Using Deep Learning in operational Data Products – Lessons learned from the IceLines Dataset on Antarctic Ice Shelf Front Change

Celia Baumhoer*,[1], Andreas Dietz[1], Jan-Karl Haug[1] and Claudia Kuenzer[1,2]
[1]German Aerospace Center (DLR), Earth Observation Center (EOC), Oberpfaffenhofen, Germany (*celia.baumhoer@dlr.de)
[2]University of Würzburg, Institute of Geography and Geology, Würzburg, Germany

Motivation

Antarctica's ice shelves are the floating extensions of the ice sheet. The discharge of the Antarctic ice sheet increases if ice shelf areas with strong buttressing forces are lost. This has direct implications on Antarctica's contribution to global sea level rise. Therefore, it is important to have an operational product constantly providing data on ice shelf front position to locate and track changes in ice shelf extent. Here, we present the workflow of the IceLines dataset showcasing a processing pipeline from acquired satellite data to a deep learning (DL) derived data product.

HED-UNet

HED-UNet is a deep convolutional neural network (CNN) combining segmentation and edge detection in one task. To improve the HED-UNet performance for front detection, deep supervision is applied. This widens the receptive field for better capturing larger features. Finally, hierarchical attention merging heads allow to attend to different resolution levels. That means coarse predictions are used for marginal areas and high-resolution predictions for the coastline itself. This deep learning approach delivers accurate results (209 ± 12 m, 5.2 pixel) on dual polarized imagery and 432 ± 21m (8.8 pixel) on single-pol imagery for a variety of Antarctic ice shelves.

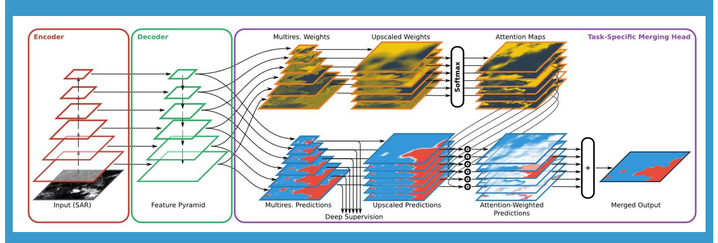


Fig. 1 HED-UNet architecture for the segmentation merging head (Heidler et al. 2021)

Lessons Learned

- CNN predictions remain accurate even after large calving events not represented in the training dataset
- Unavailability of Sentinel-1B impacts and subsequent changes in the acquisition scheme decreased data availability. Integration of IW high-resolution data would now be beneficial
- Usage of Docker ensures reliable processing on a consistent runtime environment to run the application every month. Only updates on GPU drivers can create issues.

Today, the IceLines dataset includes over 19.500 front positions and provides valuable input for Antarctic studies focusing on oceanography, sea ice, polynyas, ocean/terrestrial ecology, ice sheet modelling, iceberg calving and mapping.

DLR

IceLines Data Download

IceLines Map Visualization

1 Sentinel-1

Sentinel-1A was launched in 2014 and acquires radar data over the Antarctic coastline on a regular basis (every ~12 days, 6 days with Sentinel-1B). Continuous data availability is secured as SAR sensor acquisitions are independent from solar illumination and clouds.

2 Data Access

Sentinel-1 SAR data archive is available through the ASF (Alaska Satellite Facility) Distributed Active Archive Center. For each monitored ice shelf an automated data download is triggered on a monthly basis via a cronjob. Filtering for specific products (GRD EW, path, frame) reduces the data amount.

monthly data download

3 HPC Cluster

Raw Sentinel-1 data is stored on a HPC cluster with 63 DataNodes (32/64 GB RAM) having a total storage capacity of 1.6 PB. The master node manages the Hadoop distributed file system (HDFS) and assigns work to the DataNodes.

HED-UNet Training

HED-UNet was trained on 81 Sentinel-1 radar satellite images over the Antarctic coastline for single- and dual-polarized images. The weights of the best model are stored on the GPU cluster for prediction.

stored weights for single- and dual-polarized images

4 Pre-Processing

The Sentinel-1 data is pre-processed with the graph processing tool of the ESA Toolbox allowing command line processing. Docker containers are used to run SNAP on the HPC cluster.

5 Prediction

The pre-processed scenes are copied to the GPU cluster and front positions are predicted based on the pre-trained weights. Selection of single/dual-pol weights happens automatically.

6 Post-Processing

Post-Processing improves the prediction results by physically determined assumptions such as elevation threshold and class connectivity. Finally, the raster is binarized and converted to a vector shapefile.

