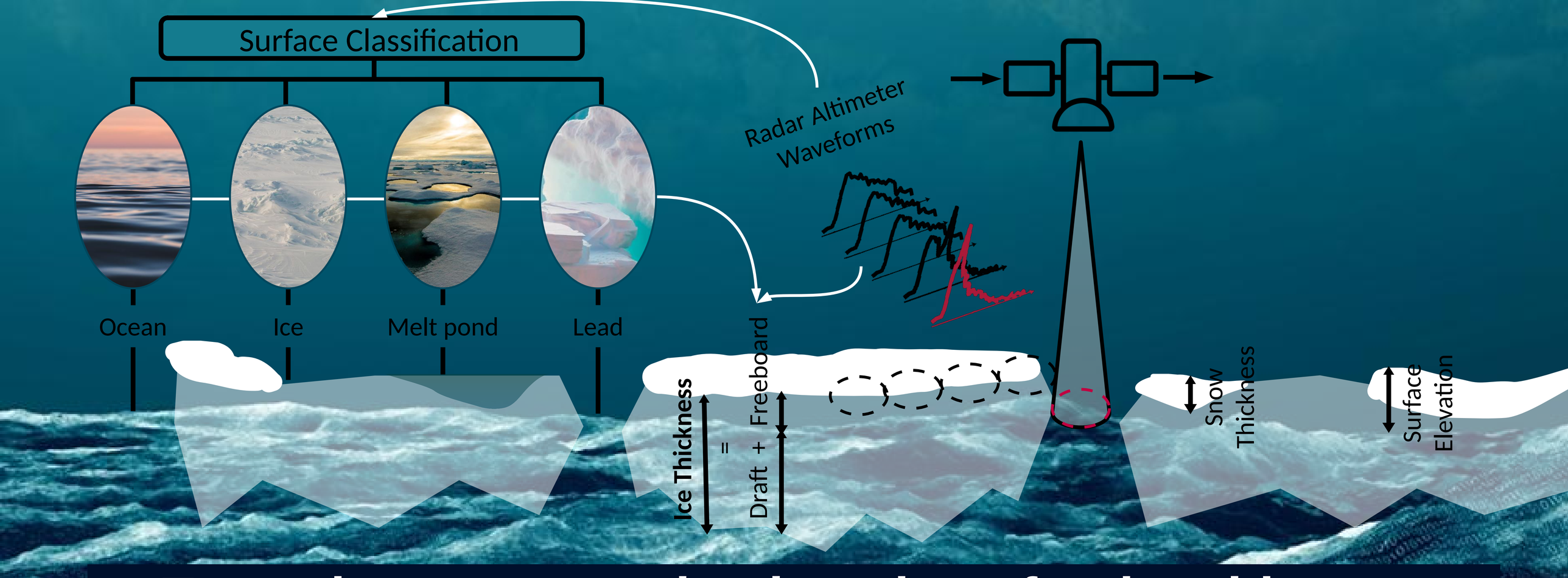


Towards representation learning of radar altimeter waveforms for sea ice surface classification

Lena Happ[1,2,4], Dr. Stefan Hendricks[1,4], Sonali Patil[3], Lars Kaleschke[1], Riccardo Fellegara[3], Prof. Dr. Andreas Gerndt[2,3] lena.happ@awi.de

1 ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG
2 Universität Bremen
3 DLR
4 MarDATA HELMHOLTZ SCHOOL FOR MARINE DATA SCIENCE

Introduction:

We are aiming for the classification of of polar ocean surface types (e.g. different stages of development of sea ice) based on satellite radar altimeter waveforms. Deriving this information would allow for a better parametrization of ice densities and thus an improved estimate of sea ice thickness measurements. However, machine learning models for sea ice classification often depend on supervised training, which is vulnerable to uncertainties in labeled data, especially in polar regions. On the other hand, unsupervised clustering cannot be easily applied to high dimensional data. Thus, we are adapting self-supervised contrastive learning frameworks to find a lower dimensional representation of the radar altimeter waveforms that is optimized for the separation of surface types. In particular we are defining an application specific batch sampling strategy and similarity distribution to draw positive pairs. With this strategy we can improve the downstream task of ice type classification compared to tradition representations by waveform parameters.

