https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

The Cryosphere Discussions
Open Access
EGU

Calving Front Machine (CALFIN): Glacial Termini Dataset and Automated Deep Learning Extraction Method for Greenland, 1972-2019

Daniel Cheng[1], Wayne Hayes[1], Eric Larour[2], Yara Mohajerani[1,3], Michael Wood[2], Isabella Velicogna[1,2], and Eric Rignot[1,2]

[1]University of California at Irvine, Irvine CA, USA
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena CA, USA
[3]University of Washington, eScience Institute and Department of Civil and Environmental Engineering, Seattle, WA, 98195, USA

Correspondence: Daniel Cheng (dlcheng@uci.edu)

Abstract. We present Calving Front Machine (CALFIN), an automated method for extracting calving fronts from satellite images of marine-terminating glaciers. The results use Landsat imagery from 1972 to 2019 to generate 22,678 calving front lines across 66 Greenlandic glaciers. The method uses deep learning, and builds on existing work by Mohajerani et al., Zhang et al., and Baumhoer et al. Additional post-processing techniques allow for accurate segmentation of imagery into Shapefile outputs. This method is uniquely robust to the impact of clouds, illumination differences, ice mélange, and Landsat-7 Scan Line Corrector errors. CALFIN provides improvements on the current state of the art. A model inter-comparison is performed to evaluate performance against existing methodologies. CALFIN's ability to generalize to SAR imagery is also evaluated. CALFIN's fronts are often indistinguishable from manually-curated fronts, deviating by 2.25 pixels (86.76 meters) from the true front on a diverse set of 162 testing images. The current implementation offers a new opportunity to explore sub-seasonal trends on the extent of Greenland's margins, and supplies new constraints for simulations of the evolution of the mass balance of the Greenland Ice Sheet and its contributions to future sea level rise.

1 Introduction

The evolution of Greenland's tidewater glaciers is an important constraint on the evolution of the Greenland Ice Sheet (Nick et al., 2013). Likewise, changes in Greenland are important in tracking and predicting future sea level rise over the next century (Andersen et al., 2015; Fürst et al., 2015; van den Broeke et al., 2016). Constraining Greenland's glacial evolution is thus an important part of improving the understanding of the earth system as a whole. One constraint on glacial evolution is the position of glacial calving fronts and ice margins over time (King et al., 2018). Currently, most calving front delineation is done with time-consuming manual labor (Carr et al., 2017; Bunce et al., 2018; Catania et al., 2018). This results in the under-utilization of available satellite imagery, and causes gaps in seasonal records that introduce uncertainty when modeling past and projected climate change (Catania et al., 2020). Significant efforts have been made to improve this situation, which include the ESA-CCI dataset of 26 Greenlandic glaciers from 1990-2016, the PROMICE dataset of 47 glaciers from 1990-2018, and the MEaSUREs

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

dataset of 200+ glaciers from 2000-2017 (ENVEO, 2017; Andersen et al., 2019; Joughin et al., 2015). Yet the increasing availability of new datasets through missions like Landsat 8 and the release of old datasets through improved reprocessing call for new automated ways of detecting calving front delineations. In particular there is a a strong need for these automated ways to be robust, specifically against cloud cover, ice mélange, and shadows. Previous techniques such as edge detection and texture analysis have significant challenges with respect to these issues (Malik et al., 2001; Seale et al., 2011; Paravolidakis et al., 2016). Modern machine learning techniques and deep neural networks provide a robust, scalable, and accurate solution to these processing challenges.

In this study, Sect. 2 covers the data source along with the spatial and temporal coverage. Sect. 3 examines the CALFIN algorithm and method for processing the data. Sect. 4.1 validates the algorithm through error analysis. Sect. 5 and Sect. 6 shows as well as discusses the results - the calving front dataset and algorithm.

2 Data Source and Scope

Several potential data sources are first evaluated for use, including Terra/MODIS, TerraSAR-X, Landsat, and Sentinel (see Table 1). Landsat is selected for its long time-series availability and reasonable spatial distribution/resolution.

Table 1. Potential Data Sources: A comparison of selected data sources available for use. We initially focus on a single source for the study.

Potential Data Sources					
Name	Resolution(s)	Time Series	Repeat Cycle	Sensor	Seasonal Coverage
Landsat	30 m , 60 m	1972-present	16 day	Optical	Spring-Fall
Terra (MODIS)	250 m , 500 m , 1000 m	1999-present	1, 8, 16 day	Optical	Spring-Fall
Sentinel	10 m , 20 m , 60 m	2014-present	10, 12 day	SAR	Spring-Winter
TerraSAR-X	1 m , 3 m , 6 m	2007-present	3-11 day	SAR	Spring-Winter

Here the area of interest is restricted to Greenland, in particular the calving fronts for 66 Greenlandic basins shown in Fig. 1, spanning the 1972 to 2019 time period shown in Fig. 2. The basins are selected for their high drainage volume and wide spatial distribution.

3 Methodology

3.1 Preprocessing

The method begins with an automated pipeline (see Fig. 3) that prepares the raw data for processing with the neural network.

The first step is to collect all the input raster images centered around one of 9 primary glacial basins. These basins include Kong Oscar, Hayes, Rink Isbrae, Upernavik, Jakobshavn, Kangiata Nunaata, Helheim, Kangerlussuaq, and Petermann. Next, all L1TP (precision and terrain corrected) rasters from Landsats 1-8 with low cloud coverage (<20%) are collected. A few L1GS/L1GT (non-corrected) products are also selected, which are manually georeferenced, and used to fill in Landsat 1-2 time

