The Cryosphere, 17, 4957–4977, 2023
https://doi.org/10.5194/tc-17-4957-2023

The Cryosphere
Open Access
EGU

Out-of-the-box calving-front detection method using deep learning

Oskar Herrmann[1], Nora Gourmelon[2], Thorsten Seehaus[1], Andreas Maier[2], Johannes J. Fürst[1], Matthias H. Braun[1], and Vincent Christlein[2]

[1]Institute of Geography, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[2]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Correspondence: Oskar Herrmann (oskar.herrmann@fau.de)

Received: 20 February 2023 – Discussion started: 28 February 2023
Revised: 28 September 2023 – Accepted: 13 October 2023 – Published: 24 November 2023

Abstract. Glaciers across the globe react to the changing climate. Monitoring the transformation of glaciers is essential for projecting their contribution to global mean sea level rise. The delineation of glacier-calving fronts is an important part of the satellite-based monitoring process. This work presents a calving-front extraction method based on the deep learning framework nnU-Net, which stands for no new U-Net. The framework automates the training of a popular neural network, called U-Net, designed for segmentation tasks. Our presented method marks the calving front in synthetic aperture radar (SAR) images of glaciers. The images are taken by six different sensor systems. A benchmark dataset for calving-front extraction is used for training and evaluation. The dataset contains two labels for each image. One label denotes a classic image segmentation into different zones (glacier, ocean, rock, and no information available). The other label marks the edge between the glacier and the ocean, i.e., the calving front. In this work, the nnU-Net is modified to predict both labels simultaneously. In the field of machine learning, the prediction of multiple labels is referred to as multi-task learning (MTL). The resulting predictions of both labels benefit from simultaneous optimization. For further testing of the capabilities of MTL, two different network architectures are compared, and an additional task, the segmentation of the glacier outline, is added to the training. In the end, we show that fusing the label of the calving front and the zone label is the most efficient way to optimize both tasks with no significant accuracy reduction compared to the MTL neural-network architectures. The automatic detection of the calving front with an nnU-Net trained on fused labels improves from the baseline mean distance error (MDE) of 753 ± 76 to 541 ± 84 m. The scripts for our experiments are published on GitHub (https://github.com/ho11laqe/nnUNet_calvingfront_detection, last access: 20 November 2023). An easy-access version is published on Hugging Face (https://huggingface.co/spaces/ho11laqe/nnUNet_calvingfront_detection, last access: 20 November 2023).

1 Introduction

Unlike the large majority of land-terminating glaciers, marine and lake-terminating (MALT) glaciers reach a water body at a low elevation. The contact surface is often referred to as the calving front. Their ice is lost by sub-marine melting or calving, i.e., ice that breaks off, gets de-connected, and starts to float freely in the form of icebergs or ice floes. Both processes, sub-marine melting and calving, determine the total frontal ablation, i.e., the mass loss at the calving front. Frontal ablation is often a dominant factor in the total mass budget of MALT glaciers (McNabb et al., 2015; Shepherd et al., 2018; Minowa et al., 2021). Besides its importance in the total glacier mass balance, the representation of processes controlling frontal ablation is currently a pressing task for numerical glacier models (Beer et al., 2021). Neglecting frontal ablation can introduce an important bias. Recinos et al. (2019) analyzed the impact on ice thickness reconstruction (based on mass conservation) in Alaska and reported an underestimation of 19 % on regional scales and up to 30 % on glacier scales. Various successful approaches exist to parameterize frontal ablation for individual glaciers. Still, the implementation in large-scale or global models is limited by the amount and quality of measurements for constraining the models (Recinos et al., 2019). Thus, large-scale measurements (ideally time series) of frontal ablation are de-

manded by the modeling community (Recinos et al., 2021). Driving forces of frontal ablation are, on the one hand, ice flux (higher flux, e.g., due to bed lubrication by meltwater, can trigger calving events) and, on the other hand, marine factors such as ocean temperature, fjord bathymetry, and sea ice or ice mélange conditions (Carr et al., 2014; Straneo et al., 2013). For example, the persistence of the ice mélange in front of the glacier can stabilize the calving front and affect the glacier dynamics. In contrast, the breakup of the ice mélange can lead to increased calving and ice flow at the glacier terminus (Amundson et al., 2010; Kneib-Walter et al., 2021; Rott et al., 2020). Moreover, a significant frontal retreat can also indicate a retreat of the grounding zone (Friedl et al., 2018). The retreat of the grounding zone of a glacier with retrograde bedrock formation will lead to further grounding-zone retreat, resulting in increased ice loss and destabilization of the glacier or ice stream (Robel et al., 2016). Thus, information on the temporal variability of the calving-front position provides fundamental information on the state of the glacier or ice stream. Therefore, the glacier area has been defined as an essential climate variable (ECV) product by the World Meteorological Organization (WMO). Calving-front positions were usually manually mapped using different remote sensing imagery (Baumhoer et al., 2018). Only a few studies applied automatic or semi-automatic approaches. In polar regions, the ocean downstream of the glaciers is often covered by sea ice and calved-off icebergs, forming the so-called ice mélange, making calving-front delineation a challenging task, even when captured by hand. Deep learning approaches have shown high potential for carrying out such complex segmentation tasks, e.g., on medical imagery (Jang and Cho, 2019). In recent years, the application of deep learning techniques for glacier front detection started (Zhang et al., 2019; Cheng et al., 2021; Baumhoer et al., 2019; Hartmann et al., 2021; Mohajerani et al., 2019; Baumhoer et al., 2021; Zhang et al., 2021; Heidler et al., 2021; Marochov et al., 2021; Loebel et al., 2022). Calving fronts can be located in both optical and synthetic aperture radar (SAR) imagery. In optical imagery, calving fronts are more easily distinguishable, whereas synthetic aperture radar imagery has a higher scene availability as this is independent of daytime, season, and cloud coverage (Baumhoer et al., 2018). A direct comparison between the results of existing deep-learning-based calving-front extraction studies is not possible as the models have been trained on different data, tested on different test sets, and evaluated using slightly differing metrics.

The benchmark dataset published by Gourmelon et al. (2022a) provides 681 SAR images of calving fronts. SAR imagery is independent of sunlight and cloud coverage, enabling continuous temporal coverage of the observation area, but compared to optical data, it has only one channel and has more speckle noise. For every SAR image, two labels are provided. One label provides four classes: ocean, glacier, rock, and no information available (e.g., radar shadow and layover areas, areas outside the swath). The other label marks the calving front with a one-pixel-wide line. Based on the training set of the dataset, Gourmelon et al. (2022a) train a modified U-Net for each label. One U-Net solves the task of glacier segmentation, and one detects the calving front. On the two test glaciers of the dataset, the segmentation model achieves an IoU (Intersection over Union) of 67.7 ± 0.6. By taking the boundary between ocean and glacier, they extract a prediction of the calving front from the zone prediction with a mean distance error (MDE) of 753 ± 76 m. The model trained directly on the front label achieved an MDE of 887 ± 189 m.

In this work, we present a method that utilizes both labels of the dataset instead of training separate models for each task (see Fig. 1). Multi-task learning (MTL) is a technique for machine learning algorithms that uses one model to tackle multiple tasks. In most cases, the potentially larger dataset and higher information content lead to higher performance for the individual tasks (Bischke et al., 2019; He et al., 2021; Li et al., 2019; Amyar et al., 2020; Chen et al., 2019).

Our method is based on the nnU-Net (no new U-Net) proposed by Isensee et al. (2021), which is an out-of-the-box framework for training the U-Net. The framework contains proven deep learning techniques and established hyperparameter values, as well as rule-based parameters that depend on the properties of the dataset and available GPU memory. Therefore, the nnU-Net is a powerful tool that simplifies the application of deep learning algorithms. Its performance in the segmentation of SAR data has not been tested, and the availability of two labels suggests the modification of the nnU-Net for MTL. As a baseline for our evaluation of the nnU-Net, we use the results of Gourmelon et al. (2022a).

In particular, our contributions are as follows: (1) visualizations showing the temporal and spatial distribution of the dataset by Gourmelon et al. (2022a); (2) application and evaluation of the nnU-Net for calving-front detection and zone segmentation; (3) two different modifications of the original nnU-Net to incorporate both labels for multi-task learning; (4) testing if an artificial third label improves the calving-front detection; (5) introduction of an efficient approach that fuses the two labels of the dataset; and (6) analysis of the influence of season, glacier, and satellite on the performance of the model.

The paper is organized as follows: after presenting the related work in Sect. 2, we give an overview of the dataset (Gourmelon et al., 2022a) in Sect. 3. Section 4 explains the method and our six experimental setups. Section 5 examines the results and analyses the influence of different properties of the satellite images. Section 6 summarises the work.

2 Related work

Automated monitoring of glacier-covered areas is a growing research field. Recent glacier monitoring uses deep learning

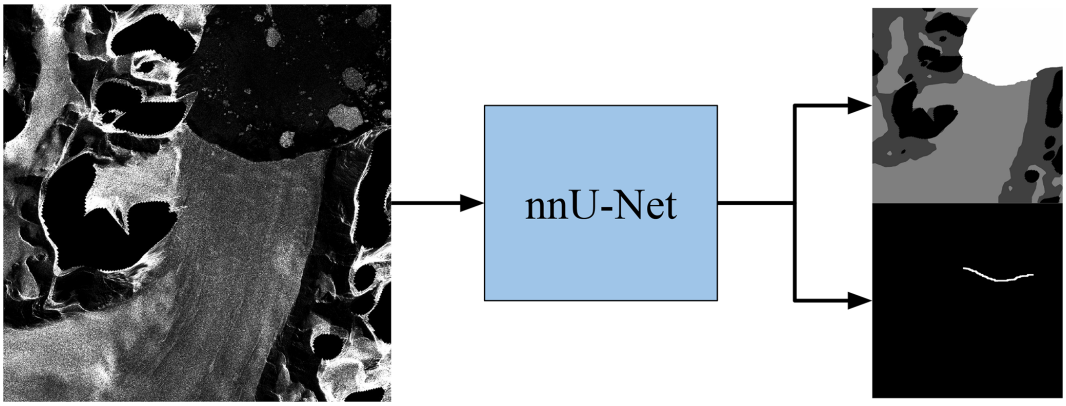


Figure 1. Illustration of the modified nnU-Net for simultaneous prediction of landscape zones and the glacier front. On the left is an exemplary satellite image of the Crane Glacier taken by TDX on 24 June 2011. On the right, the upper image shows the four segmentation classes: ocean (white), glacier (light gray), rock (dark gray), and no information available (black). The lower image shows the glacier's calving front as a white line versus the black background.

methods due to the increasing availability of satellite images and computing power. Many methods are based on the U-Net (Ronneberger et al., 2015). They modify the vanilla U-Net for better performance in calving-front detection (Loebel et al., 2022; Mohajerani et al., 2019; Zhang et al., 2019). One approach is to segment the images into different areas and extract the calving front as the border between segmentation areas (Hartmann et al., 2021; Zhang et al., 2019; Baumhoer et al., 2019; Periyasamy et al., 2022; Loebel et al., 2022). Another approach directly trains a model on the position of the calving front (Davari et al., 2022). This task suffers from severe class imbalance due to the thin calving front. Researchers approach this problem by creating a distance map from every pixel to the front line. The network is trained on the distance map instead of the thin front line. The actual front-line prediction is then extracted during post-processing.

Other works use the segmentation network DeepLabv3 (Chen et al., 2018) to detect calving fronts. The main advantage of DeepLabv3 over U-Net is the atrous spatial pyramid pooling, which makes the network adaptable to different image resolutions (Zhang et al., 2021; Cheng et al., 2021).

The Calving Front Machine (CALFIN) proposed by Cheng et al. (2021) segments optical and SAR images into the ocean–land zones and extracts the calving front during post-processing with a topography map of the area. Researchers apply MTL with a late-branching architecture. They use two labels: a binary ocean mask and a binary calving-front mask. They achieve state-of-the-art predictions with an 86 m deviation from the measured calving front. A detailed comparison to the aforementioned U-Net-based methods by Mohajerani et al. (2019) and Baumhoer et al. (2019) revealed the generalization ability of CALFIN to other glacier datasets. The images had to be down-sampled for CALFIN. Therefore, the distance in meters doubles, but comparing the pixel distance errors reveals a similar performance. With 29 million parameters, CALFIN is still a large network that needs a large amount of training data. Chen et al. (2019) propose a similar approach for medical image segmentation and show that the individual tasks benefit from MTL. Several other approaches have advanced the U-Net architecture for MTL for medical applications (Abolvardi et al., 2020; Kholiavchenko et al., 2020; Li et al., 2019; Amyar et al., 2020). The work of Heidler et al. (2021) uses MTL for the segmentation of a binary ocean–land mask and for edge detection of the Antarctic coastline, where the calving front is just part of the coastline. They add task-specific heads for the two tasks to the U-Net. They achieve results with a deviation from the reference of 345 m compared to 483 m with the vanilla U-Net. To avoid distortions of the metric from areas far from the coast, the metric is calculated within 2 km of the true coastline. A further development of the holistically nested edge detection (HED)-UNet for Antarctic ice shelf front detection is proposed by Baumhoer et al. (2023). Their post-processing includes an elevation threshold of 110 m to remove erroneous classifications in the high-altitude dry-snow zones.