7 Time Series Check

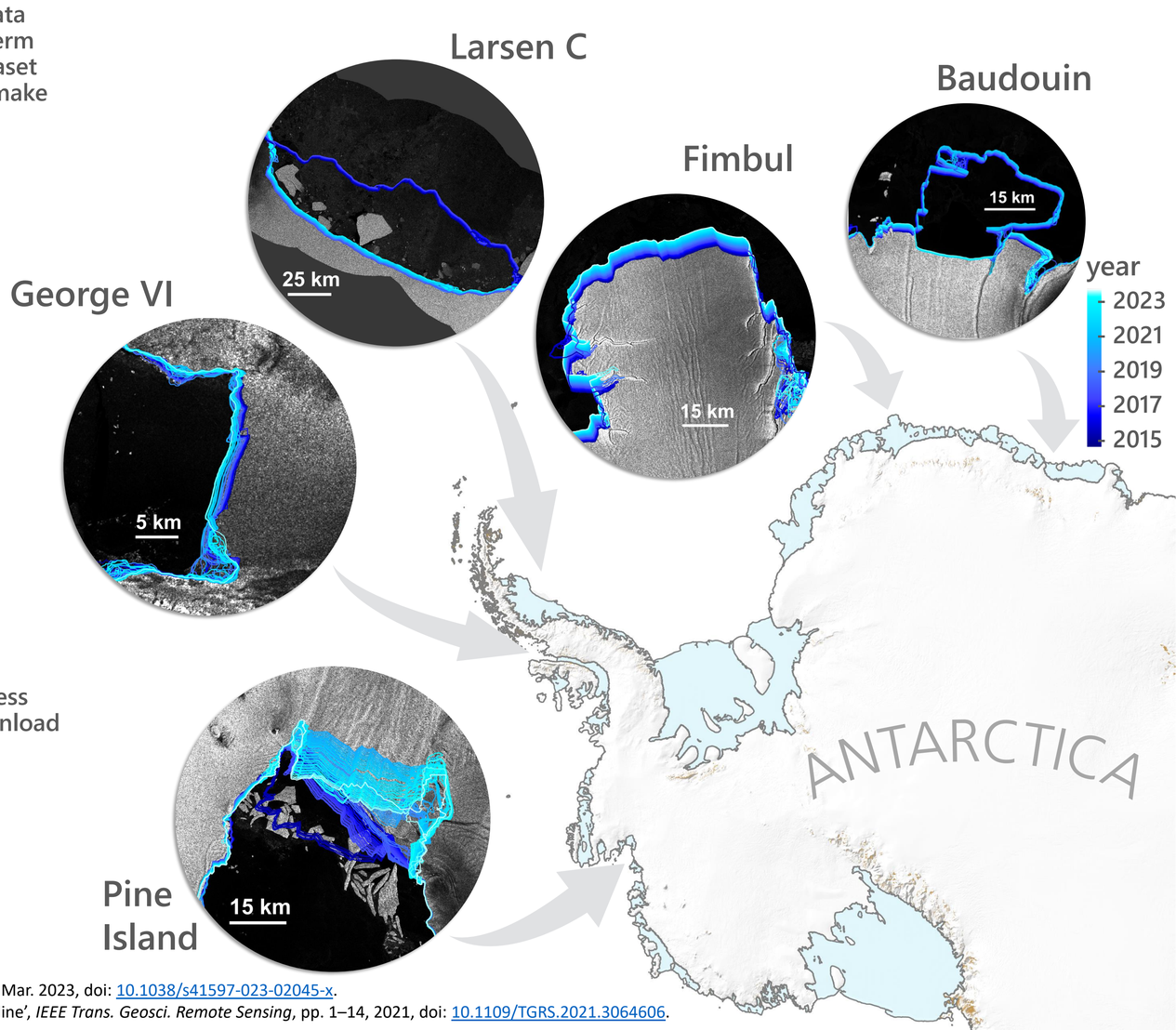
Erroneous results can be produced due to surface melt on the shelf or persistent sea ice. Potentially wrong front positions are identified by the assumption that an ice shelf front can either advance, retreat (calving event) or disintegrate. Any front positions differing from this pattern are sent for review.

8 Long-Term Storage

The ice shelf front position dataset 'IceLines' is stored on the internal data access (IDA) of DLR to ensure long-term storage and access. The ongoing dataset is registered with a persistent DOI to make it citable and accessible.

9 DLR's GeoService

DLR's Geoservice provides open access discovery, visualization, and direct download services for the IceLines dataset.



[1] C. A. Baumhoer, A. J. Dietz, K. Heidler, and C. Kuenzer, 'IceLines – A new data set of Antarctic ice shelf front positions', *Sci Data*, vol. 10, no. 1, p. 138, Mar. 2023, doi: 10.1038/s41597-023-02045-x.
[2] K. Heidler, L. Mou, C. Baumhoer, A. Dietz, and X. X. Zhu, 'HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline', *IEEE Trans. Geosci. Remote Sensing*, pp. 1–14, 2021, doi: 10.1109/TGRS.2021.3064606.