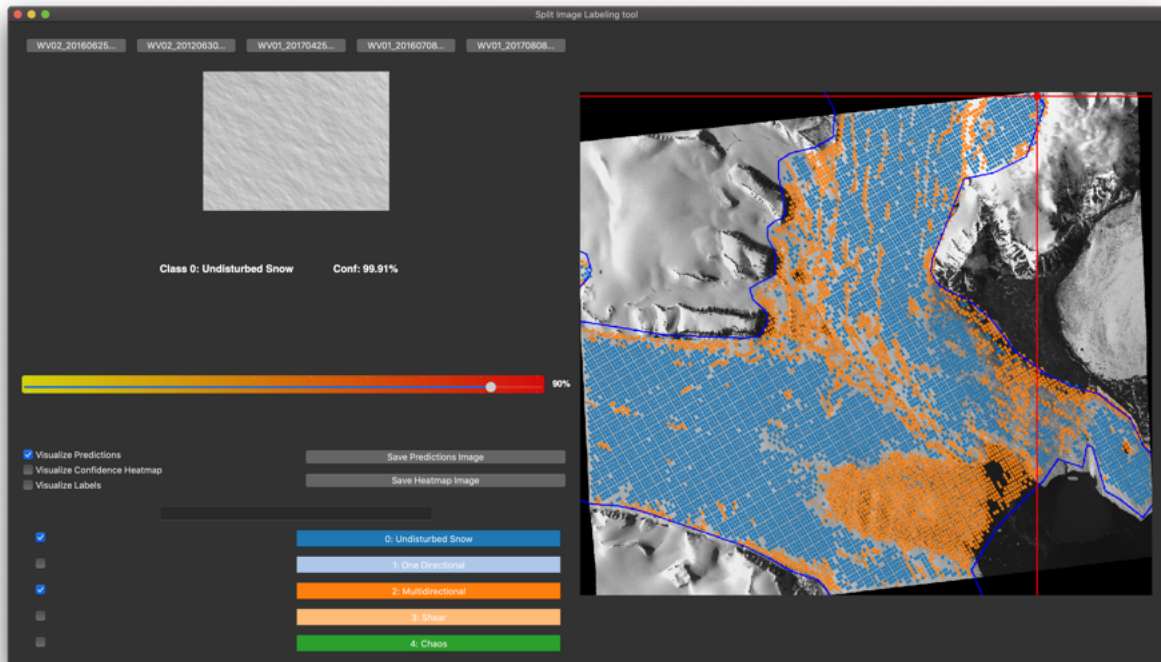


Figure (5). Split Image Explorer tool in action. Red cross-hairs select the split-image viewed in the upper left. Label options are displayed in the lower left. On the right, full classification and/or labeling results are viewed overlaying the full starting image (here a Worldview-2 image from 2016-06-25). The sliding bar at the center left allows visualization filters based on confidence levels. Finally, the tool has the options to switch between classification results for different starting images as seen in the upper left for various Worldview images.