3 Dataset

The dataset used in this work is provided by Gourmelon et al. (2022b). It contains 681 synthetic aperture radar (SAR) images of seven marine-terminating glaciers taken by six different satellites. Two glaciers are located in the Northern Hemisphere, namely Columbia Glacier in Alaska and Jacobshavn Isbrae (Sermeq Kujalleq) (JAK) in Greenland. The five glaciers in the Southern Hemisphere are all located on the Antarctic Peninsula (see Fig. 2). The Crane, Mapple, and Jorum glaciers are closest to the south pole, followed by Dinsmoore–Bombardier–Edgeworth (DBE). They are lo-

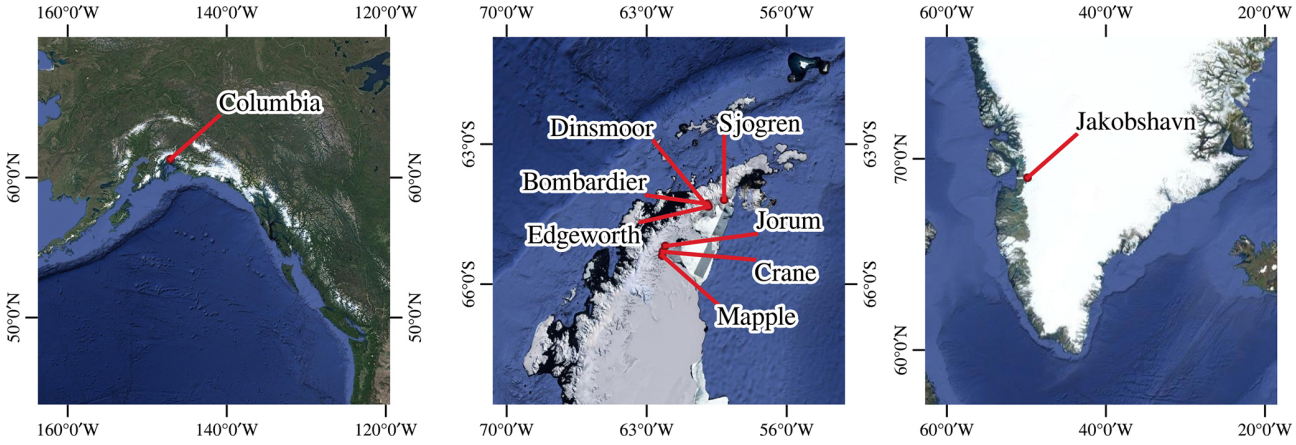


Figure 2. Location of the seven benchmark glaciers: Columbia Glacier in Alaska, Crane, Mapple, Jorum, Dinsmoore–Bombardier–Edgeworth (DBE), Sjogren Glacier in the Antarctic Peninsula, and Jakobshavn Isbrae Glacier in Greenland. The background images are taken from the Google Maps Satellite imagery layer © Google 2019.

cated so close together that they are treated as one glacier site. The last glacier is the Sjögren-Inlet Glacier.

Table 1 lists the seven glacier sites with the number of images and the area they show. The table also depicts the train–test split. The samples of the training set are used for the parameter optimization of the segmentation method, and the test set is exclusively used for evaluating the method. The test set contains the glaciers Mapple and Columbia. Columbia is the only mountain glacier in the dataset. The other glaciers are outlet glaciers of polar ice sheets. Therefore, Columbia can be seen as a benchmark glacier for the generalization capability of the segmentation method. Columbia and Mapple also differ in terms of the shape of the glacier. The images of Columbia show multiple calving fronts in one image, and the flow of the glacier arms goes in different directions. On the other hand, Mapple is a less complex glacier, moving in only one direction with one clearly defined calving front between the lateral fjord side walls.

The dataset contains two labels and a bounding box of the glacier for each SAR image. One shows a mask of the different zones of the glacier (ocean, glacier, no information available, rock). The other label contains a one-pixel-wide line representing the calving front. The bounding box is not utilized for our method. A sample of each glacier in the training set with its corresponding labels is shown in Fig. 3. Predicting the zone mask can be seen as a classic segmentation problem. The calving front can also be extracted from the zone label by taking the border between ocean and ice areas in the corresponding bounding box. The direct delineation of the calving front is a more difficult task due to the high-class imbalance. Fewer than 1 % of the pixels are labeled as front pixels. Additionally, the structure of the class region is not a convex hull but a thin line.

Corresponding to the change in glacier area, the position of the calving front changes. In Fig. 4, the retreat of Columbia is visualized by plotting the calving-front position of every sample in a different color. The bright lines represent past calving fronts, and the dark-red lines represent the more recent positions. The constant retreat of Columbia is visible through the tiered lines. The visualization of the glaciers in the Southern Hemisphere is included in Appendix B. The change in the class distribution of the zone labels over time is shown in Appendix C.

Every glacier is captured by multiple satellites for a higher temporal resolution and extended observation periods, meaning that recordings of one glacier are captured by different SAR systems with different image resolutions. The resolution refers to the ground range resolution. The Environmental Satellite (ENVISAT), European Remote Sensing Satellite 2 (ERS-2), and Sentinel-1 (S1) have a resolution of 20 m, the phased-array-type L-band SAR (PALSAR) has a 17 m resolution, and TanDEM-X (TDX) has a 7 m per pixel resolution. Unfortunately, the dataset does not provide information about the polarization. In Fig. 5, a timeline of the images of each glacier visualizes the observation time and frequency of the images. The first two rows show the glaciers of the test set.

4 Method

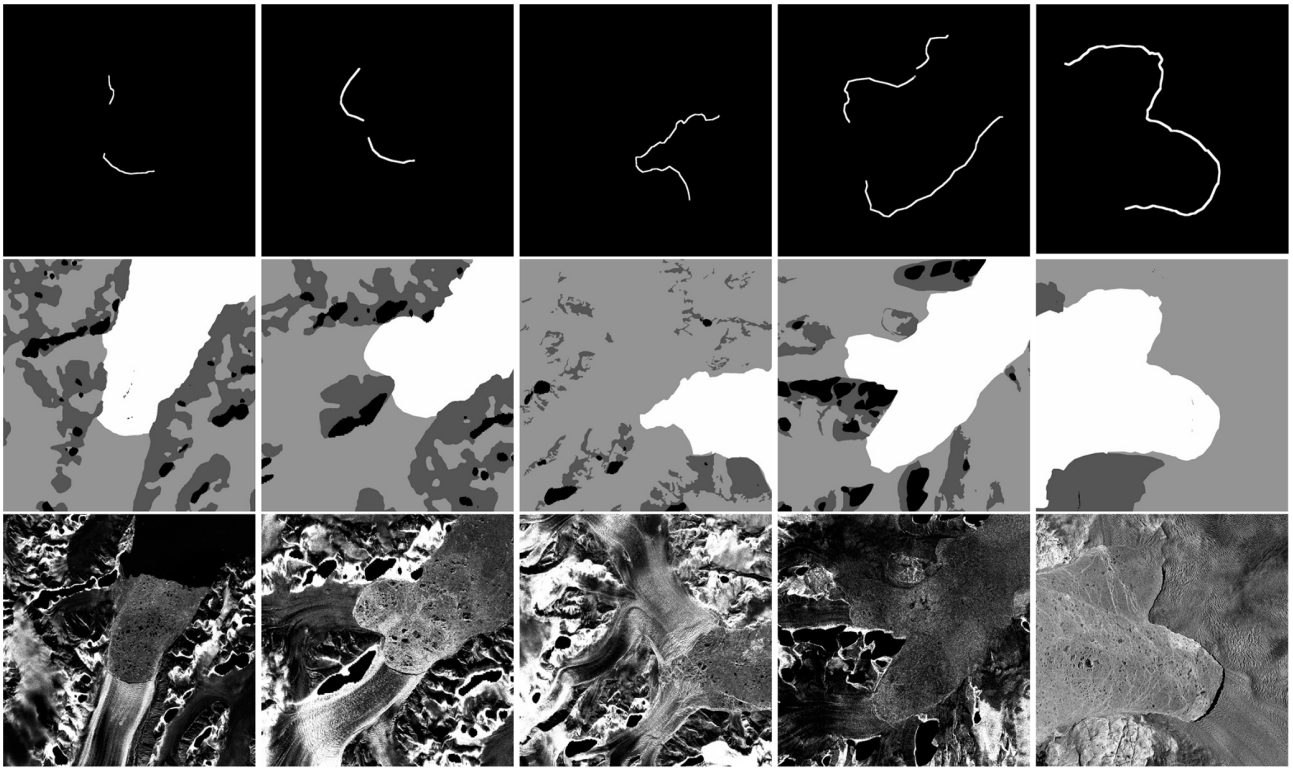
In this section, we explain our method, which includes the utilization of the nnU-Net as a framework that simplifies the training of a U-Net. We document our six experimental setups that aim to evaluate the impact of multi-task learning (MTL) on the training of the U-Net.

4.1 Background

In the field of deep learning, a lot of time and effort is put into the hyperparameter search. Isensee et al. (2021) proposes the

Table 1. Properties of the dataset, including a list of captured glaciers, train–test split, number of images per glacier, and covered area.

	Alaska	Antarctic Peninsula					Greenland
	Columbia	Mapple	Crane	Jorum	DBE	Sjögren-Inlet	Jakobshavn
Split	– test: 122 –	– train: 559 –					
No. of images	65	57	69	77	133	121	159
Area [km]	32×15	8×8	19×25	20×13	22×20	23×19	16×19



(a) Crane (b) Jorum (c) DBE (d) SI (e) Jakobshavn

Figure 3. Sample images of every glacier in the training set and their corresponding labels. The first row shows the front label with a black background and a one-pixel-wide white line representing the calving front. The second row contains the zone labels with four classes: ocean (white), glacier (light gray), rock (dark gray), and no information available (black). Black dots in the ocean zone of Crane are areas with no data available due to a small artifact at the former calving front of Crane Glacier in the digital elevation model (Cook et al., 2012) used for the orthorectification of the SAR data. SAR imagery was provided by DLR, ESA, and ASF.

nnU-Net that automates the manual tuning of hyperparameters. The nnU-Net is a framework around the U-Net architecture. It provides good default values for hyperparameters, rules to adapt hyperparameters to the dataset, and many established deep learning techniques for training and inference. Most of the rule-based hyperparameters are only relevant for three-dimensional (3D) data in the medical domain, like computer tomography (CT) and magnetic resonance imaging (MRI), and are irrelevant to our experiments. For all image modalities except CT, z-score normalization is applied. Each image is normalized independently by subtracting its mean and dividing by its standard deviation. Another dataset parameter is the median shape of the samples. The shape and the available graphics processing unit (GPU) memory determine the patch and batch size and, therefore, the network topology. The patch size is initialized with the median shape and iteratively reduced while adapting the network topology accordingly until the network can be trained with a batch size of at least two given GPU memory constraints.

In addition to the rule-based parameters, there are fixed parameters. These parameters are based on the authors' experience and generalize well across various tasks. The nnU-Net uses a poly learning rate scheduler, a combination of dice coefficient and cross-entropy as the loss function, and stochas-

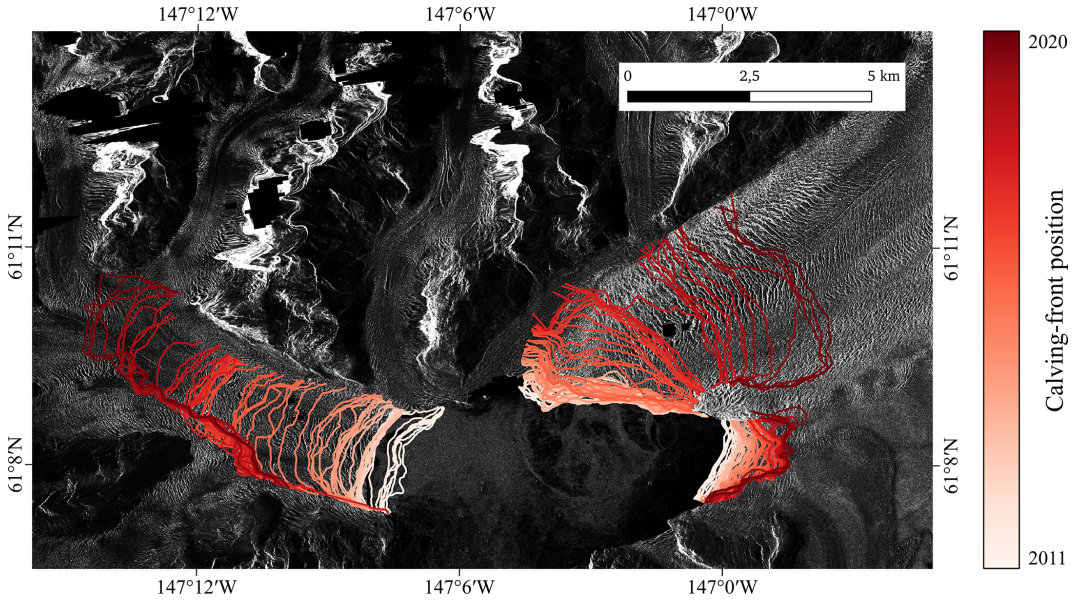


Figure 4. Evolution of the calving-front position of Columbia Glacier (Alaska). The red lines are all calving-front labels available in the dataset. Background: SAR intensity image acquired by TDX on 13 November 2011. Projection CRS EPSG:4326 – WGS 84/UTM zone 6N. SAR imagery was provided by DLR, ESA, and ASF.