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

The Cryosphere Discussions
Open Access
EGU

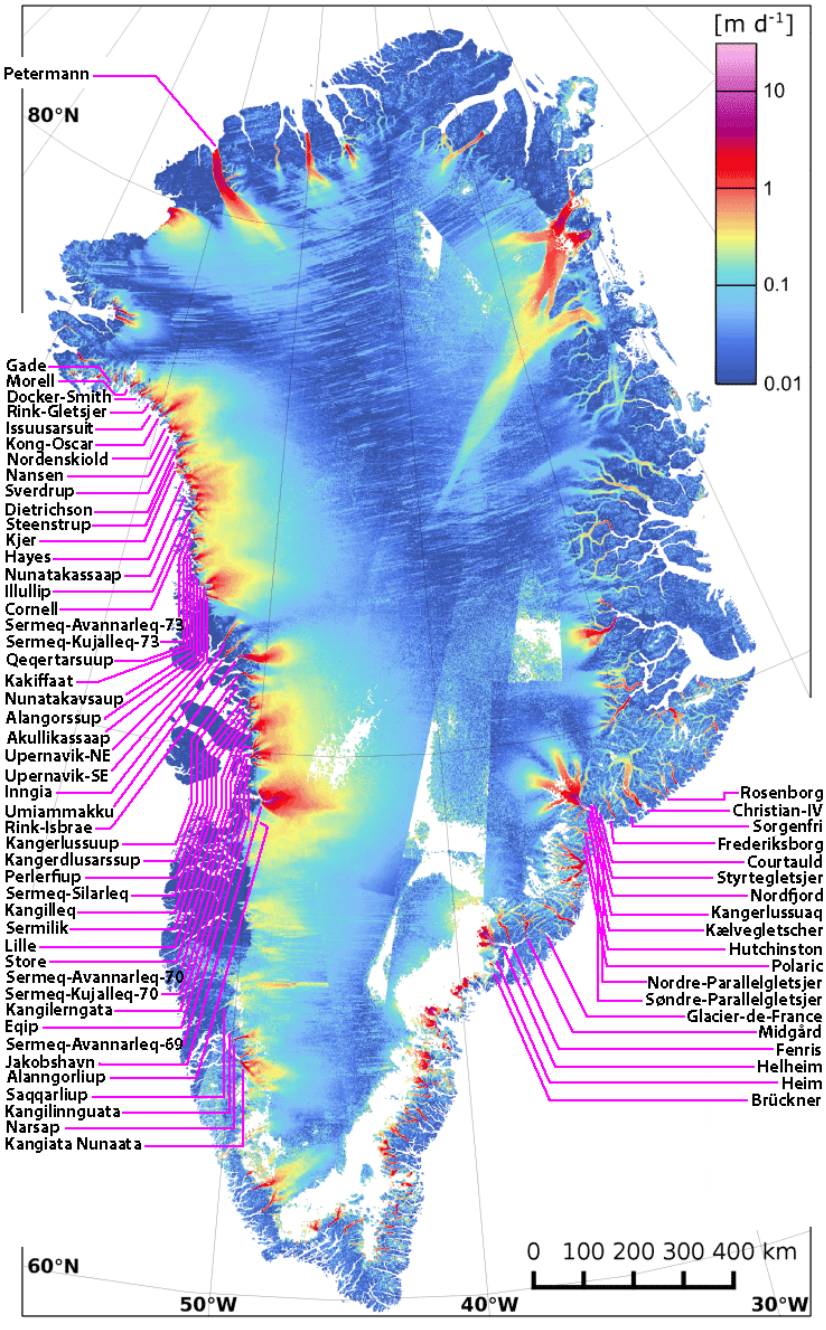

Figure 1. Spatial Coverage Map: Spatial distribution of 66 selected Greenlandic glaciers. The velocity map is taken from Nagler et al. (2015).

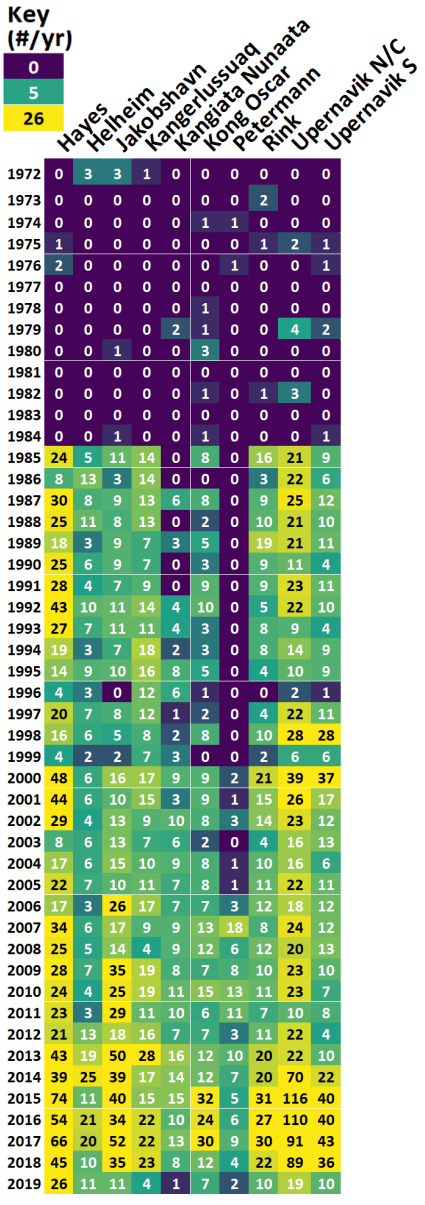

Figure 2. Temporal Coverage Map: Number of fronts per year from 1972-2019 for 10 high drainage volume basins. For the full temporal coverage map, see attached Supplement, Fig. S1.

series gaps (1972-1985). This produces a total of 4956 Landsat rasters. Next, predefined basin domain Shapefiles that enclose the terminus are used to clip the Landsat raster subsets. Additional filtering removes subsets that still contain ≥30% NODATA

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

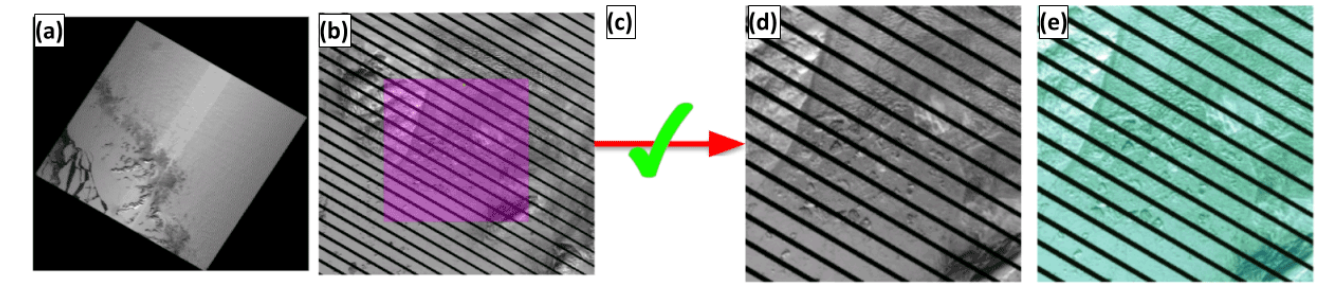

Figure 3. Preprocessing Pipeline: (a) First, input the raw Landsat GeoTIFF rasters with <20% clouds. (b) Next, subset using QGIS/GDAL and the domain Shapefile to clip each raster. (c) Then, filter the clouded/NODATA subsets. (d) Now, resize the subsets to 256x256 px. (e) Finally, enhance contrast and stack with the raw subset.

pixels or ≥20% cloud pixels, as subsets that exceed these thresholds are not likely to contain detectable fronts. At this stage, 20188 GeoTIFF subsets are accumulated. Each subset is then resized to 256x256 px, and lastly enhanced using Pseudo-HDR Toning (HDR) and Shadows/Highlights (S/H) through Adobe Photoshop. The raw, HDR, and S/H enhanced subsets are then stacked into a single RGB image. At this point, the images are ready for processing into calving front masks.

3.2 Neural Network Processing

Images are processed using the Calving Front Machine Neural Network (CALFIN-NN), as illustrated in Fig. 4. Neural networks like CALFIN-NN work by learning patterns in training data, and finding them in new data. CALFIN-NN is trained using manually delineated calving front masks, discussed in Sect. 6.1. Once trained, CALFIN-NN outputs a probability mask that shows each pixel's likelihood of lying on the coastline/calving front. CALFIN-NN also generates a land/ice-ocean probability mask as a secondary output. Once each image is processed, the calving front is ready to be extracted during post-processing.

3.2.1 Network Architecture & Modifications

Neural networks are the foundation of several automated delineation methods, including Zhang et al. (2019), Mohajerani et al. (2019), and Baumhoer et al. (2019). This method builds upon this work, and uses a modification of the DeepLabV3+ Xception neural network from Chen et al. (2018), as shown in Fig. 4. The first half, the encoder, uses the Xception-65 network to extract image features (Chollet, 2017). It does this by assembling basic features, like edges and corners, into more abstract features, such as glacier/land textures. The second half of the network, the decoder, takes the output of the encoder and up-samples the features to predict the final probability mask outputs.