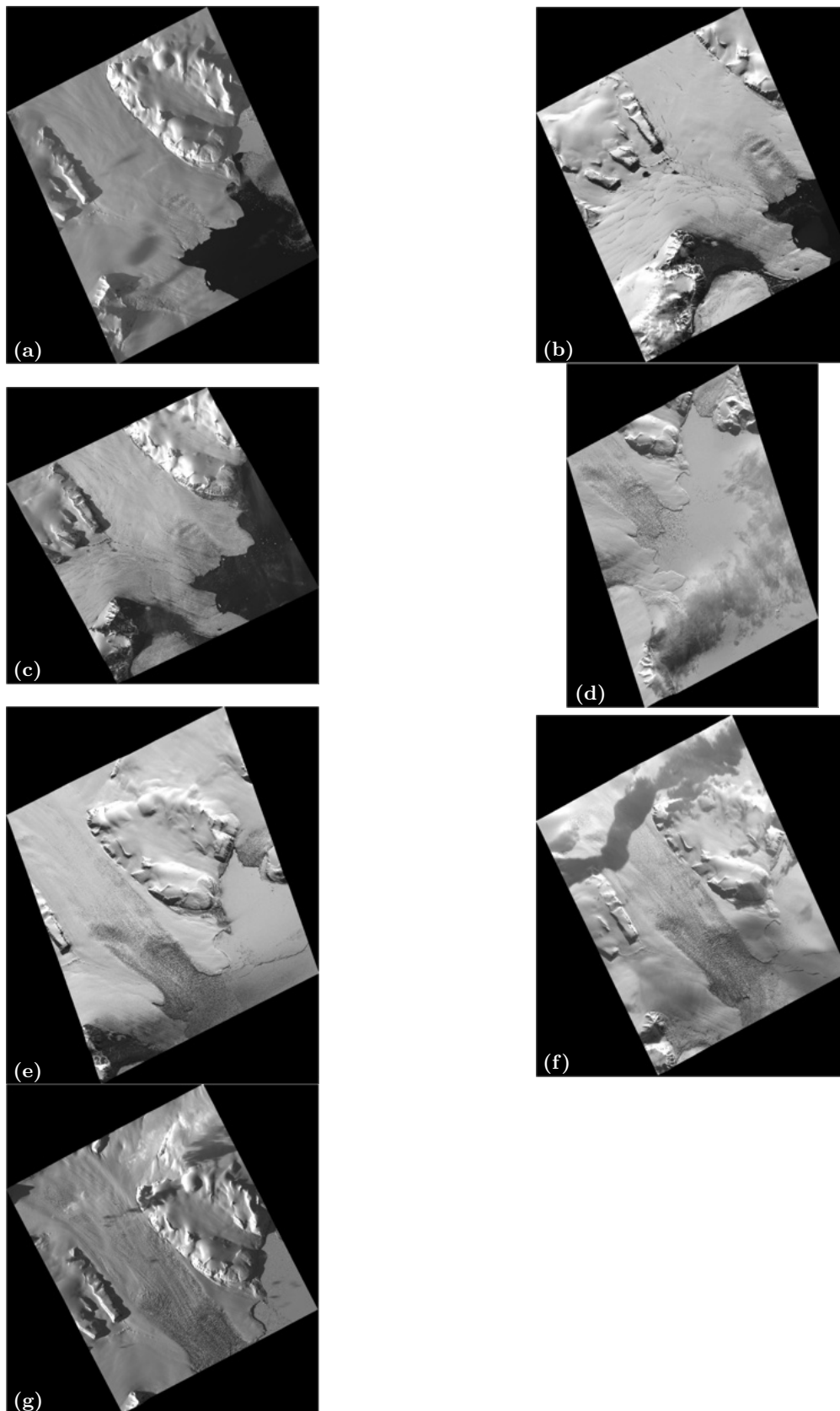


Figure (6). Full Worldview imagery used in classification time series analysis. (a) Worldview-1 acquired 2016-05-20, (b) Worldview-2 acquired 2016-06-25, (c) Worldview-1 acquired 2016-07-08, (d) Worldview-1 acquired 2017-04-25, (e) Worldview-1 acquired 2017-05-30, (f) Worldview-1 acquired 2018-04-29 and (g) Worldview-1 acquired 2018-05-26.