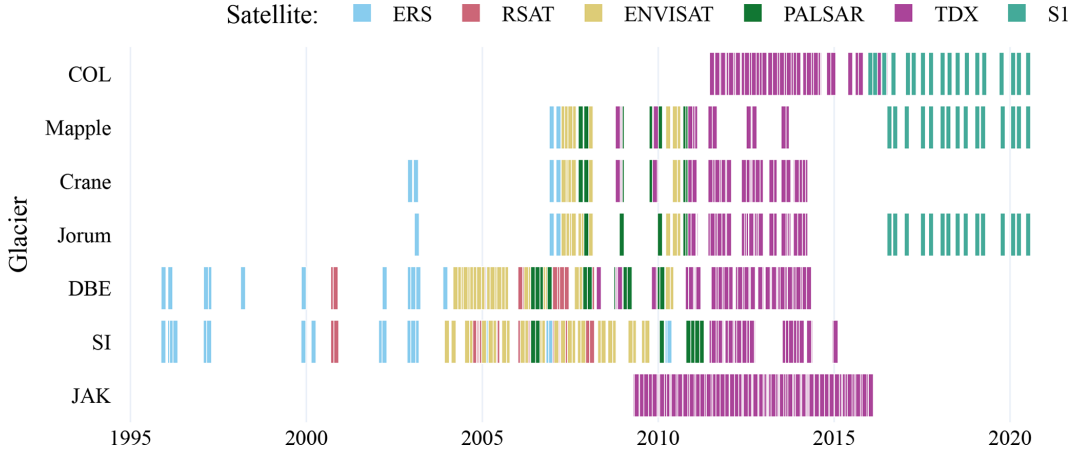


Figure 5. Distribution of the dataset's images over time. The samples are grouped by the seven glaciers and colored according to the capturing satellite. We abbreviated the following glaciers: Columbia (COL), Dinsmoore–Bombardier–Edgeworth (DBE), Sjögren-Inlet (SI), and Jakobshavn Isbrae (JAK).

tic gradient descent (SGD) with a Nesterov momentum as the optimizer. The dice coefficient also helps with the problem of class imbalance. The coefficient of each class is weighted by the number of pixels in the label that relate to the class. Additionally, deep supervision is used to avoid vanishing gradients in neural networks with many layers. Independent of the dataset size, one epoch is defined as 250 mini-batches with foreground oversampling. The foreground oversampling is especially helpful for our application because the class imbalance between the calving front and background is high. It ensures that (at least) one-third of the patches for training are guaranteed to contain a foreground class. In our case, every batch contains two patches, from which one is forced to contain calving-front pixels.

This framework achieves robust results for various medical image segmentation tasks. Overall, nnU-Net sets a new state of the art in 33 out of 53 segmentation tasks of the Kidney and Kidney Tumor Segmentation challenge (Heller et al., 2021) and otherwise shows performances that are on par with or close to the top leader board entries. But the performance of the nnU-Net in segmenting glacier SAR images is yet to be tested.

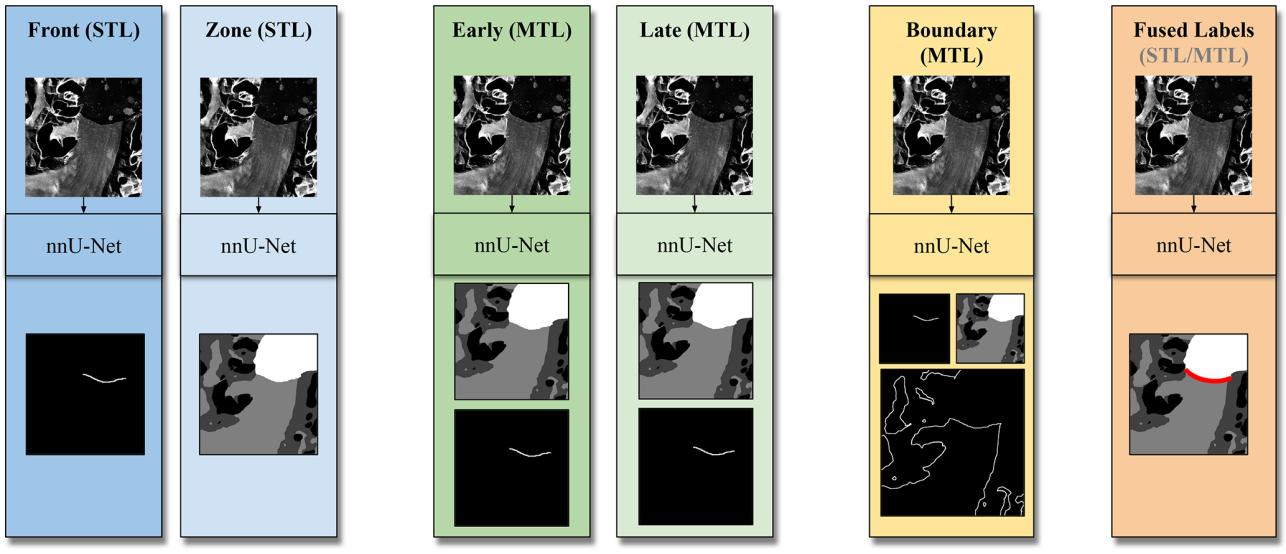


Figure 6. Illustration of the six experiments: two single-task-learning (STL) experiments (blue), two multi-task-learning (MTL) experiments with different architectures (green), one MTL experiment with an additional task of delineating the whole glacier boundary (yellow), and one experiment where the two labels are fused into one segmentation task with an additional class in the zone label (orange).

4.2 Multi-task learning with nnU-Net

The first research goal is to apply the nnU-Net out of the box to the glacier front detection and to the glacier zone segmentation. Training the nnU-Net directly on the front labels is the most straightforward approach for calving-front detection. The nnU-Net is intended to be used in the single-task-learning (STL) manner. In Fig. 6, these two baseline experiments are represented by the two blue columns on the left. The label of the calving front is dilated to the width of five pixels. Our preliminary experiments have shown that dilation makes the predictions more robust. For the training with zone labels, the post-processing includes extracting the boundary between the ocean and the glacier.

In the following experiments, the segmentation problem changes to a multi-task problem, where both labels are used to train one model. The next two experiments concern network architecture. They are represented by the two green columns in Fig. 6. The early-branching architecture uses one decoder for every label. Thus, the number of parameters increases by about 50 %. In contrast, the late-branching architecture requires only a small change to the vanilla U-Net. An additional channel of the last layer is used to predict the second label. For this architecture, only the weights of one kernel are trained in addition to the set of parameters needed for one task. The change in the total number of parameters that need to be trained is negligible.

Because architecture changes with multi-task learning were not foreseen in nnU-Net framework, we had to make changes to the framework. During the experiment's planning, i.e., the pre-processing phase, we fixed the network architecture's estimated size to be the size of the early-branching network so that late- and early-branching networks return the same value for the network size. Otherwise, early- and late-branching networks would be trained on different patch sizes. Thus, performance differences during the evaluation might arise from the different patch sizes and not the differing architectures. During training, the error of all labels is calculated and summed up with equal weighting. For the inference, we adapted only the number of channels in the case of MTL. After, the test samples are divided into patches and fed through the network, and the patch predictions are combined into the prediction of the whole image. The predictions of the zones are post-processed to get an additional result for the position of the calving front. All glacier pixels with a neighboring pixel classified as the ocean are classified as glacier front to extract the glacier front from the zone predictions. We note that the prediction for the other task is omitted in the inference.

The last two experiments concern label changes. In Fig. 6, they are colored yellow and orange. The fifth experiment of this work (see Fig. 6, yellow), extracts the boundaries between the glacier zone and all other zones as a third segmentation task for the late-branching U-Net. The label of the glacier boundaries was extracted from the zone label. All glacier pixels with a neighboring rock or shadow pixel are classified as glacier boundaries. The hypothesis is that providing more information about the same sample benefits the performance of the U-Net in the individual tasks. The third segmentation task is not considered in the final evaluation. The last experiment fuses the zone and front labels by creating a fourth class in the zone label associated with the glacier front: see Fig. 6 (orange). As the front line has a width of five pixels (35–100 m depending on the image resolution), the

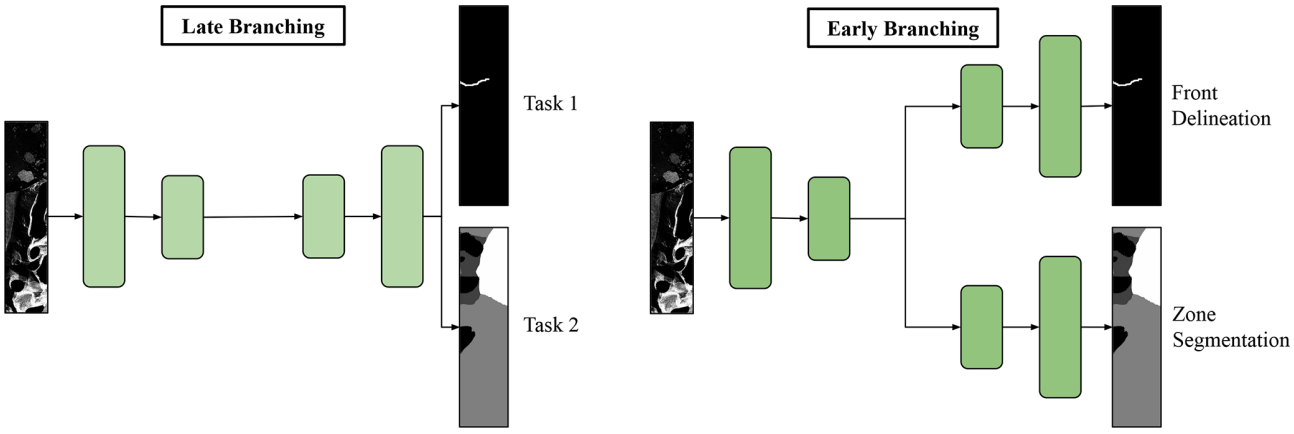


Figure 7. Illustration of early- and late-branching U-Net architectures for multi-task learning (MTL). Both architectures perform joint feature extraction. The early-branching architecture applies separate reconstruction with two decoders, one for each task. The late-branching network applies joint reconstruction with a common decoder.

other zone classes are merely impaired. A post-processing step is added to the predictions. The front pixels are isolated to generate a front prediction for comparing the results with the other experiments. To generate a prediction of the zones, pixels classified as front are assigned to the ocean, and the glacier zone is dilated once with a 7×7 kernel.

4.3 Experimental setup

The nnU-Net was trained on the NVIDIA RTX 3080 with 12 GB memory; the adapted network architecture has nine encoder blocks and eight decoder blocks. Each block consists of two convolutional layers: an instance normalization and a rectified linear unit (ReLU). The kernels of all convolutional layers have a size of 3×3. During training, one batch contains two images. The patch size of the experiments that include only one label is 1280×1024. The experiments that have more than one label have a patch size of 1024×896 because of the GPU memory limit. There are also fixed parameters that are independent of the dataset. This includes the SGD optimizer with an initial learning rate of 0.01, a Nesterov momentum of 0.99, and a weight decay of 3×10^{-5}. Training of one epoch took between 100 and 160 s. The maximum number of epochs of 500 is reached in every training instance (due to limited resources, we reduced the original maximum number of epochs from 1000 to 500). The nnU-Net defines one epoch using a fixed number of iterations (250). In each iteration, the batch is sampled depending on the class distribution of the sample to counteract the class imbalance.

5 Evaluation

In this section, we will examine our evaluation metrics, compare the results of the six proposed experiments, and evaluate the results of the fused-label experiment. We used a 5-fold cross-validation for the evaluation to eliminate the weight initialization bias and the bias caused by a single split into training and validation sets. The metric scores of the individual models are averaged to get a robust measure independent of weight initialization and split.