Several modifications are made to the DeepLabV3+ Xception network. To accurately recognize line-like features such as calving fronts, additional Atrous Spatial Pyramidal Pooling (ASPP) blocks are added in between the encoder and decoder, with the dilation scales 0, 1, 2, 3, 4, and 5. The number of Middle Blocks (MB in Fig. 4) is reduced from 16 to 8, as the extra discriminative power from those blocks is not needed. To facilitate faster training and performance, the input size is reduced from 512 px to 224 px. Since the input resolution is reduced, the encoder is also modified to remove several down-sampling

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

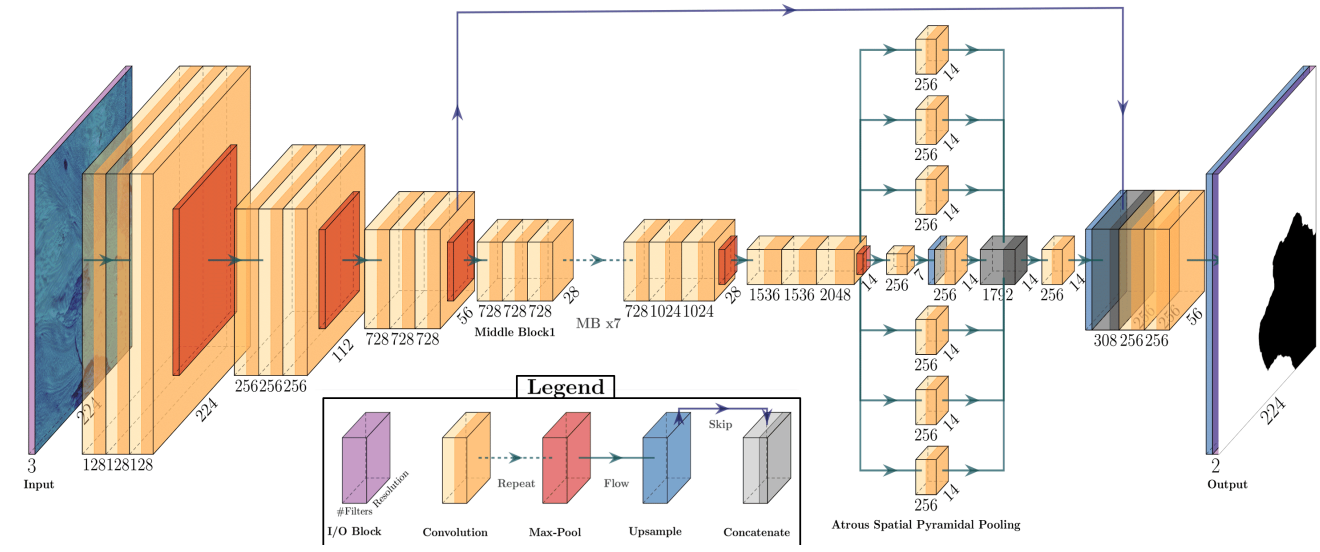

Figure 4. The CALFIN-NN Processing Architecture: Each orange "Xception" block consists of convolution kernels that detect features in the previous block. Blocks are reduced in size periodically to pool increasingly complex and numerous feature maps. "U" shaped connections help refine the probability masks during up-sampling. Note that the 7 repeated "Xception" blocks in the middle section are omitted for brevity.

"max-pool" layers. The last contribution adds a 2-channel output to the decoder, allowing for both calving front mask and ice/ocean masking. Together, these changes reduce the size of the network from 40M parameters to 29M parameters while also increasing the overall accuracy.

3.3 Post-Processing

At this stage, the 2-channel pixel mask output of CALFIN-NN is post-processed to extract useful Shapefile data products as shown in Fig. 5.

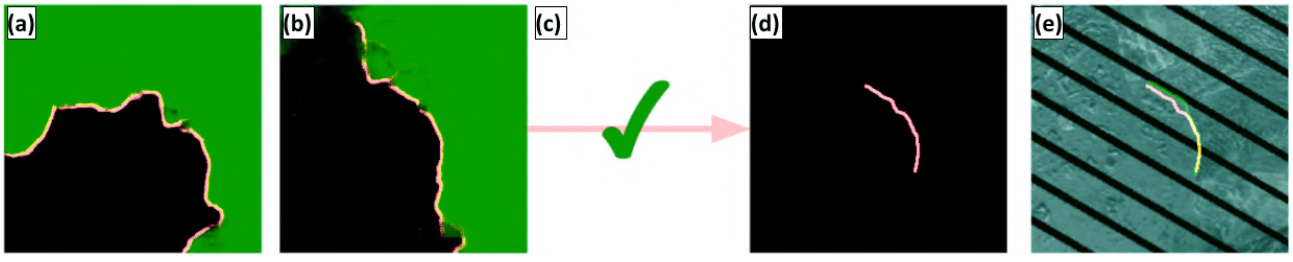

Figure 5. Postprocessing Pipeline: (a) First, get the processed image from CALFIN-NN. (b) Then, isolate and re-process each front. (c) Next, filter unconfident predictions. (d) Now, fit line and mask static coastline (see also Fig. 6). (e) Lastly, export and validate the Shapefile.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

3.3.1 Calving Front Reprocessing

Individual fronts are first isolated from the processed image and reprocessed as zoomed-in subsets of the input image wherever they are detected. The front detection method is described in Sect. 3.3.3. The nature of CALFIN-NN's output as a confidence measure is also exploited, so that generated fronts can be filtered out based on how confidently detected each one is.

3.3.2 Pixel Mask to Coastal Polyline

Next, a polyline is fit to the pixel mask to retrieve the correct coastline boundary. This is performed by converting each pixel in the mask to nodes in a graph, connecting the nearest neighboring nodes, then finding the single longest path in the graph's minimum spanning tree (MST) (Kruskal, 1956). This polyline not only corresponds with the coastline edge, but also out-performs other contour finding algorithms by eliminating noise, errors, and gaps inherited from previous steps. A visual example is given in Fig. 6a-d.

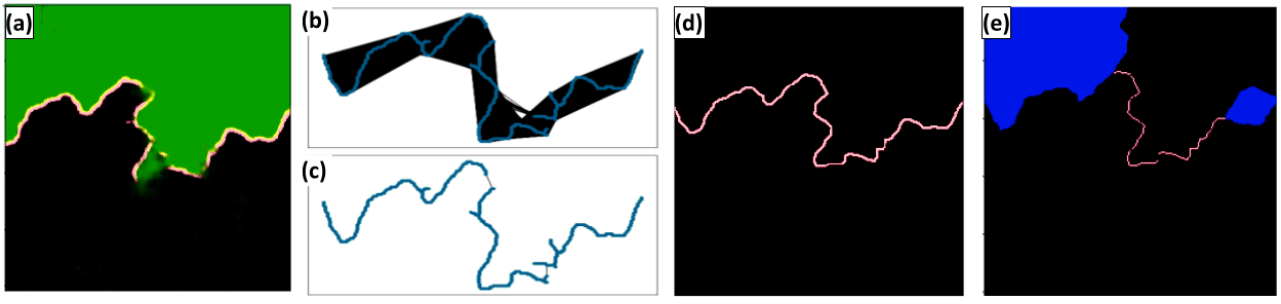

Figure 6. Mask to Polyline Algorithm: (a) First, extract the red coastline mask from the CALFIN-NN output. (b) Then create a graph, connecting each pixel (blue) to 15% of its nearest neighbors with an edge (black). (c) Next, create an MST from the graph. (d) Now, extract the longest path from the MST. (e) Finally, mask the static coastline using the fjord boundaries (blue) to extract the calving front.

3.3.3 Coastline to Calving Front

Next, the calving front is isolated from the coastline polyline. Fjord boundary masks are first created for each basin. By calculating the distance from each point in the coastline to the nearest fjord boundary pixel, then selecting the contiguous pixels which are the farthest from the fjord boundaries, the calving front can be isolated. The result of this is shown in Fig 6e.

