AI4IS - AI Forecasting for Ice Shelf Calving using Deep Learning

Jacob Alexander Hay[1], Hamze Issa[1], Daniele Fantin[1], and the AI4IS Team

[1]S&T Norway

Contact: hay@stcorp.no

Abstract

Larger calving events in the Antarctic Ice Shelf (AIS) have been seen in the past few decades, and are a potential precursor for collapse. We use Deep Learning to predict instability and calving events on the AIS. To do so we use a 4D multivariate datacube of Essential Climate Variables (ECVs) resampled to **200m** resolution, using a novel Gaussian Random Field (GRF) representation. The resulting datacube is coupled with a calving inventory from Qi et al. (2021) [1] , limited to the study site, before being processed using a UNet / VGG model. Preliminary results suggest calving being learnable with a lead time of 6 months with a calving/no-calving F1-score of $\mathbf{0.9}$, on large calving events.

Method

Samples from the datacube are aligned with the label dataset [1] at an offset in time, i.e. we use future calving events as the target, sampled with a `lead time`-argument. The labels are marked as being "calving" if there is a calving event at any point between the time of the data, t, and the time plus `lead time`, $t+T$. Initial experiments have established correlations on single-class lead times, to be later extended to multiple lead time-classes following the given definition for classes, $Y_T[t]$:

$$Y_T[t] = \bigcup \text{Labels}[t, t+T] \mid T \in [9, 6, 3],$$

Labels are assumed to be complete for supervised learning. Our first study site, **Larsen C**, has a few smaller calving events coinciding with the available data, and one very large one, the **A-68 Iceberg** of 2017. We split the data into train/test areas of interest (AOIs), where a portion of **A-68**, and some smaller calving events were reserved for validation. We train using use **Focal Loss** [2], as our target for areas that are identified as calving within the range of the given `lead time`. For data augmentation, we currently only use rotation on random samples.

Sampling & Data

We read the data in the resolution of the cube and make samples from chips and rasterised labels over the available data. The samples consist of a `(Data, Label)`-pair, where we can generate samples a tunable `lead time`-parameter.

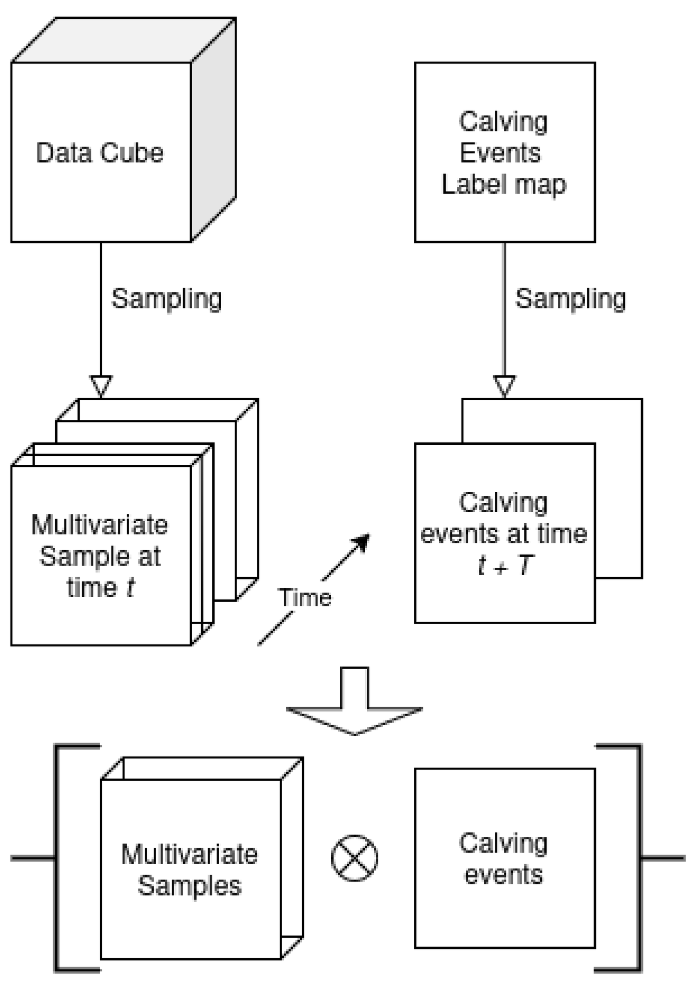

Figure 1: Sampling pipeline.

The ECVs included currently in the datacube are:

- Ice Velocity,
- Wind Speed,
- Basal melt,
- Surface Mass Balance (SMB / RACMO)
- Firn thickness & Firn Air

Training

Our sample chips have the data, sampled from the datacue at time t, and the labels sampled with an added lead-time $t+T$. The input data is all sampled to the same resolution as the input chips of 200 m. We train a simple 5-layer **Attention-UNet**, [3], (Figure 2), to predict the future calving as a semantic segmentation task. We are using a standard **AdamW** optimizer, [4], with **Focal Loss** as our target [2]. Due to the sparsity of our available label set of known calving events, the samples are kept small, at 128×128. For training/validation we split our area of interest (AOI) spatially, to avoid temporal mixing. We then sample randomly from a class-balancing weighted grid of valid samples. Since the models are trained on smaller chips, providing a full-scale prediction over the entire ice shelf is done by stitching together multiple samples. For this, we prepare a grid that covers the ice shelf, with a margin of overlap that allows us to isolate the predictions that are entirely covered by the receptive field of the feature-detection components of the model. For more robust results we randomly apply some rotation to the samples in our initial experiments, later experiments are expected to include more augmentations in the second phase of the project.