5.1 Evaluation metrics

We use the mean distance error (MDE) proposed by Gourmelon et al. (2022a) as our main measure. It measures the distance of the predicted front $\mathcal{P}$ to the ground-truth calving-front $\mathcal{Q}$ for all images in the test set $\mathcal{I}$. For every pixel in the label front $\mathcal{Q}$, the distance to the closest pixel in the predicted front $\mathcal{P}$ is determined. Additionally, to make the metric symmetric, the distance to the closest pixel in the label front $\mathcal{Q}$ is determined for every pixel in the predicted front $\mathcal{P}$. These distances are averaged and taken as the mean distance between the two lines (see Fig. 8 and Eq. 1). We note that the front MDE is also calculated for the zone segmentation. We extract the front from the zone segmentation by classifying all glacier pixels with neighboring ocean pixels as the glacier front.

$$\mathrm{MDE}(\mathcal{I}) = \frac{1}{\sum_{(\mathcal{P},\mathcal{Q})\in\mathcal{I}}(|\mathcal{P}|+|\mathcal{Q}|)} \sum_{(\mathcal{P},\mathcal{Q})\in\mathcal{I}} \cdot \left(\sum_{p\in\mathcal{P}} \min_{q\in\mathcal{Q}} ||p-q||_2 + \sum_{q\in\mathcal{Q}} \min_{p\in\mathcal{P}} ||p-q||_2 \right) \quad (1)$$

Additionally, we give classical segmentation metrics to evaluate the zone prediction. They include the Intersection over Union (IoU), which is defined as $\mathrm{IoU} = \frac{T_\mathrm{P}}{T_\mathrm{P}+F_\mathrm{P}+F_\mathrm{N}}$, i.e., the true-positive (T_P) pixels over the sum of T_P, false-positive (F_P), and false-negative (F_N) pixels. Additionally, the F1 score is computed ($\mathrm{F1} = 2 \cdot \frac{\mathrm{pr}\cdot\mathrm{re}}{\mathrm{pr}+\mathrm{re}}$), which is a combination

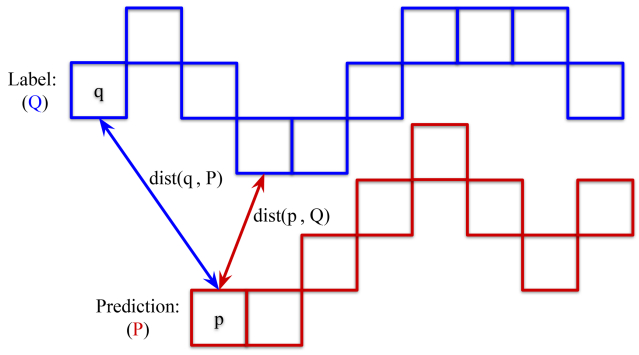


Figure 8. Visualization of the mean distance error (MDE) calculation between front label Q and front prediction P. The dist(q, P) represents the distance between q and the pixel in P that is closest to q.

of recall (re $= \frac{T_P}{T_P+F_N}$) and precision (pr $= \frac{T_P}{T_P+F_P}$). Because this segmentation task is not binary – rather, every sample has four classes – the metrics are calculated for each class individually and then averaged with equal weighting.

5.2 Results and discussion

In this section, we evaluate our experiments. In Fig. 5.2.1, we compare the performance of the six experiments described in Fig. 4.2. In Fig. 5.2.2, we analyze the prediction of the fused-label experiment and investigate the influence of season, glacier site, and satellite on the calving-front prediction performance.

5.2.1 Comparison of the five experiments

In Fig. 9, the MDE of every experiment is compared. The STL approach that is trained on the front labels has an MDE of 1103 ± 72 m, and the STL approach that is trained on the zone labels has an MDE of 1184 ± 225 m. A difference between the STL experiments arises in the performance variance, where the model trained on the zone labels is larger. The number of samples that are incorrectly predicted to have no front pixel is similar, with an average of 27.2 ± 6.0 for training on the front and 24.8 ± 12.4 out of 122 for training with zone labels.

The lowest number of samples with false non-front detections achieves the model that is trained on the fused labels (3.2 ± 2.0 for the front label out of 122 samples). The baseline of Gourmelon et al. is 1 ± 1 based on the zone segmentation and 7 ± 3 from the front prediction. All values can be seen in Table A3. All MTL models have a significantly lower MDE than STL models, with a significance level of $\alpha = 0.01$ using Student's t test. The table with the T values of all experiment pairs is given in Tables A1 and A2. The model that is additionally trained on the glacier boundary has the smallest MDE for both tasks. The MDE baseline of Gourmelon et al. (2022a) is 887 ± 189 m for the front prediction and 753 ± 76 m for the zone. Overall, the MDE is similar for all MTL approaches. Student's t test shows that the differences between the fused label and all other MTL approaches are insignificant ($\alpha = 0.33$).

The metrics for the zone segmentation, shown in Table A4, show a similar trend: an improvement in all MTL approaches compared to the STL approach and minor changes between the MTL approaches. The STL of the nnU-Net on the zone label achieves 62.4 ± 3.5 IoU and an F1 score of 71.7 ± 3.2. The highest F1 score achieves the model that is trained with the additional boundary task with 81.7 ± 0.5. The model's IoU is 72.6 ± 0.4. The baseline of Gourmelon et al. (2022a) is an F1 score of 80.1 ± 0.5 and 69.7 ± 0.6 IoU.

The fused-label approach is the most feasible method, discussed as follows. This approach's performance is on par with the other experiments. Even though the MDE of the boundary experiment is slightly lower than for the fused-label approach, Student's t test showed that the difference is not significant. Moreover, the fused-label approach requires fewer parameters and no changes to the nnU-Net framework, making its use truly out of the box. Additionally, this approach has a relatively small number ($5 \in 122$) of predictions with falsely non-detected fronts.

5.2.2 Analysis of the fused-label experiment

For the final evaluation, the zone segmentations of the five models that are trained during the 5-fold cross-validation on the fused labels are combined into a more robust ensemble prediction. The models differ by their weight initialization and train–validation split. However, they are all trained on the fused labels. The ensemble prediction is created by taking the mean of the probabilities for every class each model gives. We do not use the calving-front pixels directly from the prediction but apply the post-processing of the zone label to gain a slightly better MDE. The post-processing labels the edge of the glacier zone as a calving front that has neighboring ocean pixels. The performance of the ensemble prediction results in an MDE of 515 ± 39 m and $5 \in 122$ predictions with no fronts, which is similar to the average performance of the single models. We did not use the ensemble prediction for the comparison between the experiments in Fig. 5.2.1 to show the variance that is introduced by a different train–validation split and different weight initialization.

The distribution of the MDEs in the test set predictions is plotted in Fig. 10. In the first row, all errors of the prediction of the 117 test samples are drawn as dots. The test set contains two glaciers: Mapple and Columbia. The rows below show the distribution of the MDEs in the two test glaciers separated into summer and winter seasons. The main difference in MDE is between glaciers, similarly to the baseline results of Gourmelon et al. (2022a), with MDEs of 287 ± 48 m for Mapple and 840 ± 48 m for Columbia. The MDE of our methods is, on average, 109 ± 90 m for Mapple, while the MDE of Columbia is 930 ± 1420 m. The difference in the

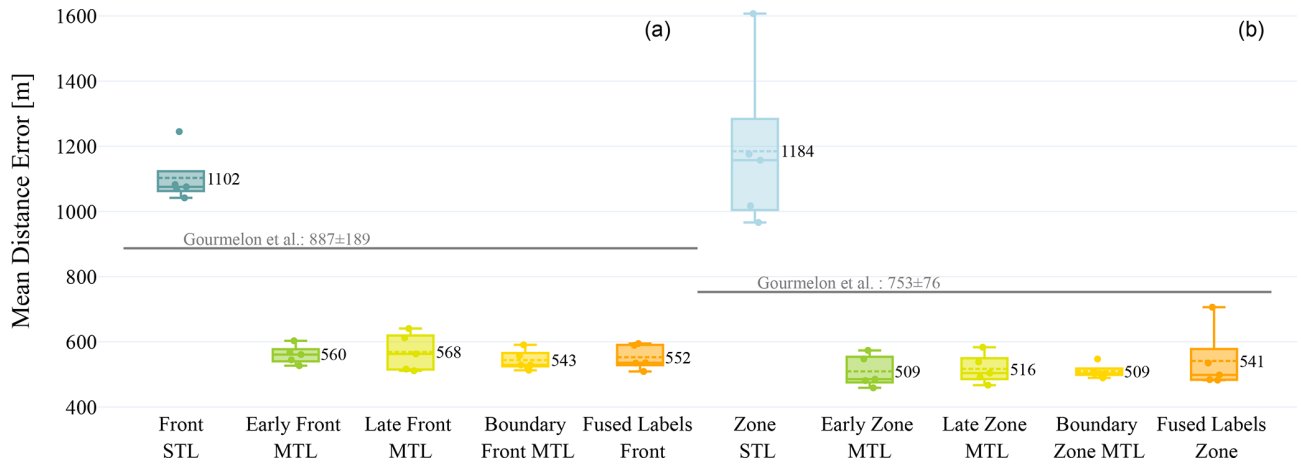


Figure 9. Boxplot of the mean distance error (MDE) of the six experiments. Two columns are colored equally for the multi-task-learning (MTL) experiments. The single-task-learning (STL) experiments are colored in dark and bright blue. Panel **(a)** represents the front delineation based on the front segmentation (task 1), and panel **(b)** represents the front delineation based on the zone segmentation (task 2). The baseline of Gourmelon et al. (2022a) is displayed as a gray line in each half.

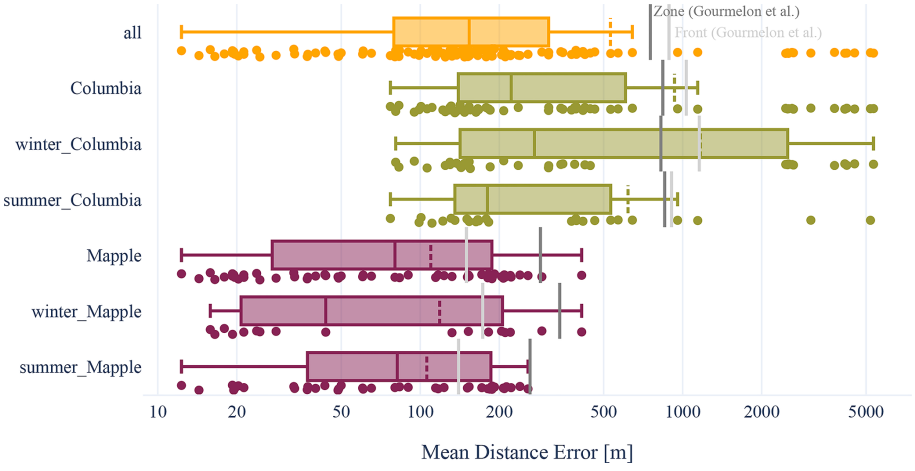


Figure 10. Mean distance error (MDE) of the test set grouped by glaciers and subdivided into seasons. The orange boxplot in the first row shows all MDEs of the test set. The olive-green plot below shows the errors from the Columbia images, subdivided into seasons in the third and fourth rows. The last three rows show errors from the Mapple images. The dark-gray line shows the baseline of the zone prediction, and the light-gray line shows the baseline of the front prediction from Gourmelon et al. (2022b). The ensemble model was trained with the fused zone and front labels. The y axis has a logarithmic scale. The median is the middle line in the rectangle, and the dashed line represents the mean. The x axis has a logarithmic scale. Otherwise, the outliers would dominate the plot. The rectangle reaches from the first quartile to the third quartile.

MDEs is caused by a group of predictions with an error > 1000 m. The median value is 222 m for Columbia and 80 m for Mapple. The reasons for the large glacier differences might be related to the different shapes. Mapple has a simple calving front, a single line constraint by a straight valley. Columbia has multiple calving fronts, a stream coming from the left side of the image, one from the top, and another from the left (see Fig. 13).

There is also a seasonal difference in the MDE. The MDEs of the front prediction during the summer of Mapple and Columbia combined have lower values (307 ± 730 m) than the samples captured during the winter months (818 ± 1389 m). However, the medians are closer together, with 150 m in the summer months and 181 m in the winter months. The MDE baseline (Gourmelon et al., 2022a) in summer is 732 ± 93 and 776 ± 65 m in winter, although most outliers are from winter seasons in Columbia. In winter, snow coverage of the glacier is more likely. The SAR signal can penetrate through several meters of snow cover, depending on the SAR frequency and the water content of the snow cover. However, most of the studied glaciers, in particular Columbia Glacier and the glaciers of the Antarctic Peninsula, are lo-

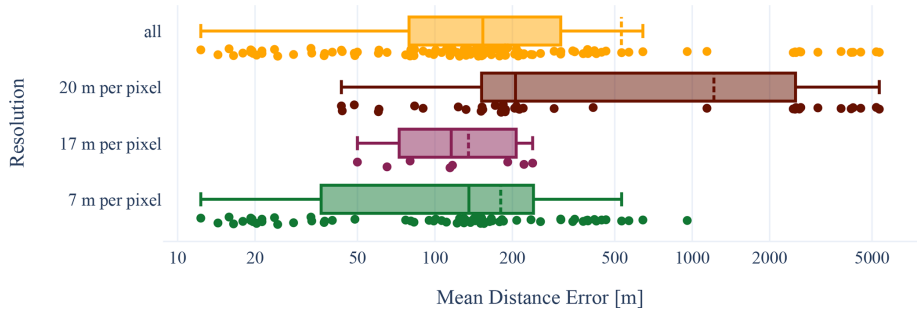


Figure 11. Mean distance error (MDE) of the test set grouped by the resolution of the synthetic aperture radar (SAR) image. The orange boxplot in the first row shows all MDEs of the test set. ENVISAT, ERS-2, and Sentinel-1 have a resolution of 20 m per pixel. MDEs of this resolution are represented by the dark-red boxplot. PALSAR has a 17 m per pixel resolution. MDEs of images of this resolution are represented in the magenta boxplot. TDX has a 7 m per pixel resolution. MDEs of images of this resolution are represented in the green boxplot. The ensemble model was trained with the fused zone and front labels. The dark-gray line shows the baseline of the zone prediction, and the light-gray line shows the baseline of the front prediction from (Gourmelon et al., 2022b). The y axis has a logarithmic scale. The median is the middle line in the rectangle, and the dashed line represents the mean. The x axis has a logarithmic scale. Otherwise, the outliers would dominate the plot. The rectangle reaches from the first quartile to the third quartile.