3.3.4 Calving Front to Shapefile

The last step is to export the polyline and corresponding polygon as geo-referenced Shapefiles. First, the polyline is smoothed to eliminate noise artifacts inherited from previous steps. Next, the smoothed polylines, fjord boundary mask, and land-ice/ocean masks are combined to create a polygonal ocean mask. Optionally, manual verification each output with the original GeoTIFF subset can be performed. This was done for all cases in this study to ensure the validity of the automated pipeline. This constrains the mean distance error to be <100 m, as covered in the following section.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

The Cryosphere Discussions
Open Access
EGU

4 Error Analysis and Quality Assessment

Two methods are used to evaluate CALFIN. For the primary method, error is estimated by calculating the Mean/Median Distance between predicted and manually delineated fronts (see Sect. 4.1). For the secondary method, the classification accuracy is calculated with the Intersection over Union metric (see Sect. 4.2). Additionally, the detection accuracy is evaluated, and the associated confusion matrix is provided (see Sect. 4.4). These metrics are evaluated on several validation sets, taken from existing studies as discussed in Sect. 6.2. These validation sets contain data that is excluded during model training. This prevents the models from memorizing data and skewing the accuracy.

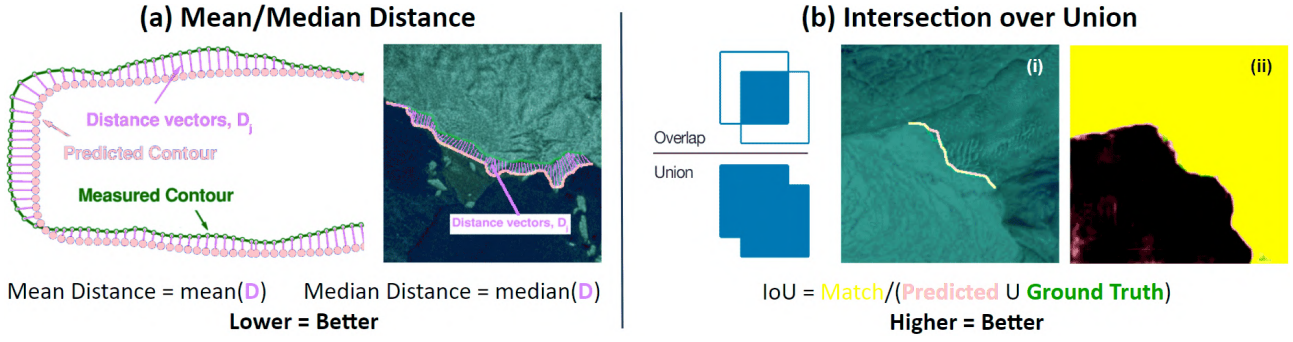

Figure 7. Error Measures: (a) A visual outline of Mean/Median Distance Error Estimation and (b) Classification Accuracy using Intersection over Union (IoU) for (i) the primary calving front, and (ii) the secondary ice/ocean mask, respectively.

4.1 Error Estimation

The primary quality assessment method is the Mean Distance Error (Mohajerani et al., 2019; Zhang et al., 2019; Baumhoer et al., 2019). Conceptually, this method resembles the numerical integration of the area between two curves, normalized by the average length of the curves (see Fig. 7a). Also referred to as the Area over Front (A/F) in literature, this method can also be seen as a generalization of the method of transects along arbitrarily oriented fronts (Mohajerani et al., 2019; Baumhoer et al., 2019). This metric is implemented by taking the mean/median of the distances between closest pixels in the predicted and manually delineated fronts. Note that pixel distance is biased to be inversely proportional to a network's input size, so the error in meters is also provided in the following analysis.

4.2 Classification Accuracy

The secondary quality assessment method calculates the Intersection over Union (IoU) (Baumhoer et al., 2019). This metric evaluates the degree of overlap between the predicted and ground truth masks of the calving front. It is calculated by dividing the number of pixels in the intersection of two masks over the number of pixels in the union of the two masks (see Fig. 7b). When calculating the IoU of 3 pixel wide edges, this measure is very strict: 1 pixel of difference results in a score of 0.5000, and scores in that range or above are indicative of human levels of accuracy. When calculating the IoU of land-ice/ocean-mélange masks, this measure is less strict, and scores in the range of 0.9000 and above are indicative of human levels of accuracy.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

4.3 Validation Results

The following subsections list tables that print the above metrics for the associated validation sets, the values from the original studies, and a subset of the outputs of CALFIN-NN on each. The primary validation set, the CALFIN validation set (CALFIN-VS), consists of 162 images with clouds, illumination differences, ice mélange, and Landsat 7 Scanline Corrector Errors (L7SCEs). The CALFIN-VS contains data from 62 Greenlandic basins, including Helheim, which was specifically excluded from CALFIN's training set for validation purposes - as done by Mohajerani et al. (2019). The CALFIN-VS ensures CALFIN-NN produces consistent results on new data, addressing concerns raised by Zhang et al. (2019) Sect. 7.3. The two validation subsets, CALFIN-VS-L7-only/none, are also evaluated so as to isolate and exclude images with L7SCEs, respectively. To allow for comparisons between studies, CALFIN-NN's performance metrics on previous studies' validation sets are also shown, where appropriate. The sets include the 10 Landsat Helheim subsets used in Mohajerani et al. (2019) (M-VS), the 6 TerraSAR-X Jakobshavn subsets used in Zhang et al. (2019) (Z-VS), and 62 Sentinel-1 Antarctic basins taken from the 11 validation scenes used in Baumhoer et al. (2019) (B-VS). Note that the error metrics are still sensitive to how each study implements them, which are nevertheless reproduced and documented for the purpose of comparison. These concerns are also addressed in the comprehensive inter-model comparison, discussed in Sect. 6.4.

4.3.1 CALFIN Validation Set

CALFIN-NN performs well on the CALFIN-VS. The true mean distance error of the CALFIN dataset is calculated to be 86.76 $\pm$ 1.43 m with 95% confidence. When including only images with L7SCEs (CALFIN-VS-L7-only), the error is 2.22 px (91.93 m), showcasing CALFIN-NN's unique robustness to L7SCEs. When excluding outliers such as Kong Oscar, the median distance error is only 1.21 px (44.59 m). For full outputs, see Supplement Fig. S4-S7.

Validation Set	Model	Mean Distance	Median Distance	IoU Calving Front	IoU Ice/Ocean
CALFIN-VS	CALFIN-NN	2.25 px, 86.76 m	1.21 px, 44.59 m	0.4884	0.9793
CALFIN-VS-L7-none	CALFIN-NN	2.27 px, 81.65 m	1.16 px, 44.01 m	0.4880	0.9819
CALFIN-VS-L7-only	CALFIN-NN	2.22 px, 91.93 m	1.33 px, 49.24 m	0.4888	0.9766

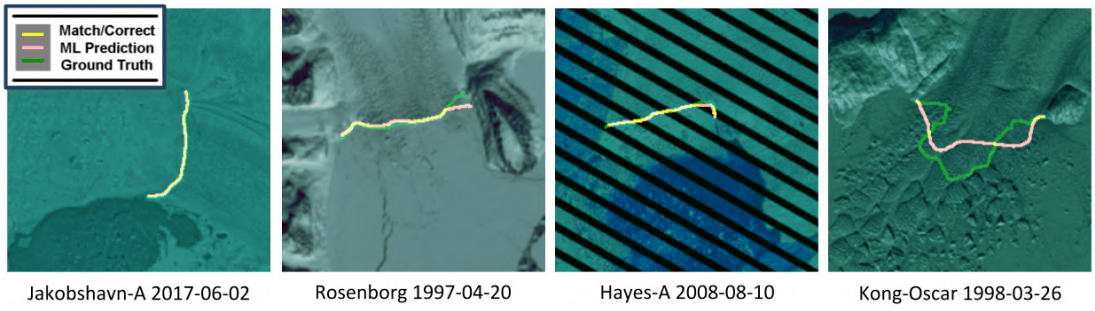

Figure 8. CALFIN-VS Validation Output Results: Yellow represents human (green) and machine (red) agreement on the front location. Note that the drop in mean pixel distance despite the increase in mean meter distance (and vice versa) comes from L7SCE images being reprocessed at lower sizes due to detection failures (see Fig. 5c), and pixel error bias being inversely related to input size (see Sect. 4.1).