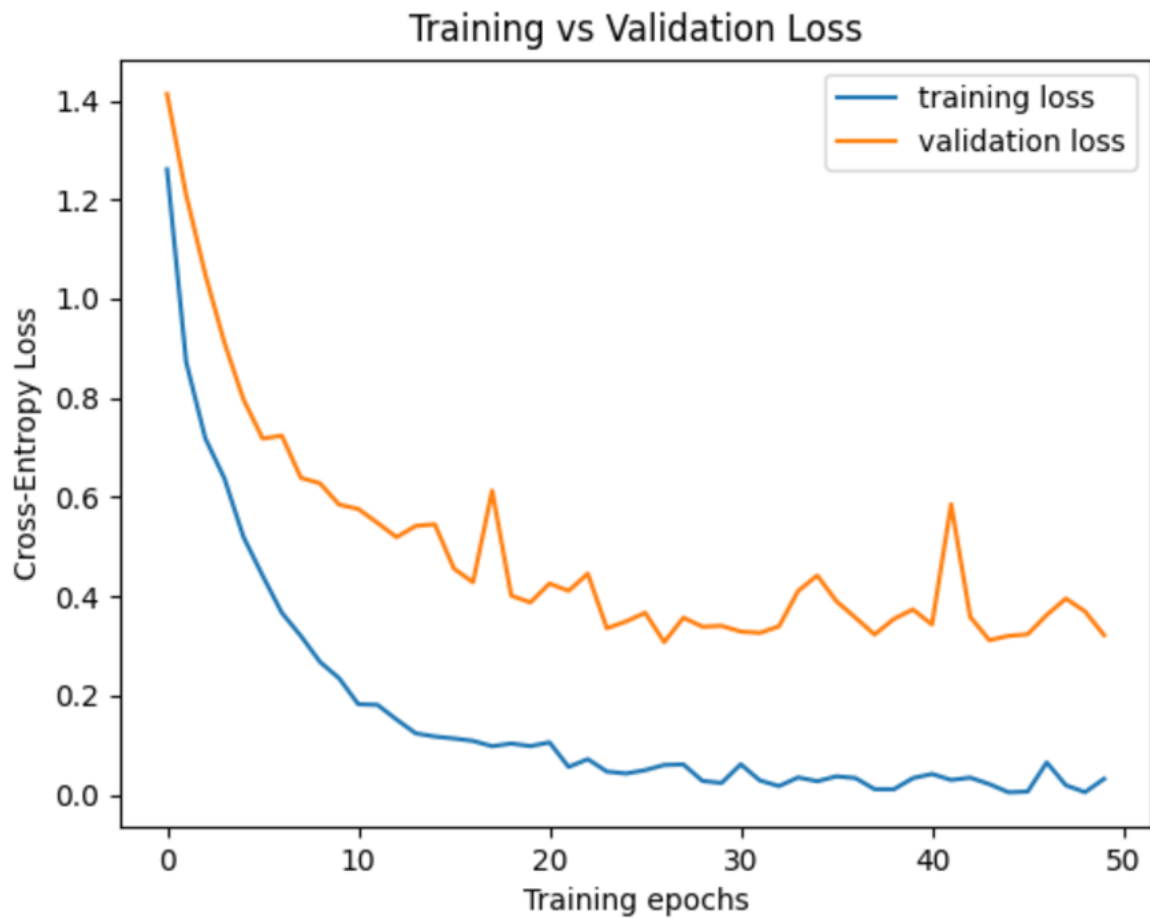


Figure (7). An example of overfitting from a test run of the Resnet18 model trained with a training dataset of 1,362 split images. The loss function used is the cross-entropy loss. The training loss approximates zero, whereas the validation loss stays high.