Data:

SRAL sensor on Sentinel-3A and -3B
October 2022 to April 2023

Class Labels for validation:
Operational Ice Charts and
OSI SAF Global Sea Ice Type Classification

New Ice – First Year Ice – Multi Year Ice
Leads excluded by waveform parameter thresholds

Results:

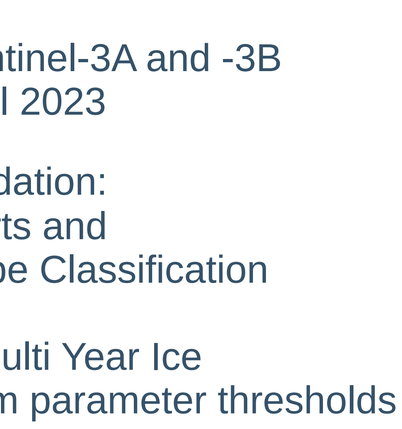
Improvement of class separability measured by accuracy of k-nearest Neighbor (kNN) based on class labels

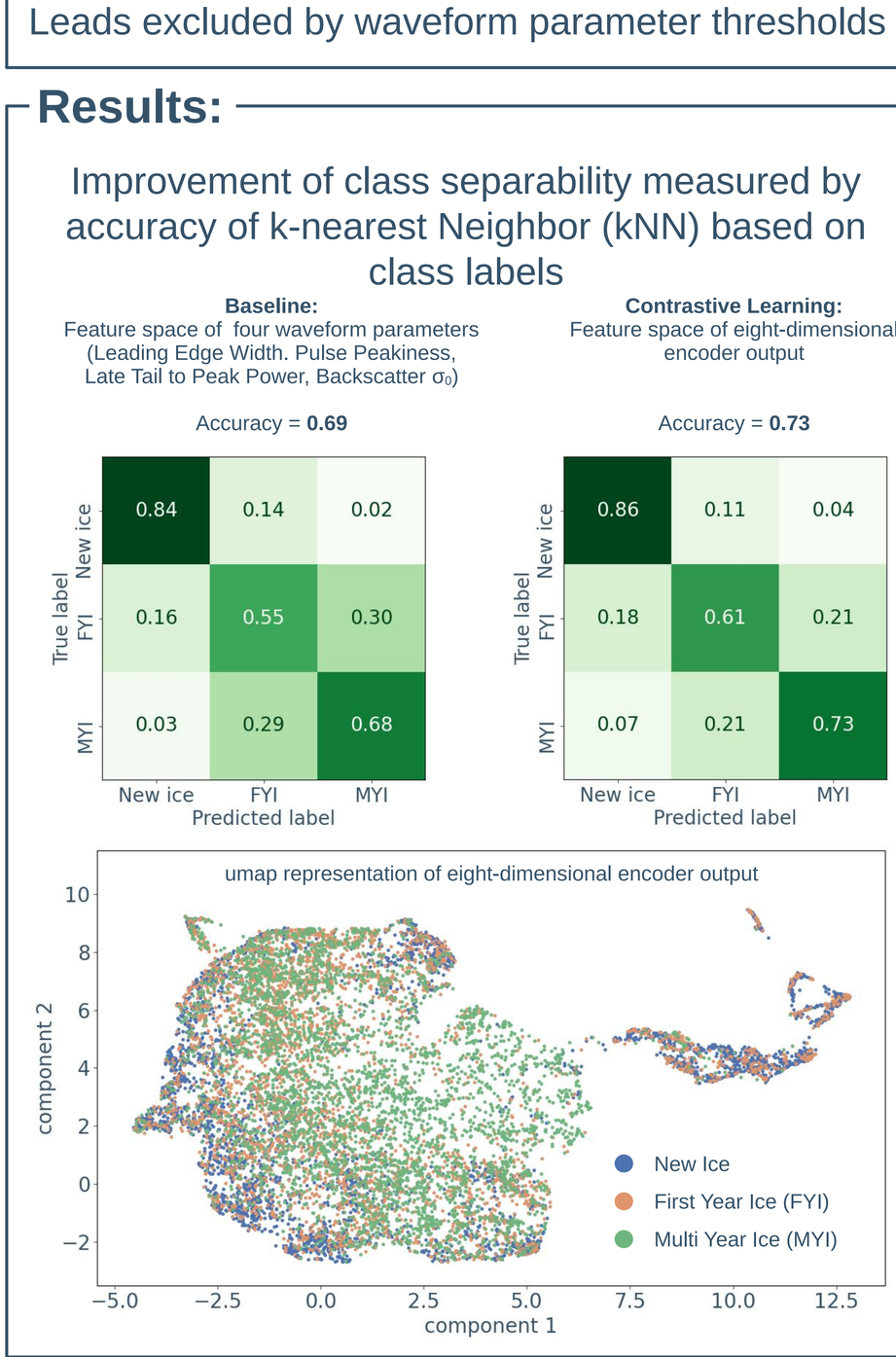
Baseline:
Feature space of four waveform parameters (Leading Edge Width. Pulse Peakiness, Late Tail to Peak Power, Backscatter σ_0)