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

The Cryosphere Discussions
Open Access
EGU

4.3.2 Mohajerani et al. (2019) Validation Set

CALFIN-NN performs well on the M-VS. This demonstrates the generalization capability of CALFIN-NN, which improves upon the Mohajerani et al. (2019) neural network (M-NN). Also note that M-NN implements distances errors differently and omits ice/ocean masks from the evaluation. This is further explored in the model inter-comparison discussed in Sect. 6.4.

Validation Set	Model	Mean Distance	Median Distance	IoU Calving Front	IoU Ice/Ocean
M-VS	CALFIN-NN	2.56 px, 97.72 m	2.55 px, 97.44 m	0.3332	N/A
M-VS	M-NN	1.97 px, 96.31 m	N/A	N/A	N/A

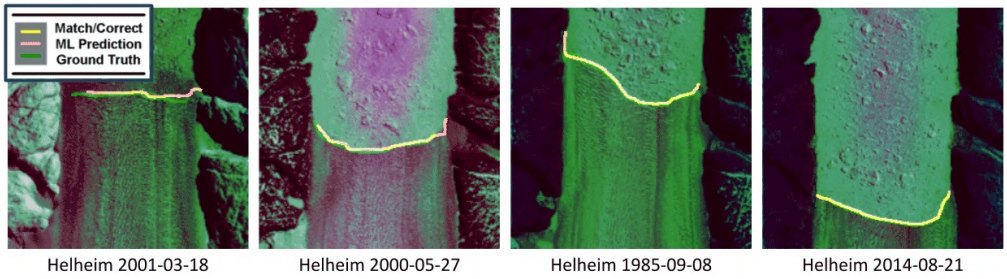

Figure 9. M-VS Validation Output Results: Note that CALFIN-NN has never trained on Helheim, but can still predict the front at multiple scales and conditions. See Fig. S8. for full outputs.

4.3.3 Zhang et al. (2019) Validation Set

CALFIN-NN performs competitively on the Z-VS. It achieves a similar mean meter distance (115.24 m vs. 104 m) despite being constrained to using lower resolution TerraSAR-X data. Note though that the Zhang et al. (2019) neural network (Z-NN) uses higher resolution input data (960×720) compared to CALFIN-NN (224x224), which skews the mean pixel distance comparison, where CALFIN-NN performs better (2.11 px vs. 17.3 px). Another source of skew comes from CALFIN-NN confidence filtering, as only 8 of 12 fronts in the set are confidently detected (see Sect. 4.4). Increasing CALFIN-NN's input resolution and training on higher resolution SAR data may enable CALFIN-NN to detect more fronts with greater accuracy.

Validation Set	Model	Mean Distance	Median Distance	IoU Calving Front	IoU Ice/Ocean
Z-VS	CALFIN-NN	2.11 px, 115.24 m	1.65 px, 77.29 m	0.3832	0.9761
Z-VS	Z-NN	17.3 px, 104 m	N/A	N/A	N/A

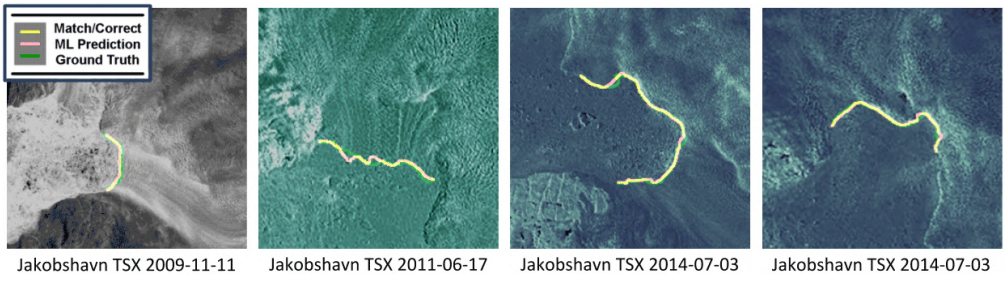

Figure 10. Z-VS Validation Output Results: CALFIN-NN works well on SAR data in addition to optical data. See Fig. S9. for full outputs.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

4.3.4 Baumhoer et al. (2019) Validation Set

CALFIN-NN performs sub-par on the B-VS. However, when comparing the mean distance error with the Baumhoer et al. (2019) equivalent Area over Front (A/F) error, easily detected static coastlines are masked out, raising the relative error. When comparing metrics that isolate the calving front, the absolute median distance error is calculated (achieving 112.75 m) whereas Baumhoer et al. (2019) uses signed median distance error (achieving 0 m), which is not directly comparable in this context, and thus omitted. Currently, the error is affected by kilometer-range deviations in very large domains like Voyeykov Ice Shelf, and differences in sea-ice mélange as seen along the Gillet and Wordie Ice Shelves, which would be consistent with findings in Baumhoer et al. (2019) Sect. 5.2. After excluding such outliers, fronts are detected in 55 out of 62 domains (88.71%), achieving median distance errors of 0.95 px (127.87 m). Intensive retraining on ice shelves may be required for CALFIN-NN to improve.

Validation Set	Model	Mean Distance	Median Distance	IoU Calving Front	IoU Ice/Ocean
B-VS	CALFIN-NN	2.35 px, 330.63 m	0.74 px, 112.75 m	0.6451	0.9879
B-VS	B-NN	2.69 px, 108 m	N/A	N/A	0.905

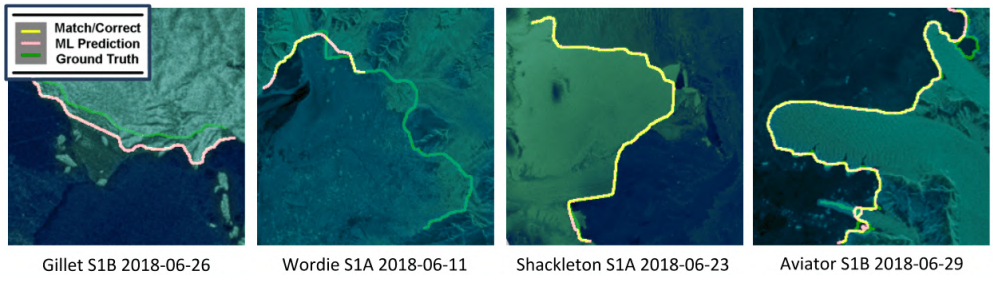

Figure 11. B-VS Validation Output Results: Similar to Z-NN, B-NN uses a high resolution input (768×768) relative to CALFIN-NN (224x224), which skews the mean pixel distance comparison undeservedly in CALFIN-NN's favor. See Fig. S10-S11 for full outputs.

4.4 Detection Accuracy

Lastly, CALFIN-NN is shown to automatically filter images that do not have detectable calving fronts. To verify this, 13 images are included in the CALFIN-VS which do not contain calving fronts discernible to the human eye. The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) rates are computed for the entire 162 image CALFIN-VS, and the associated confusion matrix is shown below. Note that CALFIN-NN does not output any false positives on the CALFIN-VS. While this ensures accurate fronts are output rather than incorrect fronts, this filtering behavior removes potentially large errors, and must be accounted for when comparing errors across other sets.