VarioCNN

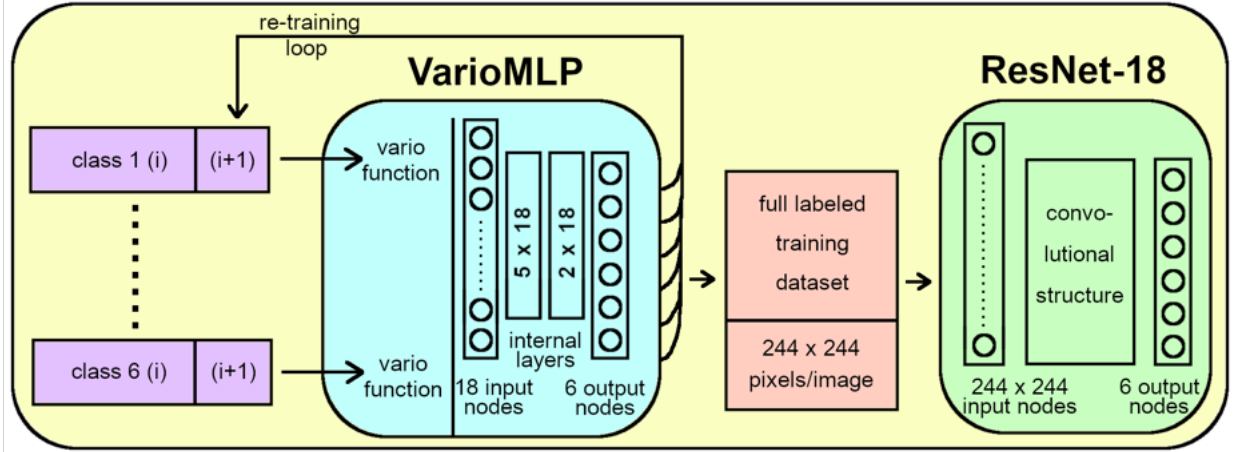


Figure (8). VarioCNN, derivation and architecture. Training data set (step(1)) derived by expert (structural glaciologist), includes split images representing 6 crevasse classes. Vario functions calculated per input image, vario values activate input nodes of VarioMLP (number of input nodes=number of vario values; number of input nodes is a variable model parameter). VarioMLP example with two internal layers of sizes [5,2]=[5 x 18, 2 x 18], output layer with 6 nodes; number of output nodes=number of crevasse classes, a variable model parameter. Retraining loop used for augmentation of training images per class from step (i) to step (i+1), for n steps. Full labeled training data set used to train ResNet-18. Note training images must be of size 244 by 244 pixels (fixed due to ResNet-18 architecture requirements).