Accuracy = **0.69**

Contrastive Learning:
Feature space of eight-dimensional encoder output

Accuracy = **0.73**



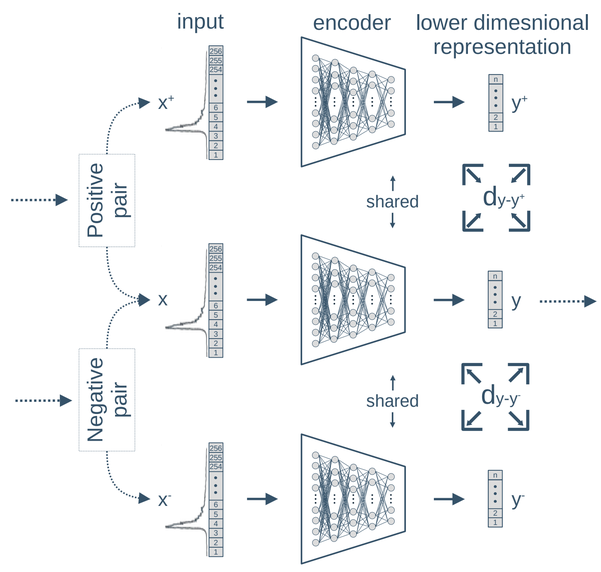


Method:

Self-Supervised **Contrastive Learning** [1,2]
with an application specific definition of positive pairs and batch sampler

Positive pairs are defined based on geographical and temporal proximity of footprints
→
A sequence of footprints in a trajectory of satellite form positive pairs, all others form negative pairs

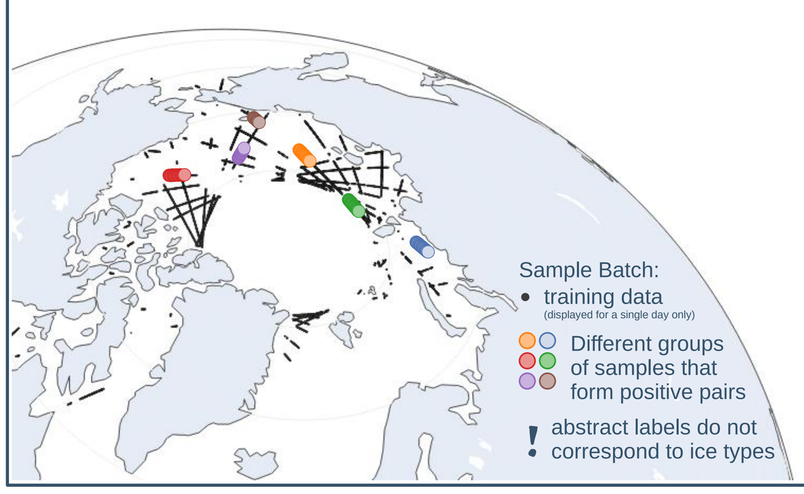
Batch Sampler:
Samples sequences of up to 200 footprints from random locations



Encoder output used in downstream task: sea ice surface classification

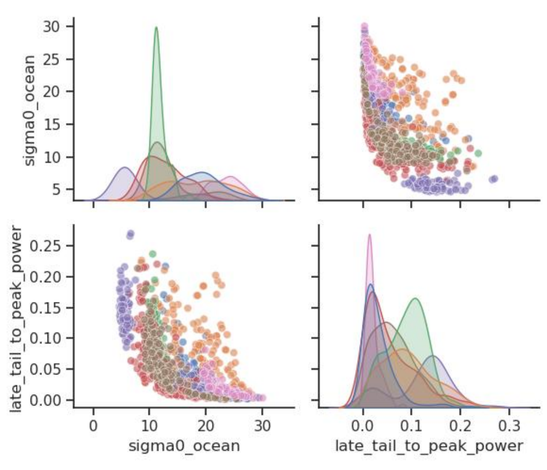
Should be optimized to separate different ice classes

We validate representation using the accuracy of a kNN algorithm on this feature space



Scatter-plot of waveform parameters for sample batch:

Within sequences of footprints waveform shapes are more homogeneous compared to ice chart labels



Discussion:

How to find subgroups in representation space, that can be linked to sub categories of ice types?

How could we validate results independent of ice chart labels?

How can we interpret the additional information in encoder output compared to traditional waveform parameter representation?

How can we refine definition of positive pairs and batch sampler to achieve an even better separability of ice classes?

[1] SimCLR: Chen et al, 2020, A simple framework for contrastive learning of visual representations, ICML'20
[2] SupCon:Khosla et al, 2020, Supervised Contrastive Learning, NeurIPS'20

The first author, Lena Happ, is funded through the Helmholtz School for Marine Data Science (MarDATA), Grant No. HIDSS-0005.