Table 2. Confusion Matrix: CALFIN-NN misses fronts in 8 of 149 valid CALFIN-VS images, but this is deemed as an acceptable trade-off.

		Front Detected?	
		Yes	No
Front Exists?	Yes	TP = 141/149 = 94.63%	FN = 8/149 = 5.76%
	No	FP = 0/13 = 0.00%	TN = 13/13 = 100.00%

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

5 Data Product Results

The code implementation of the CALFIN method is released, along with its associated calving front data products as described in the following subsections, for use within the scientific community.

5.1 CALFIN Dataset

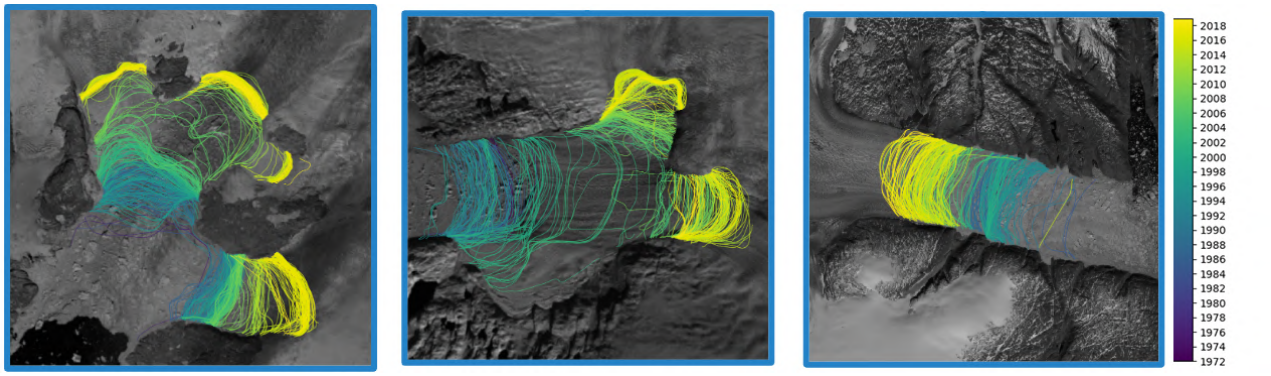

Figure 12. CALFIN Dataset Samples: Data products for Upernavik (left), Jakobshavn (center), and Helheim (right), from 1972-2019.

The CALFIN dataset spans 66 Greenlandic basins, over the period Sept. 1972 - June 2019. This consists of over 1500 manual delineations and 22,678 total calving fronts. Two levels of CALFIN data products are provided. The level 0 products include the Shapefile domains used for subsetting, the neural network training image/mask pairs, the fjord boundary masks, the full Landsat scene ID list, and the quality assurance images for validation purposes. The use cases of Level 0 products may include studies of reproducibility, validation, or training new neural networks. The level 1 products includes the calving front polyline and polygon Shapefiles. The polyline product consists of the isolated, refined, geo-referenced, and verified calving fronts for each domain. The polygon product consists of an ocean mask bounded by the domain subset, the fjord boundaries, and the calving front(s), for each domain. Both of the Shapefiles share a common metadata feature schema (see Table S2) derived from the MEaSUREs Glacial Termini Dataset (Moon and Joughin, 2008; Joughin et al., 2015), and names are derived from Bjørk et al. (2015). These products can be found at datadryad.org/stash/share/Q9guqsrdoB7v2a9JSLsgoV6HY_RS8RkCDvStx2eWsBg.

5.2 CALFIN-NN Implementation

An implementation of CALFIN-NN is available at github.com/daniel-cheng/CALFIN. Innovations as described in Sect. 3.2 can also be applied to other networks and investigations. The implementation and its associated processing scripts are written in Python 3, using the Keras & Tensorflow libraries. Note that access to the network parameters is hosted separately as part of the associated DataDryad dataset linked above (Cheng et al., 2020). For additional insight into the network training and processing requirements, see the following discussion in Sect. 6.1.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

6 Discussion

6.1 Training Insights

Throughout the course of the study, several innovations are developed to improve the performance of CALFIN-NN. To increase accuracy, a custom Intersection-over-Union based loss function is used to heavily penalize incorrect calving front predictions. To prevent over-fitting the neural network, a large set of training data was manually delineated (see Fig. S3), totalling 1541 Landsat and 232 Antarctic SAR image/mask pairs, with the SAR data taken from the same training scenes used by Baumhoer et al. (2019). Another measure to prevent over-fitting involves data augmentation, which entails performing random flips/transpositions, random Gaussian noise, random sharpen filters, random rotations of up to 12°, random crops, and random scaling on the pre-processed images during CALFIN-NN training. Through empirical testing, it is determined that excessive image padding, rotation, warping, and cropping calving fronts to close to the image bounds result in sub-optimal performance. Yet another helpful technique is the use of test-time augmentations. More specifically, each image subset is cut into 9 overlapping 224x224 image windows and processed individually, before being reassembled into the final 256x256 output mask. This allows for multiple independent classifications of the central pixels, ensuring agreement and confidence in detected calving fronts.

After integrating these improvements, CALFIN-NN is trained for a total of 80 epochs, with 4000 batches per epoch, and 8 images per batch. Training is carried out on a K40 Nvidia Tesla GPU with 12GB of VRAM, with each epoch taking about 126 minutes to complete, and almost 1 week in total to obtain the optimal weights at epoch 65. Once trained, an NVIDIA GTX1080 with 6GB VRAM for off-line data processing. The CALFIN algorithm (excluding preprocessing, but including post-processing) is capable of handling about 4 subsets per minute, taking about 3.5 days to process all 20188 GeoTIFF subsets into calving fronts. Future investigations should thus consider the trade-offs between the processing time of large networks, their accuracy, and the required computational resources, as existing works may prove suitable for their needs.

6.2 Existing Works

Mohajerani et al. (2019) is an example of an existing study that uses deep neural networks to detect calving fronts. The study pioneers the UNet-style network for application towards the Greenlandic glacial basins Jakobshavn, Helheim, Sverdrup, and Kangerlussuaq. While the methodology is restricted by its preprocessing requirements and inability to handle branching/non-linear calving fronts, it nonetheless supports the viability of the neural network method. Zhang et al. (2019) and Baumhoer et al. (2019) evaluate modified UNet architectures, as applied to SAR data in Jakobshavn and Antarctica, respectively. Their studies incorporate large spatial context in order to capture high resolution and potentially whole-coastline delineations. These larger networks, on the order of 960x720 and 768x768 pixels respectively, support the viability of both training and applying large networks to new data. The CALFIN-NN method builds on these studies by improving on the network design, capability, and post-processing methods. The following section shows a data analysis example as performed in Zhang et al. (2019), and similarly showcases a possible application of calving front datasets in advancing the understanding of the Greenland Ice Sheet.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

6.3 Data Analysis and Usage Example

With the new data available to use in the CALFIN dataset, it is possible to explore a subset and validate the evolution of Helheim Glacier against existing ESA-CCI, MEaSUREs, and PROMICE data products (ENVEO, 2017; Joughin et al., 2015; Andersen et al., 2019). Similar to Zhang et al. (2019), the relative change in position of the calving front along the fjord centerline from 1972 to June 2019 is graphed. For Joughin et al. (2015), if a date range is given, the same relative change at both start and end dates (Moon and Joughin, 2008) is plotted. For Andersen et al. (2019), August 15th is used as the "end-of-melt-season" date of delineation, as the date is otherwise not specified in the provided data. Fig. 13 shows the length change of the calving front along the basin centerline, relative to its Sept. 6, 1972 position.