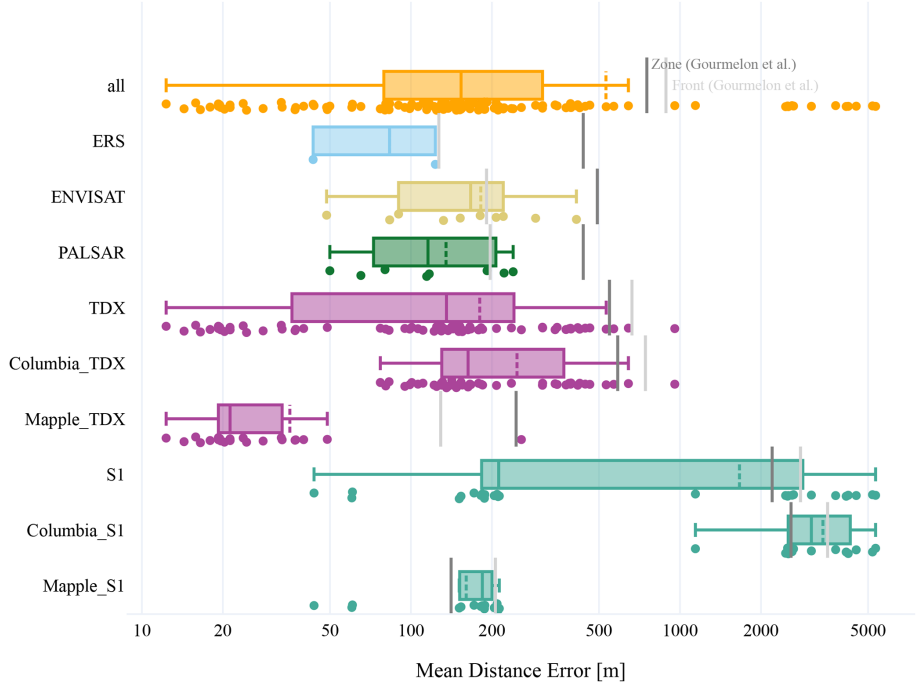


Figure 12. Mean distance error (MDE) of the test set grouped by satellites. The dark-gray line shows the baseline of the zone prediction, and the light-gray line shows the baseline of the front prediction from (Gourmelon et al., 2022b). The images from ERS-2 (light blue), ENVISAT (yellow), and PALSAR (green) only capture the Mapple Glacier. The MDEs from TDX (pink) and Sentinel-1 (turquoise) are subdivided into the two glaciers of the test set. The ensemble model was trained with the fusion of zone and front labels. The y axis has a logarithmic scale. The median is the middle line in the rectangle, and the dashed line represents the mean. The x axis has a logarithmic scale. Otherwise, the outliers would dominate the plot. The rectangle reaches from the first quartile to the third quartile.

cated in temperate marine environments, making wet-snow conditions also likely during winters. Thus, snow can cover useful artifacts like crevasses and rock structures that can be useful for the pattern recognition of the nnU-Net. However, more important is that the ocean next to the glaciers is covered more often by ice mélange during winters, reducing the contrast between ocean and glacier areas. Even for experienced human mappers, it can be a challenging task to distinguish between the glacier tongue and the ice mélange. Thus, we hypothesize that the network has similar issues, leading to reduced performance during winter. The difference between seasonal MDE also depends on the sensor. The MDE of the

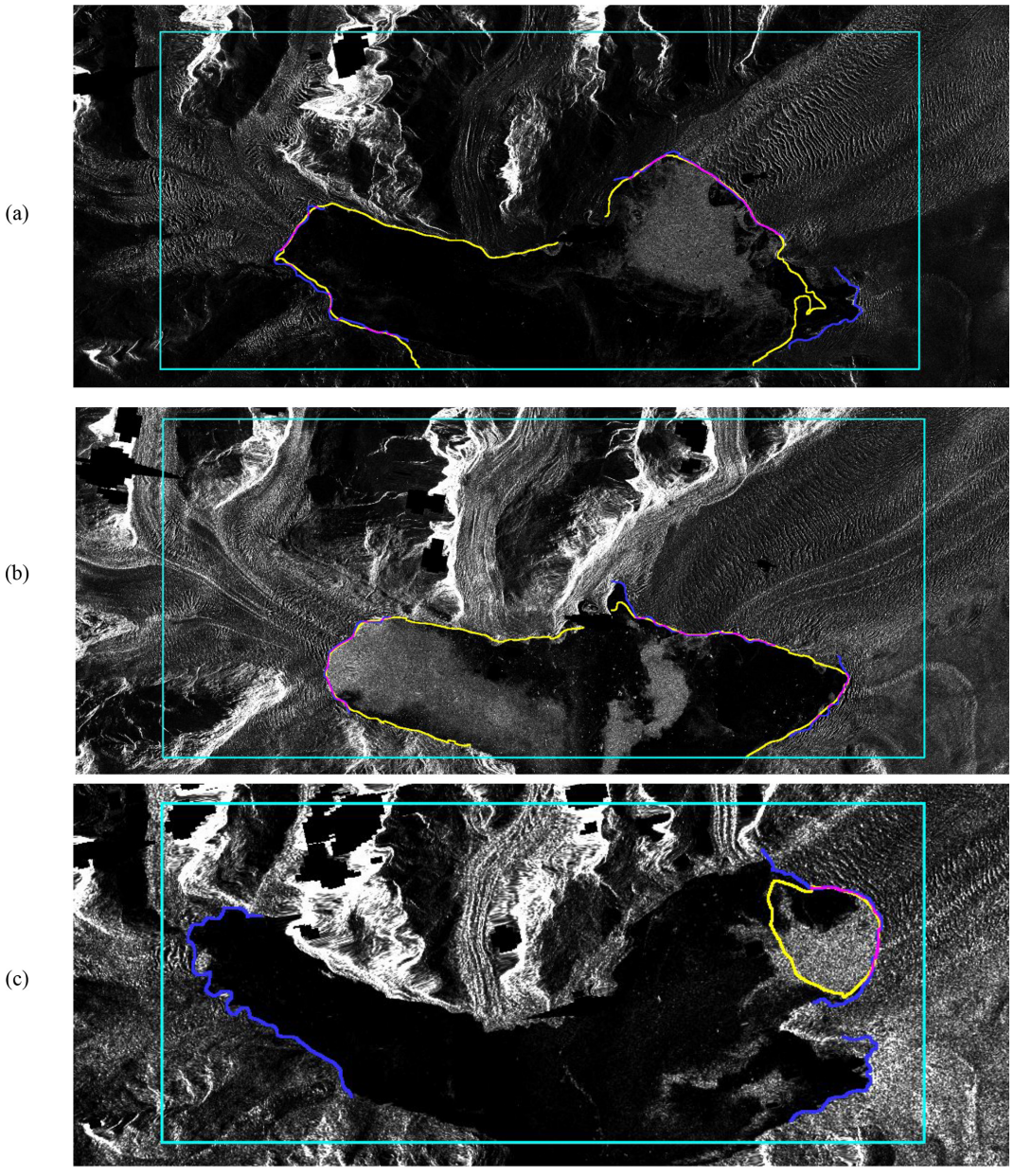


Figure 13. Calving front predictions of Columbia on **(a)** 11 February 2016 and **(b)** 22 June 2014 taken by TDX with 7 m pixel resolution (after multi-looking with a factor of 3×3); **(c)** was taken by Sentinel-1 on 10 September 2019 with 20 m pixel resolution. Note the ground truth (blue), the prediction (yellow), and the overlap of ground truth and prediction (magenta). SAR imagery was provided by DLR, ESA, and ASF.

low-resolution sensors (Sentinel-1 , ENVISAT) is lower in the summer months. The MDE of the high-resolution sensor (TDX) is lower in winter months, but the difference is much smaller. A table of the MDE grouped by satellite and season is in Appendix A5.

An overview of the impact of the sensor resolution is given in Fig. 11. The images with a resolution of 7 m per pixel have a mean MDE of 180 m. The images with 17 m resolution have a mean MDE of 135 m, but this class contains only images of Mapple Glacier taken by PALSAR. The images with 20 m per pixel have a mean MDE of 1214 m. The distribution has a cluster of MDEs that are larger than 1000 m. These large errors only stem from images of Columbia (see Fig. 12, row Columbia_S1).

In Fig. 12, the MDE is grouped by satellites. The predictions for the samples captured by ERS, ENVISAT, and PALSAR have a similar average error between 150 and 300 m. However, the ERS, ENVISAT, and PALSAR samples only represent the Mapple Glacier. TDX captures both test sites and has an MDE of 180 ± 179 m. The samples captured by

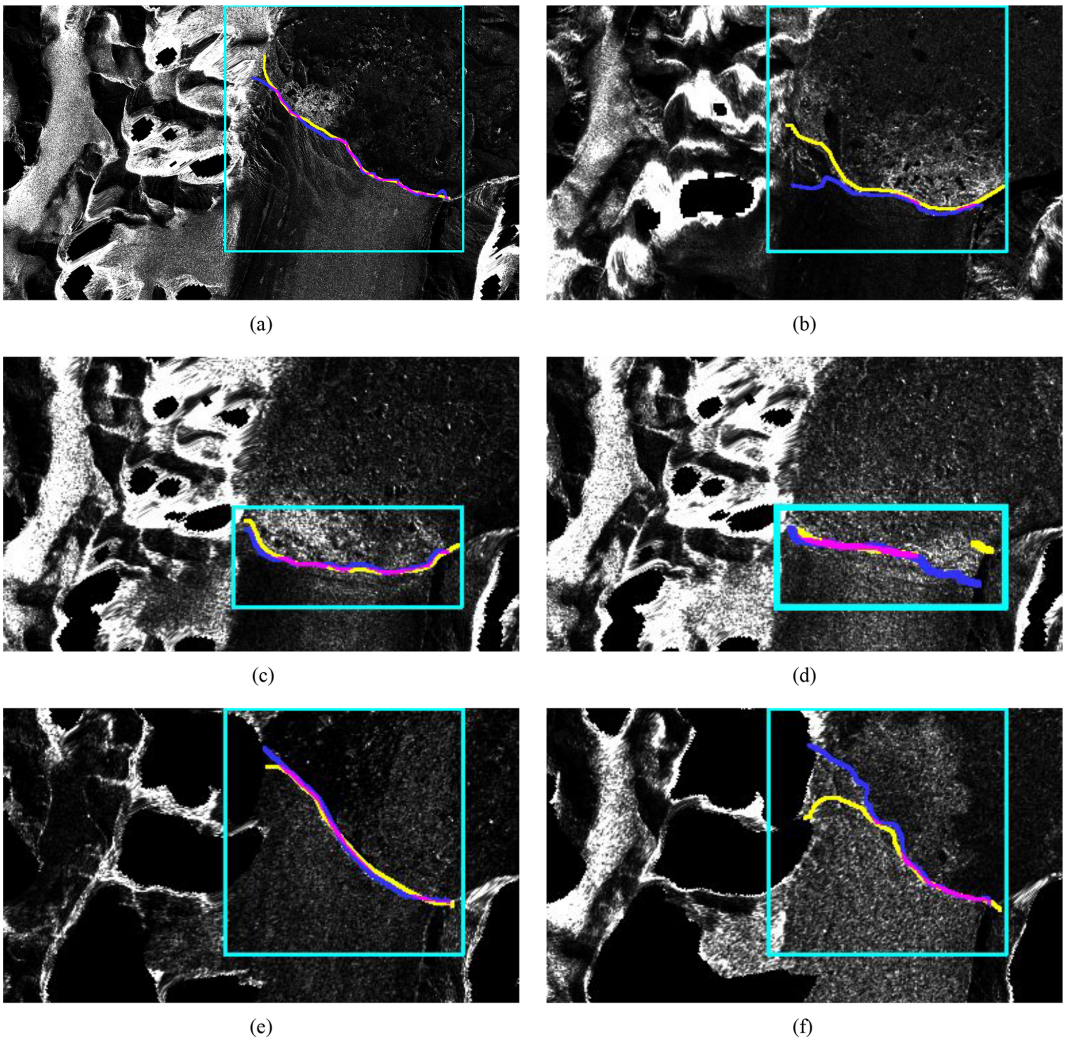


Figure 14. Calving-front predictions of Mapple Glacier. Note the ground truth (blue), the prediction (yellow), the overlap of ground truth and prediction (magenta), and the bounding box (cyan). Panel **(a)** was taken by TDX on 13 October 2008. Panel **(b)** was taken by PALSAR on 29 November 2008. Panel **(c)** was taken by Sentinel-1 on 3 July 2016. Panel **(d)** was taken by Sentinel-1 on 2 March 2019. Panel **(e)** was taken by ERS on 5 February 2007. Panel **(f)** was taken by ENVISAT on 22 September 2007. SAR imagery was provided by DLR, ESA, and ASF.