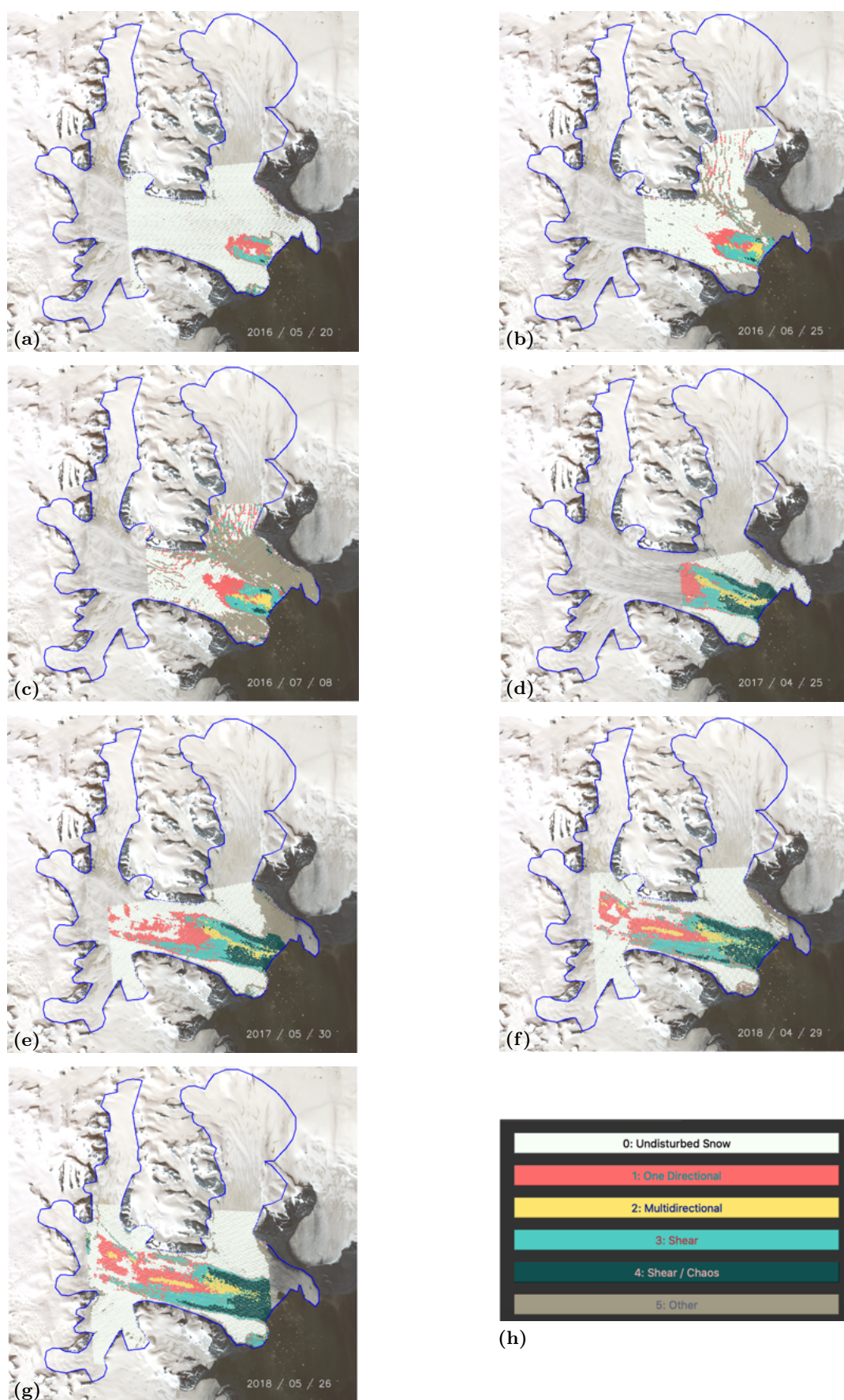


Figure (9). Time series of crevasse classes showing surge evolution. Classification of the following imagery: (a) Worldview-1 acquired 2016-05-20, (b) Worldview-2 acquired 2016-06-25, (c) Worldview-1 acquired 2016-07-08, (d) Worldview-1 acquired 2017-04-25, (e) Worldview-1 acquired 2017-05-30, (f) Worldview-1 acquired 2018-04-29 and (g) Worldview-1 acquired 2018-05-26. (h) Crevasse-type legend for the classification time series.

TABLES

	Worldview-1	Worldview-2
Dates of Operation	18 September 2007 - September 2023	8 October 2009 - present
Sensor Name	WorldView-60 camera	WorldView-110 camera
Channel(s)	(1) Panchromatic channel, spectral range: 450 nm - 900 nm	(1) Panchromatic channel, spectral range: 450 nm - 800 nm, (2) 8-band Multispectral channel, spectral range: 400 nm - 1050 nm
Resolution	0.50 m - 0.55 m Ground Sampling Distance (GSD)	(1) 0.46 m - 0.52 m GSD, (2) 1.8 m - 2.4 m GSD
Swath Width	17.9 km	16.4 km
Field of View	2.12°	1.28°
Orbit Altitude	496 km	767 km
Inclination	97.2°	97.8°
Orbit Period	94.6 minutes	100.2 minutes
Revisit Time	1.7 days - 5.9 days	~1.1 days

Table (1). Instrument specifications for the Worldview-1 and Worldview-2 satellites (Maxar). Both satellites provide high-resolution imagery using the pushbroom scanning technique.