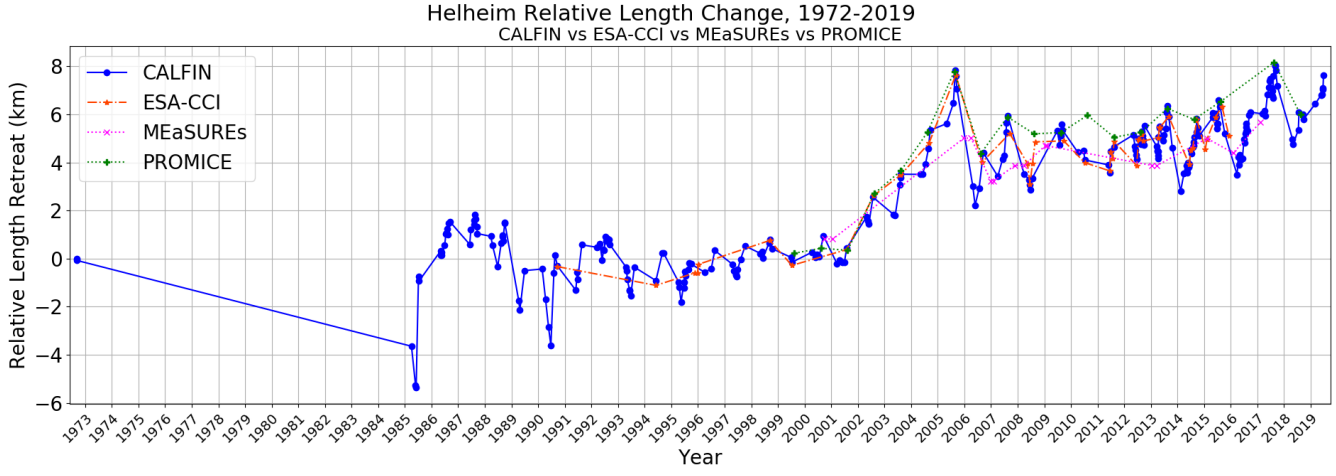

Figure 13. Helheim Terminus Length Change Over Time. Positive length change represents retreat relative to the 1972 position. Seasonal variations are captured by CALFIN (blue). Time series for other studies span 1995-2016 (ESA-CCI), 2005-2017 (MEaSUREs), and 1999-2019 (PROMICE). See Fig. S12 for an enlarged CALFIN/ESA-CCI 1995-2016 comparison.

Overall, there is high agreement between CALFIN and existing data products on the evolution of Helheim over the available time series. Note that while Helheim is relatively easy to accurately and automatically delineate, all of the above data is still produced without manual input outside of visual verification. Thus with this context in mind, this comparison with existing data products helps validate the applicability of this study's outputs.

6.4 Inter-model Comparison

To similarly reinforce the validity of the study, and address the shortcomings of different error metric comparisons (as discussed in Sect. 4.3), a comprehensive inter-model comparison is conducted between CALFIN-NN and the model developed by Mohajerani et al. (2019) (M-NN). This experiment seeks to understand how both models perform, holding all other variables constant. In particular, this experiment seeks to determine if the M-NN model, and by extension other UNet models, perform on par with the CALFIN-NN model, given the same training data. To perform this task, the M-NN is retrained using CALFIN training data, process validation data, and compare the results. For the fairest results, only images without L7SCEs are evalu-

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

ated, which is within the known capabilities of the M-NN. Furthermore, the same pre- and post-processing is applied to both models.

Table 3. Model Inter-comparison Error Table: Metrics for the CALFIN-NN and M-NN models on all non-Landsat 7 test images in the CALFIN validation set.

Validation Set	Training Set	Model	Mean Distance	Median Distance	IoU Front	IoU Ice/Ocean
CALFIN-VS-L7-none	CALFIN	CALFIN-NN	2.27 px, 81.65 m	1.16 px, 44.01 m	0.4880	0.9819
CALFIN-VS-L7-none	CALFIN	M-NN	4.45 px, 201.35 m	1.25 px, 50.52 m	0.4935	0.9699

Across all non-Landsat 7 test images in the CALFIN validation set, CALFIN-NN attains a 2.27 pixel (81.65 meter) mean distance between the predicted and the ground truth fronts. This exceeds the level of accuracy achieved by the model from Mohajerani et al. (2019), which after retraining on CALFIN training data, is 4.45 pixels (201.35 meters). Note again that Landsat 7 images were excluded during reevaluation for the M-NN. This supports the findings that the CALFIN-NN architecture is an improvement over existing UNet models.

With this added context, the validation table is reproduced from Sect. 4.3.2, Fig. 9, and the error analysis is continued below. To reemphasize the differences in mean distance error calculation between different studies, Mohajerani et al. (2019) begins by breaking each delineated front to 1000 smaller segments within a small buffer from the fjord walls and calculating the mean deviation between the segments of the true and delineated fronts. The method begins by averaging the mean distance between each pixel of the delineated front and the closest pixel of the true front as detailed in Sect 4.1. While the line-segment methodology of Mohajerani et al. (2019) provides a stricter estimate by enforcing close agreement between corresponding front segments, the CALFIN method allows for non-aligned evaluation of the mean distance error. Although both implementations quantify the differences between the lines, the differences in implementation should still be considered when evaluating the comparison below.

Table 4. M-VS Validation Output Results: Accuracy and error metrics for the CALFIN-NN and the M-NN models on the M-VS. Again, some metrics are not provided by Mohajerani et al. (2019), so they are omitted from this table.

Validation Set	Training Set	Model	Mean Distance	Median Distance	IoU Front	IoU Ice/Ocean
M-VS	CALFIN	CALFIN-NN	2.56 px, 97.72 m	2.55 px, 97.44 m	0.3332	N/A
M-VS	Mohajerani	M-NN	1.97 px, 96.31 m	N/A	N/A	N/A

Across all 10 test images in the M-VS, CALFIN-NN attains a 2.56 pixel (97.72 meter) mean distance between the predicted and the ground truth fronts. This approaches the level of accuracy achieved in the original study, which is 1.97 pixels (96.31 meters). This supports the findings that the CALFIN-NN architecture generalizes to new data well. Note that CALFIN-NN's larger network size requires additional training data to avoid overfitting, which could explain the slightly lesser accuracy when compared to the M-NN. In summary, this comprehensive model inter-comparison supports the hypothesis that the CALFIN-NN model improves on existing studies and is generalizing well.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

7 Conclusion

Overall, the goal of automatically delineating calving fronts from satellite imagery is accomplished. The CALFIN method uses the cutting-edge in deep learning architectures, allowing for robustness to minor cloud cover, Landsat 7 Scanline Corrector Errors, and illumination changes. Future work may entail accuracy improvements, expansion of included domains, usage of SAR data sources, and near-real time data products. Within the community, the benefits of standardized training, validation sets, and outputs/metadata are anticipated. The community's development of new automated extraction studies, such as grounding line delineation, iceberg tracking, and sea ice mélange measurements, is also anticipated. A key takeaway is the maturation of neural networks for automated calving front detection. Specifically, a well trained network now approaches human levels of accuracy in picking arbitrary glacial calving fronts. This reinforces existing studies on the viability of the methodology, and paves the way for applications on other data processing tasks. Ultimately, this work showcases the state-of-the-art in automated calving front detection, and provides a new database of glacial termini positions for the cryosphere community.

Code and data availability. The code used to automate the implement the CALFIN pipeline is freely available at github.com/daniel-cheng/CALFIN. The data generated by CALFIN is currently available at datadryad.org/stash/share/Q9guqsrdoB7v2a9JSLsgoV6HY_RS8RkCDvStx2eWsBg.