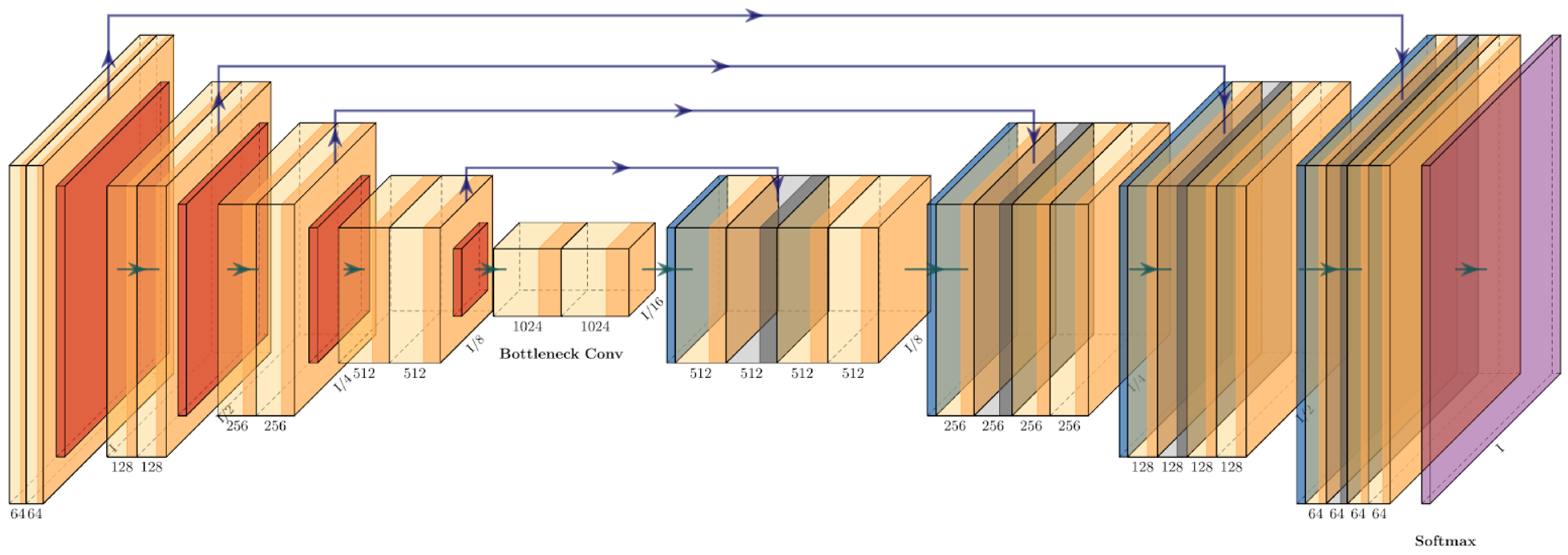

Figure 2: Basic UNet - model architecture.

Explaining

The final component of the experiments is to test the trained model through an eXplainable AI (XAI), which highlights the most significant input features for a given prediction. Our approach is initially based on **Grad-CAM**, [5], which makes use of false labels and backpropagation to identify key input variables, or areas in the multivariate input which have strong correlations to the predicted output. The resulting heat map has a 1:1 relationship with the input data, which allows us to later compare side by side the predictions and the input. Ideally this results in a situation where scientists can self-evaluate the predictions based on the significance placed on the inputs. The saliency map does not mean much without the set of value names of its weights. As such, it is interpretable as a correlation map. Comparisons of these maps should also establish the degree to which robustness follows from longer training, and delayed generalisation, **grokking** [6], which is believed to be an emergent feature of neural networks.

Results

Our first models using semantic segmentation have an F1-score of 0.9 over the test sections, which include the eastern portion of the A-68 iceberg. These results indicate that the included variables cover the required information to predict future calving events on the Larsen C ice shelf. Later experiments will establish to what degree this approach generalises across other ice shelves.