Sentinel-1 have a higher MDE with 1664 ± 1807 m. The outliers are all from samples of Columbia captured by Sentinel-1. The neural network can not generalize from the training set to this set of samples. The 14 outliers have a front-line delineation error of > 1000 m. These are predictions from images of Columbia taken by Sentinel-1. The low-resolution artifacts, e.g., cracks in the ice, are not visible. Figure 13c shows that the model falsely predicts ocean pixels as the front and does not predict the left calving front at all. The MDE for this sample is 4221 ± 5026 m. We suspect that the complex calving front of Columbia requires a high resolution for accurate detection or more training data from Sentinel-1. The group of outliers heavily increases the overall MDE but can be separated clearly by specifying the satellite and glacier. It shows the generalization limits of our method. For a visual inspection, we provide the predictions of the test set as single images and as an animation on Zenodo (https://doi.org/10.5281/zenodo.8379954, Herrmann, 2023a).

The MDE is negatively influenced by calving-front predictions extending over the labeled calving front on the coastline, which can be seen in Fig. 13a and b. The MDE might give a wrong implication with regard to the usability of these predictions. The prediction of Fig. 13b has a relatively high MDE (393 m) because it predicts the coastline as a calving front. Despite this, the prediction can still be used to monitor the change in the position of the calving front and to calculate the total frontal ablation rate or surface area. The prediction can be corrected easily if the coastline is known. The misclassification of the coastline as a calving front holds true for

nearly all images of Columbia taken by TDX. In the northern part of Columbia, it is particularly difficult to distinguish between the coastline and calving front because the glacier retreated so far that it transitioned from a marine-terminating glacier to a land-terminating glacier. Without prior knowledge, it is even difficult for humans to distinguish between the coastline and the calving front.

The model predictions of the calving front in images of Mapple have a high overlap with the label, even with a low resolution of ≥ 20 m (ERS) (see Fig. 14). There are also some outliers for Mapple, but they are not as severe as the outliers for Columbia. The prediction of Mapple with a high resolution of 7 m per pixel is close to the ground truth (Fig. 14).

6 Conclusions

This work explores the use of the nnU-Net by Isensee et al. (2021) in segmenting glacier-calving fronts. The nnU-Net promises an out-of-the-box application of deep learning for segmentation tasks. We evaluate this claim using a dataset of glacier images provided by Gourmelon et al. (2022b). The dataset contains two tasks: the calving-front detection and the glacier zone segmentation. We try different modifications of multi-task learning (MTL) with two different neural-network architectures to tackle both tasks simultaneously. The results show that combining both tasks increases each task's performance. No significant difference between the two MTL architectures exists. Adding more domain-specific tasks like glacier boundary delineation does not further improve the previous tasks. Due to the small area of the front line, the two labels can be fused into one label. The fusion of labels decreases the number of parameters used, shortens the training time, and reduces the deep learning expertise needed as the nnU-Net can be used without modifications. This approach achieves an average MDE of 541 ± 84 m. We provide the code and the pre-trained model for application on further SAR images of glacier fronts (see "Code and data availability" section). The predictions need to be filtered manually since there can be outliers, as our results show. However, this will provide an initial prediction, which eases the task of glacier front delineation.

To improve the average MDE, future work should focus on reducing the model performance with low-resolution images, such as with Sentinel-1 with 20 m per pixel resolution. This can be done by including more images taken by Sentinel-1 in the training set or by implementing an oversampling strategy of the low-resolution images.

The framework nnU-Net is well suited to segmenting SAR images of glaciers and calving-front delineation. The modification of the nnU-Net for MTL improves the results compared to STL experiments, where only one label is used. However, it is not necessary for glacier segmentation and calving-front delineation because both labels can be fused without losing a significant amount of information. Our findings highlight the suitability of the nnU-net for glacier front segmentation with multi-mission SAR remote sensing data, which will facilitate an efficient, extended spatiotemporal mapping of tidewater glacier terminus changes. Our findings also promote the out-of-the-box application of the nnU-Net for other segmentation tasks based on satellite imagery because we did not need to modify it for the calving-front detection.

Appendix A: Evaluation results of all experiments

This section provides more detailed values of the evaluation. Tables A1 and A2 contain the T values for significant and insignificant differences between models. Table A3 contains the zone MDE and the front MDE of the six experiments. The MDE is averaged over the five model results. The baseline from Gourmelon et al. (2022a) is also provided in the first two rows. The values correspond to Fig. 9. We provide Fig. A6 to compare the two methods (nnU-Net and Gourmelon et al., 2022a). Table A5 contains the MDE of the fused-label training grouped by satellite and season.

Table A1. T value for the significance of different front positions extracted from front predictions. T values that surpass the threshold of 3.36 mean that the probability for the difference to be random is below 1 %. T values below 1 mean that the difference is by chance, with a probability of 35 %, meaning that the difference is insignificant.

	Early MTL	Late MTL	Boundary MTL	Fused labels
STL	15.78	13.46	16.14	14.29
Early MTL	0	0.31	1.02	0.43
Late MTL		0	0.95	0.58
Bound. MTL			0	0.46

Table A2. T value for the significance of different front positions extracted from zone predictions. T values that surpass the threshold of 3.36 mean that the probability for the difference to be random is below 1 %. T values below 1 mean that the difference is by chance, with a probability of 35 %, meaning that the difference is insignificant.

MTL	Early MTL	Late MTL	Boundary labels	Fused
STL	6.57	6.51	6.66	5.96
Early MTL	0	0.29	0.03	0.75
Late MTL		0	0.35	0.58
Bound. MTL			0	0.81

Table A3. MDE of all six experiments. Every experiment setup is trained five times with different weight initializations and train–validation splits and is then averaged. The ∅ columns show the number of images for which no front was falsely detected out of the 122 test samples.

Model	Experiment	Modality	MDE front ↓ [m]	∅ front	MDE zone ↓ [m]	∅ zone
Gourmelon et al. (2022a)	Front	STL	887 ± 189	7 ± 3	–	–
	Zone	STL	–	–	753 ± 76	1 ± 1
nnU-Net	Front	STL	1102 ± 72	27.2 ± 6.0	–	–
	Zone	STL	–	–	1184 ± 255	24.8 ± 12.4
	Early	MTL	560 ± 25	8.6 ± 3.0	509 ± 43	2.2 ± 0.4
	Late	MTL	568 ± 51	8.4 ± 3.5	516 ± 40	4.0 ± 1.1
	Boundary	MTL	543 ± 27	5.6 ± 1.4	509 ± 19	4.4 ± 1.8
	Fused	MTL	552 ± 33	3.2 ± 2.0	541 ± 84	3.4 ± 1.3

Table A4. Segmentation metrics of all six experiments. Every experiment setup is trained five times with different weight initializations and train–validation splits. The metric results of each run are then averaged.

Model	Experiment	Modality	Precision↑	Recall↑	F1↑	IoU↑
Gourmelon et al. (2022a)	Zone	STL	84.2 ± 0.5	79.6 ± 0.9	80.1 ± 0.5	69.7 ± 0.6
nnU-Net	Front	STL	–	–	–	–
	Zone	STL	81.2 ± 2.4	71.7 ± 3.2	71.9 ± 3.7	62.4 ± 3.5
	Early	MTL	87.4 ± 0.4	80.9 ± 0.7	81.6 ± 0.6	72.6 ± 0.9
	Late	MTL	86.4 ± 0.4	79.9 ± 0.6	80.7 ± 0.7	71.1 ± 0.7
	Boundary	MTL	87.5 ± 0.3	80.8 ± 0.4	81.7 ± 0.5	72.6 ± 0.4
	Fused	MTL	87.0 ± 0.2	79.1 ± 1.7	80.7 ± 0.1	70.8 ± 1.8

Table A5. MDE of the fused-label experiments grouped by sensor and season.

Satellite	S1	ENVISAT	ERS	PALSAR	TDX	all
Winter	2529 ± 1719 m	241 ± 101 m	–	–	160 ± 125 m	818 ± 1389 m
Summer	798 ± 1442 m	122 ± 61 m	83 ± 39 m	135 ± 68 m	197 ± 213 m	307 ± 730 m

Table A6. Comparison of the U-Net training setup and hyperparameters of Gourmelon et al. (2022a) and the nnU-Net (Isensee et al., 2021).

Hyperparameter	Gourmelon et al. (2022a)	nnU-Net (Isensee et al., 2021)
Number of convolutional layers	10	34
Pooling	Max pooling	Strided convolution
Activation function	ReLU	Leaky ReLU
Patch size	256×256	1280×1024
Batch size	16	2
Optimizer	Cyclic learning rate scheduler (Smith, 2017) in combination with the Adam optimizer (Kingma and Ba, 2014)	SGD with a Nesterov momentum of 0.99 and a weight decay of 3×10^{-5}
Initial learning rate	4×10^{-5}	1×10^{-3}
Gradient clipping	True	False
Deep supervision	False	True
Loss function	Combination of dice and cross-entropy	Combination of dice and cross-entropy

Appendix B: Calving-front labels

The other six glacier sites in this section are displayed complementarily to Fig. 4. The corresponding calving fronts of the different time steps are colored with a gradient from bright for the past to dark red for the most recent fronts.

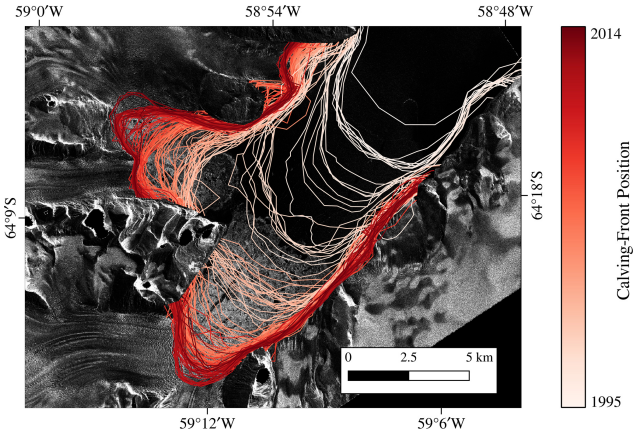


Figure B1. Jakobshavn Isbrae Glacier from 1995 to 2014. SAR imagery was provided by DLR, ESA, and ASF.

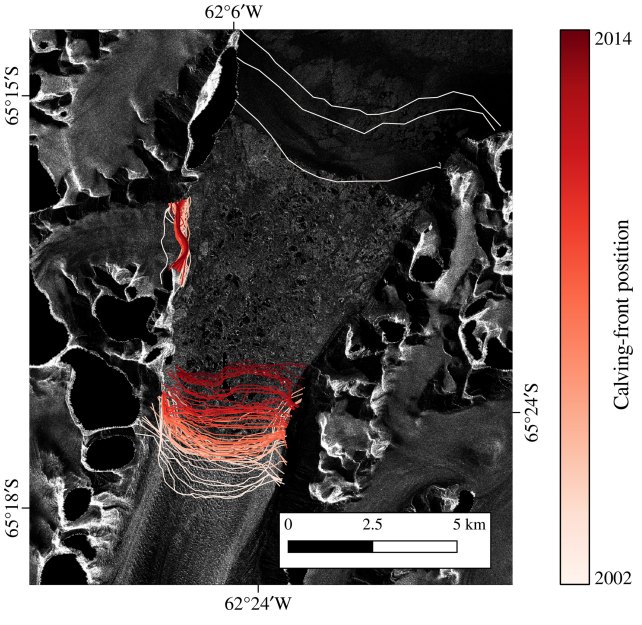


Figure B2. Mapple Glacier from 2006 to 2020. SAR imagery was provided by DLR, ESA, and ASF.