Author contributions. DC developed the code/model, created the training data, carried out the data processing/error analysis, and wrote the majority of the manuscript. WH provided input on the processing methodology, post-processing algorithms, error analysis, discussion topics, and writing the manuscript. EL provided key direction for the overall study, error analysis, outputs, and writing the manuscript. YM performed the model inter-comparison and assisted with the writing of the manuscript. MW performed the data preprocessing for the model inter-comparison. IV assisted in organizing collaborators and the model inter-comparison. ER contributed suggestions regarding the error analysis and inter-comparison. WH, EL, MW, and YM revised the manuscript and results.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work was conducted as a collaboration between NASA's Jet Propulsion Laboratory and the University of California, Irvine. The CALFIN neural network architecture implementation is derived from Emil Zakirov's Deeplabv3+ Xception codebase at github.com/bonlime/keras-deeplab-v3-plus (last access: 13 August 2020). We acknowledge the USGS for providing Landsat-1-8 images, the ESA for their Sentinel-1 images, as well as the ESA-CCI, PROMICE, and MEaSUREs programs for providing calving front data used in this study.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

References

Andersen, J. K., Fausto, R. S., Hansen, K., Box, J. E., Andersen, S. B., Ahlstrøm, A. P., Dirk, Citterio, M., Colgan, W., Karlsson, N. B., and et al.: Update of annual calving front lines for 47 marine terminating outlet glaciers in Greenland (1999–2018), GEUS Bulletin, 43, https://doi.org/10.34194/GEUSB-201943-02-02, 2019.

Andersen, M., Stenseng, L., Skourup, H., Colgan, W., Khan, S., Kristensen, S., Andersen, S., Box, J., Ahlstrøm, A., Fettweis, X., and Forsberg, R.: Basin-scale partitioning of Greenland ice sheet mass balance components (2007–2011), Earth and Planetary Science Letters, 409, 89 – 95, https://doi.org/10.1016/j.epsl.2014.10.015, 2015.

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 Sensing, 11, 2529, https://doi.org/10.3390/rs11212529, 2019.

Bjørk, A. A., Kruse, L. M., and Michaelsen, P. B.: Brief communication: Getting Greenland's glaciers right – a new data set of all official Greenlandic glacier names, The Cryosphere, 9, 2215–2218, https://doi.org/10.5194/tc-9-2215-2015, 2015.

Bunce, C., Carr, J. R., Nienow, P. W., Ross, N., and Killick, R.: Ice front change of marine-terminating outlet glaciers in northwest and southeast Greenland during the 21st century, Journal of Glaciology, 64, 523–535, https://doi.org/10.1017/jog.2018.44, 2018.

Carr, J. R., Stokes, C. R., and Vieli, A.: Threefold increase in marine-terminating outlet glacier retreat rates across the Atlantic Arctic: 1992–2010, Annals of Glaciology, 58, 72–91, https://doi.org/10.1017/aog.2017.3, 2017.

Catania, G. A., Stearns, L. A., Sutherland, D. A., Fried, M. J., Bartholomaus, T. C., Morlighem, M., Shroyer, E., and Nash, J.: Geometric Controls on Tidewater Glacier Retreat in Central Western Greenland, Journal of Geophysical Research: Earth Surface, 123, 2024–2038, https://doi.org/10.1029/2017JF004499, 2018.

Catania, G. A., Stearns, L. A., Moon, T. A., Enderlin, E. M., and Jackson, R. H.: Future Evolution of Greenland's Marine-Terminating Outlet Glaciers, Journal of Geophysical Research: Earth Surface, 125, e2018JF004 873, https://doi.org/10.1029/2018JF004873, 2020.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, European Conference on Computer Vision, pp. 801–818, http://arxiv.org/abs/1802.02611, 2018.

Cheng, D., Hayes, W., Larour, E., Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E.: CALFIN: Calving Front Dataset for East/West Greenland, 1972-2019, https://doi.org/10.7280/D1FH5D, 2020.

Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions, Computer Vision and Pattern Recognition, pp. 1800–1807, https://doi.org/10.1109/cvpr.2017.195, 2017.

ENVEO: Greenland Calving Front Dataset, 1990 - 2016, v3.0, http://products.esa-icesheets-cci.org/products/downloadlist/CFL/, 2017.

Fürst, J. J., Goelzer, H., and Huybrechts, P.: Ice-dynamic projections of the Greenland ice sheet in response to atmospheric and oceanic warming, The Cryosphere, 9, 1039–1062, https://doi.org/10.5194/tc-9-1039-2015, 2015.

Joughin, I., Moon, T., Joughin, J., and Black, T.: MEaSUREs Annual Greenland Outlet Glacier Terminus Positions from SAR Mosaics, Version 1, https://doi.org/10.5067/DC0MLBOCL3EL, 2015.

King, M. D., Howat, I. M., Jeong, S., Noh, M. J., Wouters, B., Noël, B., and van den Broeke, M. R.: Seasonal to decadal variability in ice discharge from the Greenland Ice Sheet, The Cryosphere, 12, 3813–3825, https://doi.org/10.5194/tc-12-3813-2018, 2018.

Kruskal, J. B.: On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem, Proceedings of the American Mathematical Society, 7, 48–50, https://doi.org/10.2307/2033241, 1956.

Malik, J., Belongie, S., Leung, T., and Shi, J.: Contour and Texture Analysis for Image Segmentation, Int. J. Comput. Vision, 43, 7–27, https://doi.org/10.1023/A:1011174803800, 2001.

https://doi.org/10.5194/tc-2020-231
Preprint. Discussion started: 14 October 2020

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

Moon, T. and Joughin, I.: Changes in ice front position on Greenland's outlet glaciers from 1992 to 2007, Journal of Geophysical Research: Earth Surface, 113, https://doi.org/10.1029/2007JF000927, 2008.

Nagler, T., Rott, H., Hetzenecker, M., Wuite, J., and Potin, P.: The Sentinel-1 Mission: New Opportunities for Ice Sheet Observations, Remote Sensing, 7, 9371–9389, https://doi.org/10.3390/rs70709371, 2015.

Nick, F. M., Vieli, A., Andersen, M. L., Joughin, I., Payne, A., Edwards, T. L., Pattyn, F., and van de Wal, R. S. W.: Future sea-level rise from Greenland's main outlet glaciers in a warming climate, Nature, 497, 235–238, https://doi.org/10.1038/nature12068, 2013.

Paravolidakis, V., Moirogiorgou, K., Ragia, L., Zervakis, M., and Synolakis, C.: COASTLINE EXTRACTION FROM AERIAL IMAGES BASED ON EDGE DETECTION, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, III-8, 153–158, https://doi.org/10.5194/isprsannals-III-8-153-2016, 2016.

Seale, A., Christoffersen, P., Mugford, R. I., and O'Leary, M.: Ocean forcing of the Greenland Ice Sheet: Calving fronts and patterns of retreat identified by automatic satellite monitoring of eastern outlet glaciers, Journal of Geophysical Research: Earth Surface, 116, https://doi.org/10.1029/2010JF001847, 2011.

van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P. Y., van de Berg, W. J., van Meijgaard, E., and Wouters, B.: On the recent contribution of the Greenland ice sheet to sea level change, The Cryosphere, 10, 1933–1946, https://doi.org/10.5194/tc-10-1933-2016, 2016.

Zhang, E., Liu, L., and Huang, L.: Automatically delineating the calving front of Jakobshavn Isbræ from multi-temporal TerraSAR-X images: a deep learning approach, The Cryosphere, 2019, 1–20, https://doi.org/10.5194/tc-2019-14, 2019.