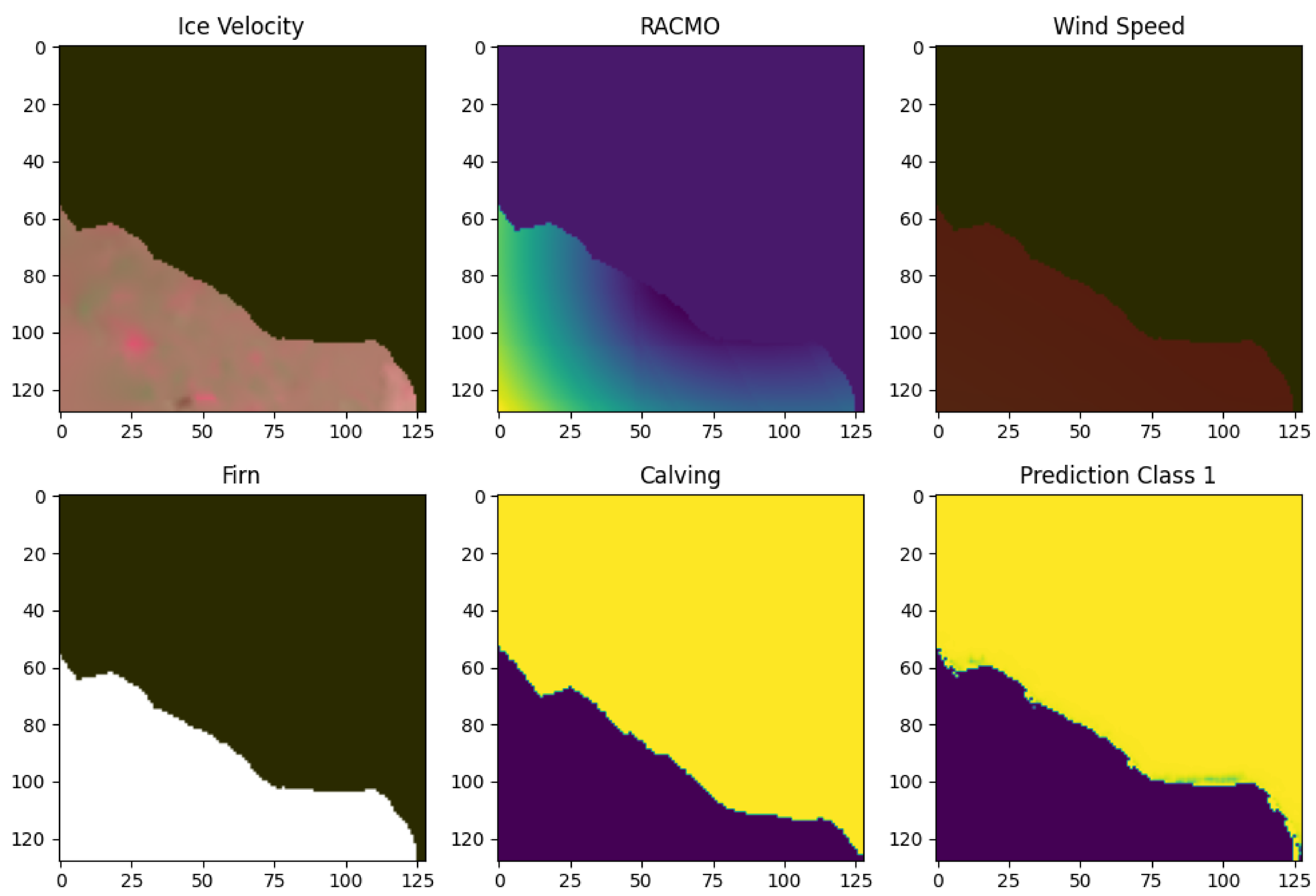

Figure 3: Example from model training

Figure 3 shows output during training, where the model is starting to fit the curve of the calving event. The model is here at an F1 of 0.916.

Future work

The work will continue on a second study site over Pine Island, where we will also test transfer learning across the two sites/ice-shelves. As the models converge we will also include XAI heat-maps at regular intervals to monitor the saliency of the different inputs to the model, as well as the changes over these as the models improve. Finally the model prediction and associated heat maps will be evaluated by domain experts from Lancaster University, who will evaluate how reasonable the model correlations are, based on our current understanding of ice shelf dynamics. The AI4IS project will also aim at future scientific publication, which will be shared on the project website.

References & Acknowledgements

This work was funded by the European Space Agency (ESA) as part of the AI4Science ITT (ESA Contract No 4000143299/23/I-DT)

[1] M. Qi, Y. Liu, J. Liu, X. Cheng, Y. Lin, Q. Feng, Q. Shen, and Z. Yu. A 15-year circum-Antarctic iceberg calving dataset derived from continuous satellite observations. *Earth System Science Data*, 13(9):4583–4601, 2021. doi: 10.5194/essd-13-4583-2021.

[2] Isaac Newton, Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 1:2980–2988, 2017. doi: 10.1109/TPAMI.2018.2858826.

[3] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas. (arXiv:1804.03999), May 2018. doi: 10.48550/arXiv.1804.03999.

[4] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. (arXiv:1711.05101), January 2019. doi: 10.48550/arXiv.1711.05101.

[5] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, 128(2):336–359, February 2020. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-019-01228-7.

[6] Ahmed Imtiaz Humayun, Randall Balestriero, and Richard Baraniuk. Deep Networks Always Grok and Here is Why. (arXiv:2402.15555), June 2024. doi: 10.48550/arXiv.2402.15555.