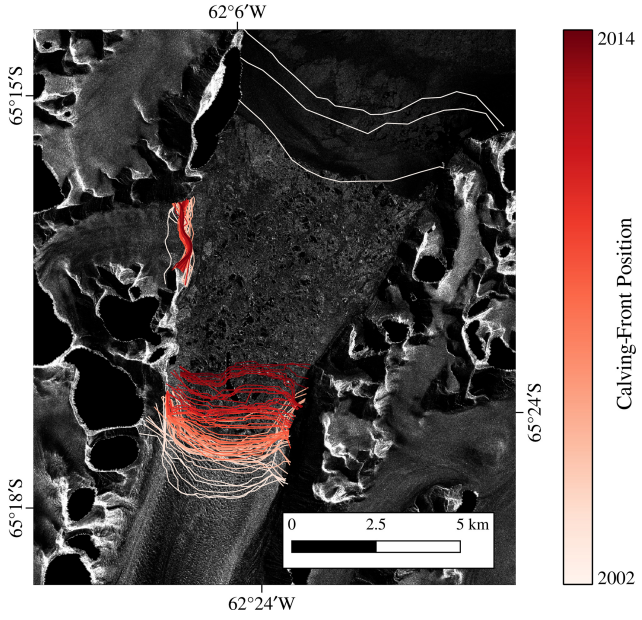


Figure B3. Crane Glacier from 2002 to 2014. SAR imagery was provided by DLR, ESA, and ASF.

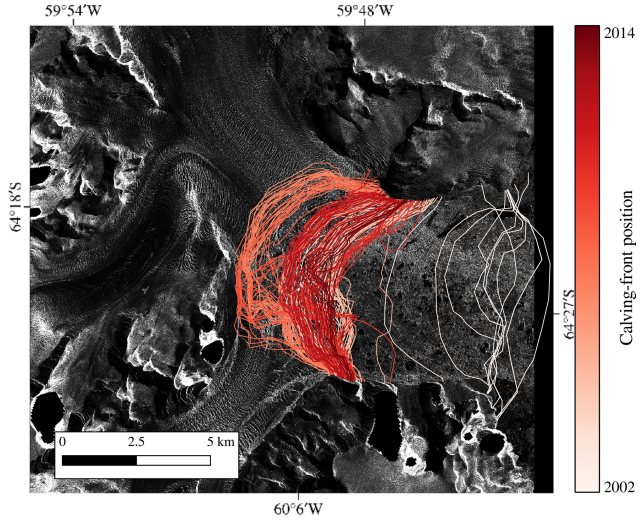


Figure B4. Dinsmoore–Bombardier–Edgeworth glaciers from 1995 to 2014. SAR imagery was provided by DLR, ESA, and ASF.

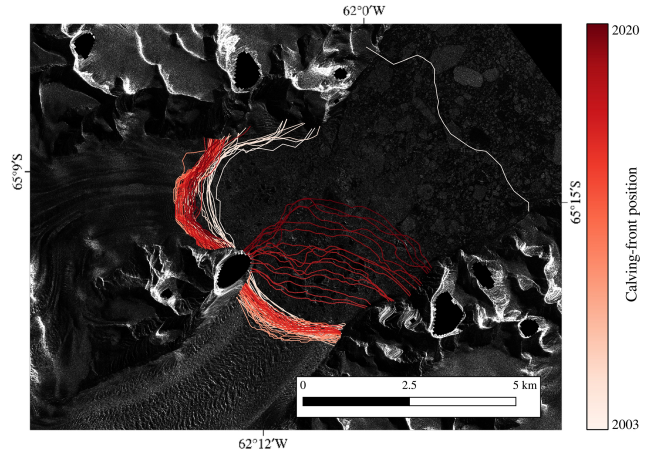


Figure B5. Jorum Glacier from 2003 to 2020. SAR imagery was provided by DLR, ESA, and ASF.

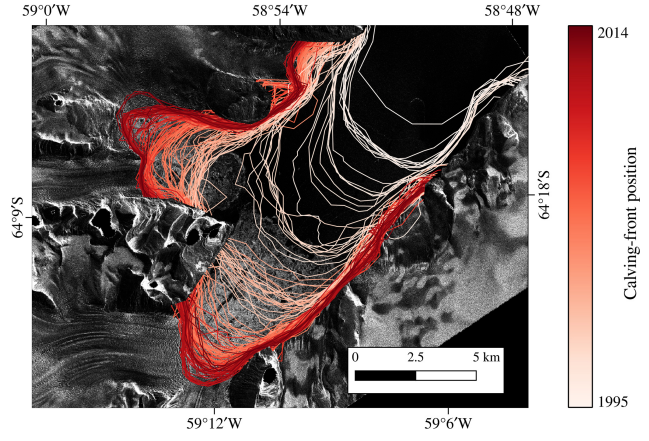


Figure B6. Sjögren-Inlet Glacier from 1995 to 2014. SAR imagery was provided by DLR, ESA, and ASF.

Appendix C: Temporal distribution of zone label

Figure C1 shows the temporal distribution of the zone label. Jacobshavn and Columbia, the two glaciers in the Northern Hemisphere, show an increase in ocean area with a decreasing glacier area. In particular, the set of images of the glaciers on the Antarctic Peninsula has samples with large areas with no available information. Low partial coverage by the radar swath causes prominent peaks of the no-information-available class. Jacobshavn's area distribution shows a repetitive structure of increasing and decreasing glacier area. This pattern represents seasonal changes.

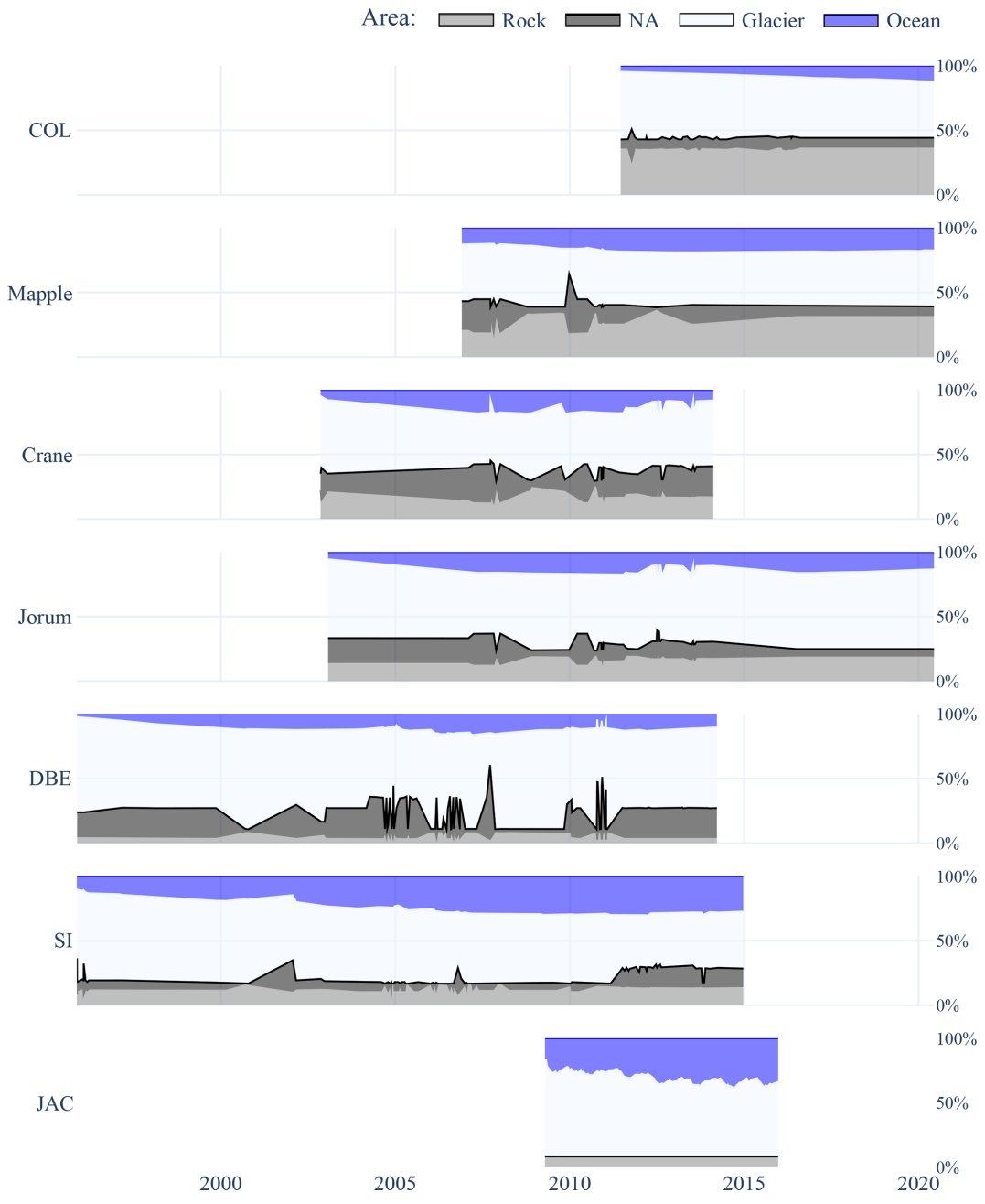


Figure C1. Changes of the class distributions in the zone labels of each glacier over time. The ocean is colored in blue, the glacier area is colored in white, the rock is colored in gray, and the area with no available information (NA) is colored in black. There is no radar reflection in the NA area due to terrain elevation causing shadows or due to limited coverage by the radar swath.

Code and data availability. The code is fundamentally based on the nnU-Net by Isensee et al. (2021), which is also publicly available (https://github.com/MIC-DKFZ/nnUNet/tree/nnunetv1, Isensee, 2019). We modified the neural-network architecture and the corresponding pre- and post-processing. The modified version of the nnU-Net and our evaluation and visualization scripts are available on Zenodo (https://doi.org/10.5281/zenodo.10168770, Herrmann and Gourmelon, 2023), and a demo version is provided on Hugging Face (https://doi.org/10.5281/zenodo.10169965, Herrmann, 2023c). The calving-front predictions of the test set are available on Zenodo (https://doi.org/10.5281/zenodo.8379954, Herrmann, 2023a), as well as the pre-trained model (https://doi.org/10.5281/zenodo.7837300, Herrmann, 2023b). The dataset is provided by Gourmelon et al. (2022b) (https://doi.org/10.1594/PANGAEA.940950).

Author contributions. OH and NG were responsible for conceptualizing the study. OH developed and implemented the methodology and software for the experiments. NG and TS provided the dataset and a benchmark score. NG and VC supervised the study and reviewed and edited the paper. JJF acquired financial support. OH created visualizations and prepared the paper with contributions from all the co-authors. JJF, VC, and NG reviewed and edited the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *The Cryosphere*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We acknowledge the free provision of the SAR data via various proposals from ESA, ASF, and DLR.

Financial support. This research has been supported by the European Research Council H2020 European Research Council (grant no. 948290), the Staedtler Stiftung (Emerging Field Initiative (EFI), TAPE), the Bayerisches Staatsministerium für Wissenschaft und Kunst (Measuring and Modelling Mountain Glaciers in a Changing Climate (IDP M^3OCCA)), the Deutsche Forschungsgemeinschaft (National High-Performance Computing Center (NHR@FAU)), and the Universitätsbund Erlangen-Nürnberg (Open Access Publication Funding).

Review statement. This paper was edited by Stef Lhermitte and reviewed by two anonymous referees.

References

Abolvardi, A. A., Hamey, L., and Ho-Shon, K.: UNET-Based Multi-Task Architecture for Brain Lesion Segmentation, in: Digital Image Computing: Techniques and Applications (DICTA), 1–7, https://doi.org/10.1109/DICTA51227.2020.9363397, 2020.

Amundson, J. M., Fahnestock, M., Truffer, M., Brown, J., Lüthi, M. P., and Motyka, R. J.: Ice mélange dynamics and implications for terminus stability, Jakobshavn Isbrse, Greenland, J. Geophys. Res.-Earth Surf., 115, F01005, https://doi.org/10.1029/2009JF001405, 2010.

Amyar, A., Modzelewski, R., Li, H., and Ruan, S.: Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation, Comput. Biol. Med., 126, 104037, https://doi.org/10.1016/j.compbiomed.2020.104037, 2020.

Baumhoer, C. A., Dietz, A. J., Dech, S., and Kuenzer, C.: Remote sensing of antarctic glacier and ice-shelf front dynamics-a review, Remote Sens., 10, 1445, https://doi.org/10.3390/rs10091445, 2018.

Baumhoer, C. A., Dietz, A. J., Kneisel, C., and Kuenzer, C.: Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning, Remote Sens., 11, 2529, https://doi.org/10.3390/rs11212529, 2019.

Baumhoer, C. A., Dietz, A. J., Kneisel, C., Paeth, H., and Kuenzer, C.: Environmental drivers of circum-Antarctic glacier and ice shelf front retreat over the last two decades, The Cryosphere, 15, 2357–2381, https://doi.org/10.5194/tc-15-2357-2021, 2021.

Baumhoer, C. A., Dietz, A. J., Heidler, K., and Kuenzer, C.: IceLines – A new data set of Antarctic ice shelf front positions, Sci. Data, 10, 138, https://doi.org/10.1038/s41597-023-02045-x, 2023.

Beer, C., Biebow, N., Braun, M., Döring, N., Gaedicke, C., Gutt, J., Hagen, W., Hauck, J., Heinemann, G., Herata, H., Holfort, J., Jung, T., Kassens, H., Klenzendorf, S., Läufer, A., Lohmann, G., Nixdorf, U., Plass, S., Quillfeldt, P., Rhein, M., Rachold, V., Riedel, A., Sachs, T., and Wendisch, M.: Forschungsagenda Polarregionen im Wandel, 79, Bundesministerium für Bildung und Forschung (BMBF), Germany, 2021.