Source image	Total split images generated	Labeled split images final dataset	Time series subfigure
WV01_20160520215109	10,142	67	(a)
WV02_20160625170309	11,834	608	(b)
WV01_20160708211932	10,379	164	(c)
WV01_20170425150931	4797	380	(d)
WV01_20170530144716	9209	319	(e)
WV01_20170808153226	8448	23	N/A
WV01_20180429140517	11,002	479	(f)
WV01_20180429140558	7614	543	N/A
WV01_20180526211954	12,191	634	N/A
WV01_20180526211859	13,647	228	(g)
WV01_20180530151039	9370	488	N/A
Total	108,623	3933	

Table (2). List of WoldView satellite image data sets and distribution of split images per source files in the final labeled training dataset of 3993 split images.

		Design	Implementation	Software Testing	Release v1.0
(A) Utility Scripts	Universal coordinate system transform	x	x	x	x
	Vis. of image/glacier overlap	x	x	x	x
	Directional variogram	x	x	x	x
	Roughness maps	x	x	x	x
(B) Geostatistical Methods: Functions	GUI-based creation and manipulation of Region of Interest and export of UTM coordinates	x	x	x	x
	First-order vario function	x	x	x	x
	Second-order vario function	x	x		
	First-order residual vario function	x	x	x	x
(C) Geostatistical Methods: Feature Vectors	Second-order residual vario function	x	x		
	Other functions	x			
	Direct use of vario function	x	x	x	x
	Classification parameters	x			
(D) Machine Learning Models	Feature vector built from classification params	x			
	Multi-layer perceptron model	x	x	x	x
	Resnet-18 CNN model	x	x	x	x
	Alternative machine learning models	x			
(E) Machine Learning Training and Validation	Deterministic class association	x			
	Universal training script	x	x	x	x
	Universal testing script	x	x	x	x
	Overall classification visualization	x	x	x	x
(F) Dataset Management	Visualization of losses	x	x	x	x
	Threshold visualization of classification confidence values	x	x	x	x
	Visualization of split image location from classifications	x	x	x	x
	Visualization of split images from classifications	x	x	x	x
(F) Dataset Management	Generalized split image dataset class	x	x	x	x
	Generalized split image dataset creation tool	x	x	x	x
	Universal split image dataset loader	x	x	x	x
	GUI-based labeling of individual split images	x	x	x	x
(F) Dataset Management	GUI-based batch-labeling of split images	x	x	x	x
	GUI and visualizations dynamically adapt to creation of new classes	x	x	x	x
	NetCDF-formatted output of labels and classifications	x	x	x	x
	Integration of multiple satellite images	x	x	x	x

Table (3). GEOCLASS-image software system, components design, implementation, testing and release, updated November 30th, 2023.

Class	Total labeled examples
Undisturbed Snow	538
One-Directional	953
Multi-Directional	614
Shear Holes	742
Shear / Chaos	564
Other	522
Total	3933

Table (4). Number of split images for each of the 6 crevasse classes in the final labeled training data set of 3933 split-images, derived using VarioMLP and several training and reinforcement loops.

Lag steps in variogram	Min. loss achieved
10	1.56
12	0.61
14	0.88
16	1.14
18	0.54
20	1.48

Table (5). Minimum cross entropy loss achieved by VariogramMLP model with different values for number of lag steps in the variogram step. Hidden layer shape was fixed at $[5, 2]$ for all runs.

Hidden layer shape	Min. validation loss
[2, 2]	0.54
[5, 2]	0.52
[5, 5, 2]	1.21
[10, 5, 2]	0.80
[10, 10, 2]	0.78

Table (6). Minimum cross entropy loss achieved by VarioMLP model for each of the 5 hidden layer shapes tested. Variogram lag steps was fixed at 18 for all runs. It is interesting to note that after [5, 2], adding both depth and width to the hidden layers results in decreased validation performance.

Batch size	Min. validation loss achieved
1	0.169
2	0.115
4	0.245
8	0.925

Table (7). Minimum validation loss achieved on the full-sized validation dataset of 786 split images using batch sizes of 1, 2, 4 and 8. The best result was achieved from a batch size of 2, a slight improvement from 1.