Bischke, B., Helber, P., Folz, J., Borth, D., and Dengel, A.: Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks, in: International Conference on Image Processing (ICIP), 1480–1484, IEEE, Taipei, ISBN 978-1-5386-6249-6, https://doi.org/10.1109/ICIP.2019.8803050, 2019.

Carr, J. R., Stokes, C., and Vieli, A.: Recent retreat of major outlet glaciers on Novaya Zemlya, Russian Arctic, influenced by fjord geometry and sea-ice conditions, J. Glaciol., 60, 155–170, https://doi.org/10.3189/2014JoG13J122, 2014.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, in: European conference on computer vision (ECCV), edited by: Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., 833–851, Springer International Publishing, Cham, ISBN 978-3-030-01234-2, https://doi.org/10.1007/978-3-030-01234-2_49, 2018.

Chen, S., Bortsova, G., García-Uceda Juárez, A., van Tulder, G., and de Bruijne, M.: Multi-task Attention-Based Semi-supervised Learning for Medical Image Segmentation, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, edited by: Shen, D., Liu, T., Peters, T. M., Staib, L. H.,

Essert, C., Zhou, S., Yap, P.-T., and Khan, A., Lecture Notes in Computer Science, 457–465, Springer International Publishing, Cham, ISBN 978-3-030-32248-9, https://doi.org/10.1007/978-3-030-32248-9_51, 2019.

Cheng, D., Hayes, W., Larour, E., Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E.: Calving Front Machine (CALFIN): glacial termini dataset and automated deep learning extraction method for Greenland, 1972–2019, The Cryosphere, 15, 1663–1675, https://doi.org/10.5194/tc-15-1663-2021, 2021.

Cook, A. J., Murray, T., Luckman, A., Vaughan, D. G., and Barrand, N. E.: A new 100-m Digital Elevation Model of the Antarctic Peninsula derived from ASTER Global DEM: methods and accuracy assessment, Earth Syst. Sci. Data, 4, 129–142, https://doi.org/10.5194/essd-4-129-2012, 2012.

Davari, A., Baller, C., Seehaus, T., Braun, M., Maier, A., and Christlein, V.: Pixel-wise Distance Regression for Glacier Calving Front Detection and Segmentation, IEEE T. Geosci. Remote, 60, 1–10, https://doi.org/10.1109/TGRS.2022.3158591, 2022.

Friedl, P., Seehaus, T. C., Wendt, A., Braun, M. H., and Höppner, K.: Recent dynamic changes on Fleming Glacier after the disintegration of Wordie Ice Shelf, Antarctic Peninsula, The Cryosphere, 12, 1347–1365, https://doi.org/10.5194/tc-12-1347-2018, 2018.

Gourmelon, N., Seehaus, T., Braun, M., Maier, A., and Christlein, V.: Calving fronts and where to find them: a benchmark dataset and methodology for automatic glacier calving front extraction from synthetic aperture radar imagery, Earth Syst. Sci. Data, 14, 4287–4313, https://doi.org/10.5194/essd-14-4287-2022, 2022a.

Gourmelon, N., Seehaus, T., Braun, M. H., Maier, A., and Christlein, V.: CaFFe (CAlving Fronts and where to Find thEm: a benchmark dataset and methodology for automatic glacier calving front extraction from sar imagery), PANGAEA [data set], https://doi.org/10.1594/PANGAEA.940950, 2022b.

Hartmann, A., Davari, A., Seehaus, T., Braun, M., Maier, A., and Christlein, V.: Bayesian U-Net for Segmenting Glaciers in Sar Imagery, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 41, 3479–3482, https://doi.org/10.1109/IGARSS47720.2021.9554292, iSSN: 2153-7003, 2021.

He, K., Lian, C., Zhang, B., Zhang, X., Cao, X., Nie, D., Gao, Y., Zhang, J., and Shen, D.: HF-UNet: Learning Hierarchically Inter-Task Relevance in Multi-Task U-Net for Accurate Prostate Segmentation in CT Images, IEEE T. Med. Imaging, 40, 2118–2128, https://doi.org/10.1109/TMI.2021.3072956, 2021.

Heidler, K., Mou, L., Baumhoer, C., Dietz, A., and Zhu, X. X.: HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline, IEEE T. Geosci. Remote, 60, 1–14, https://doi.org/10.1109/TGRS.2021.3064606, 2021.

Heller, N., Isensee, F., Maier-Hein, K. H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y., Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., Zhong, C., Ma, J., Rickman, J., Dean, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Kaluzniak, H., Raza, S., Rosenberg, J., Moore, K., Walczak, E., Rengel, Z., Edgerton, Z., Vasdev, R., Peterson, M., McSweeney, S., Peterson, S., Kalapara, A., Sathianathen, N., Papanikolopoulos, N., and Weight, C.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge, Med. Image Anal., 67, 101821, https://doi.org/10.1016/j.media.2020.101821, 2021.

Herrmann, O.: Out-of-the-box calving front detection method using deep learning (Version 3), Zenodo [data set], https://doi.org/10.5281/zenodo.8379954, 2023a.

Herrmann, O.: Pretrained_nnUNet_calvingfront_detection, Zenodo [code], https://doi.org/10.5281/zenodo.7837300, 2023b.

Oskar Herrmann: nnUNet_calvingfront_detection, Zenodo [code], https://doi.org/10.5281/zenodo.10169965, 2023c.

Herrmann, O. and Gourmelon, N.: nnUNet_calvingfront_detection, Zenodo [code], https://doi.org/10.5281/zenodo.10168770, 2023.

Isensee, F.: nnU-Ne, GitHub [code], https://github.com/MIC-DKFZ/nnUNet/tree/nnunetv1 (last access: 21 November 23), 2019.

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods, 18, 203–211, https://doi.org/10.1038/s41592-020-01008-z, 2021.

Jang, H.-J. and Cho, K.-O.: Applications of deep learning for the analysis of medical data, Arch. Pharm. Res., 42, 492–504, https://doi.org/10.1007/s12272-019-01162-9, 2019.

Kholiavchenko, M., Sirazitdinov, I., Kubrak, K., Badrutdinova, R., Kuleev, R., Yuan, Y., Vrtovec, T., and Ibragimov, B.: Contour-aware multi-label chest X-ray organ segmentation, Int. J. Comput. Ass. Rad., 15, 425–436, https://doi.org/10.1007/s11548-019-02115-9, 2020.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint, arXiv:1412.6980, 2014.

Kneib-Walter, A., Lüthi, M. P., Moreau, L., and Vieli, A.: Drivers of Recurring Seasonal Cycle of Glacier Calving Styles and Patterns, Front. Earth Sci., 9, 667717, https://doi.org/10.3389/feart.2021.667717, 2021.

Li, X., Wang, Y., Tang, Q., Fan, Z., and Yu, J.: Dual U-Net for the Segmentation of Overlapping Glioma Nuclei, IEEE Access, 7, 84040–84052, https://doi.org/10.1109/ACCESS.2019.2924744, 2019.

Loebel, E., Scheinert, M., Horwath, M., Heidler, K., Christmann, J., Phan, L. D., Humbert, A., and Zhu, X. X.: Extracting glacier calving fronts by deep learning: the benefit of multi-spectral, topographic and textural input features, IEEE T. Geosci. Remote, 60, 1–12, https://doi.org/10.1109/TGRS.2022.3208454, 2022.

Marochov, M., Stokes, C. R., and Carbonneau, P. E.: Image classification of marine-terminating outlet glaciers in Greenland using deep learning methods, The Cryosphere, 15, 5041–5059, https://doi.org/10.5194/tc-15-5041-2021, 2021.

McNabb, R. W., Hock, R., and Huss, M.: Variations in Alaska tidewater glacier frontal ablation, 1985–2013, J. Geophys. Res.-Earth Surf., 120, 120–136, https://doi.org/10.1002/2014JF003276, 2015.

Minowa, M., Schaefer, M., Sugiyama, S., Sakakibara, D., and Skvarca, P.: Frontal ablation and mass loss of the Patagonian icefields, Earth Planet. Sc. Lett., 561, 116811, https://doi.org/10.1016/j.epsl.2021.116811, 2021.

Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E.: Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study, Remote Sens., 11, 74, https://doi.org/10.3390/rs11010074, 2019.

Periyasamy, M., Davari, A., Seehaus, T., Braun, M., Maier, A., and Christlein, V.: How to Get the Most Out of U-Net for Glacier Calving Front Segmentation, IEEE J. Sel. Top. Appl. Earth Obs.,

15, 1712–1723, https://doi.org/10.1109/JSTARS.2022.3148033, 2022.

Recinos, B., Maussion, F., Rothenpieler, T., and Marzeion, B.: Impact of frontal ablation on the ice thickness estimation of marine-terminating glaciers in Alaska, The Cryosphere, 13, 2657–2672, https://doi.org/10.5194/tc-13-2657-2019, 2019.

Recinos, B., Maussion, F., Noël, B., Möller, M., and Marzeion, B.: Calibration of a frontal ablation parameterisation applied to Greenland's peripheral calving glaciers, J. Glaciol., 67, 1177–1189, https://doi.org/10.1017/jog.2021.63, 2021.

Robel, A. A., Schoof, C., and Tziperman, E.: Persistence and variability of ice-stream grounding lines on retrograde bed slopes, The Cryosphere, 10, 1883–1896, https://doi.org/10.5194/tc-10-1883-2016, 2016.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by: Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., 9351, 234–241, Springer International Publishing, Cham, ISBN 978-3-319-24573-7, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rott, H., Wuite, J., Rydt, J. D., Gudmundsson, G. H., Floricioiu, D., and Rack, W.: Impact of marine processes on flow dynamics of northern Antarctic Peninsula outlet glaciers, Nat. Commun., 11, 2969, https://doi.org/10.1038/s41467-020-16658-y, 2020.

Shepherd, A., Ivins, E., Rignot, E., Smith, B., Broeke, M. V. D., Velicogna, I., Whitehouse, P., Briggs, K., Joughin, I., Krinner, G., Nowicki, S., Payne, T., Scambos, T., Schlegel, N., Geruo, A., Agosta, C., Ahlstrøm, A., Babonis, G., Barletta, V., Blazquez, A., Bonin, J., Csatho, B., Cullather, R., Felikson, D., Fettweis, X., Forsberg, R., Gallee, H., Gardner, A., Gilbert, L., Groh, A., Gunter, B., Hanna, E., Harig, C., Helm, V., Horvath, A., Horwath, M., Khan, S., Kjeldsen, K. K., Konrad, H., Langen, P., Lecavalier, B., Loomis, B., Luthcke, S., McMillan, M., Melini, D., Mernild, S., Mohajerani, Y., Moore, P., Mouginot, J., Moyano, G., Muir, A., Nagler, T., Nield, G., Nilsson, J., Noel, B., Otosaka, I., Pattle, M. E., Peltier, W. R., Pie, N., Rietbroek, R., Rott, H., Sandberg-Sørensen, L., Sasgen, I., Save, H., Scheuchl, B., Schrama, E., Schröder, L., Seo, K. W., Simonsen, S., Slater, T., Spada, G., Sutterley, T., Talpe, M., Tarasov, L., Berg, W. J. V. D., Wal, W. V. D., Wessem, M. V., Vishwakarma, B. D., Wiese, D., and Wouters, B.: Mass balance of the Antarctic Ice Sheet from 1992 to 2017, Nature, 558, 219–222, https://doi.org/10.1038/s41586-018-0179-y, 2018.

Smith, L. N.: Cyclical Learning Rates for Training Neural Networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 24–31 March 2017, Sanra Rosa, CA, USA, 464–472, https://doi.org/10.1109/WACV.2017.58, 2017.

Straneo, F., Heimbach, P., Sergienko, O., Hamilton, G., Catania, G., Griffies, S., Hallberg, R., Jenkins, A., Joughin, I., Motyka, R., Pfeffer, W. T., Price, S. F., Rignot, E., Scambos, T., Truffer, M., and Vieli, A.: Challenges to Understanding the Dynamic Response of Greenland's Marine Terminating Glaciers to Oceanic and Atmospheric Forcing, B. Am. Meteorol. Soc., 94, 1131–1144, https://doi.org/10.1175/BAMS-D-12-00100.1, 2013.

Zhang, E., Liu, L., and Huang, L.: Automatically delineating the calving front of Jakobshavn Isbræ from multitemporal TerraSAR-X images: a deep learning approach, The Cryosphere, 13, 1729–1741, https://doi.org/10.5194/tc-13-1729-2019, 2019.

Zhang, E., Liu, L., Huang, L., and Ng, K. S.: An automated, generalized, deep-learning-based method for delineating the calving fronts of Greenland glaciers from multi-sensor remote sensing imagery, Remote Sens. Environ., 254, 112265, https://doi.org/10.1016/j.rse.2020.112265, 2021.