

1 Model development log and notes

1.1 Two day forecast

Data sources used are Sea Ice charts from Nick initiated at 15:00 as well as Arome Arctic initiated at 18:00 [1, 2]. For a given date, the current Ice Chart is used as a predictor for the model, while the Ice Chart drawn two days later is supplied as the model target. Moreover, the Arome Arctic data is structured such that the period between lead time 0 - 18 and 18 - 42 is stored as mean products. Thus, the period covering lead hours 3 - 21 and 21 - 45 relative to the Ice Charts are covered (and meaned). The 4d variables used from AA (which feature a temporal dimension) are T2M, Xwind and Ywind, the 3d variable used is SST, which is available in the data, though the variable itself is only utilized to force the Data Assimilation system of AA and not a product of the prediction system. Finally, a land sea mask present on previously available ice charts on lustre is adopted to fit the current data, and is used both as a predictor as well as a mask when computing metrics and visualizing results.

The model developed for the two day prediction is based on the SimpleUNET architecture, though with a different sized Input layer to accommodate for the changed dataloader. The dataloader has subsequently been changed to appropriately select the correct fields from the .hdf5 samples and appoint them as input or target variables. As a result of using three variables of two days mean AA forecast, as well as sst, land-sea-mask and current time-step ice chart, the total number of predictors fed into the model is 9. Moreover, the resolution of all fields are kept at 1km, though their spatial extent is limited to (1920 x 1840). This resolution and spatial size conserves (almost) the entirety of the west-east axis of the AA domain. However, the southern border is raised by 450km compared to the AA domain. There are two main motivations behind readjusting the spatial extent of the predictors and targets.

1. The spatial extent of the input domain has to be divisible by the reducing factor enforced by the MaxPooling operation performed in the encoding component of the UNET.
2. The southern latitudes covered by AA has a proportionally skewed Sea Ice / Ice Free open water ratio, as exemplified in Figure (1). Increasing the southern bounding latitude of the subdomain thus decreases the number of guaranteed ice free pixels, which in turn decreases the skewness towards the ice free open water class for the UNET.

1.2 Model Architecture

The model architecture follows an encoder - decoder structure, commonly referred to as a U-NET [3] due to its shape funnelling the spatial data to coarser resolution, which resembles the letter "U". The current U-NET implementation follows that of Ronneberger et.al, though it has been modified with batch normalization after each convolution operation to ensure a more stable gradient flow. The weights of the model are Kaiming-He initialized [4], as the activation function used throughout the network is the ReLU function [5]. The final output of the model is a (1920, 1840, 7) tensor containing softmaxed probabilities along its final axis.

1.3 CategoricalCrossEntropy-Loss

As the title suggests, these runs of the model involved using CategoricalCrossEntropy as the loss function for multi-class image segmentation. Categorical Cross Entropy loss is defined as

Target for two day forecast initiated 20190611

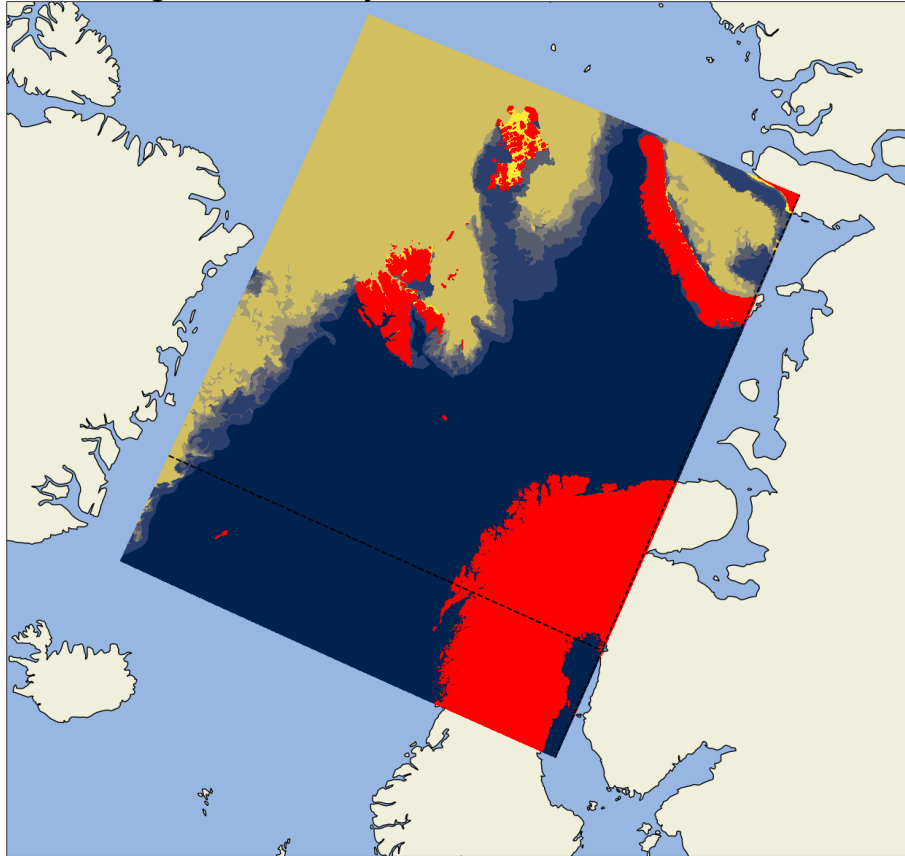


Figure 1: Example sample displaying an Ice Chart on a 1km Arome Arctic projection. Note the horizontal and vertical dashed black line which indicate the domain subsection used by the UNET

$$CE = - \sum_i^C y_i \log(\hat{y}_i) \quad (1)$$

where C denotes the number of available classes, y the ground truth and \hat{y} a prediction of y . Note that as y is onehot-encoded, the formulated function only contributes to the overall loss with the log of the predicted probability of the correct class according to the ground truth.

Two variants of the previously described model have been trained with the CategoricalCrossEntropy described in equation (1). The first model was trained with an encoder consisting of 4 convolutional blocks with channel dimensions (64, 128, 256, 512). The second model consisted of 5 convolutional blocks, with an identical architecture except for the last convolutional block increasing the channel dimension to (1024). Example outputs as well as target can be seen in Figure (2).

By inspecting Figure (2), two observations can be made. The first observation is regarding how the model complexity affects how it fit to the data. By comparing Figure (2a) with (2b), it can be seen that the latter is resolving the finer structures of the ice edge to larger extent than the prior. Though the overall correctness is left to be discovered, this shows that increasing the depth of the encoder (increasing the trainable parameter count from 7 million to 31 million) is reflected by the model preserving the details of the ice edge structure. Though it is non-trivial to say why the 1024-model preserves the details to a larger extent than the 512-model, it does follow from the U-Net architecture that a deeper encoder (higher channel count and more convolutional blocks) is better at describing "WHAT" is in the image compared to the shallow-layers, which include a larger amount of spatial information and tells the model to a larger extent "WHERE" things are in the model.

The second observation made from inspecting both forecasts is their inability to represent classes 2 and 3. This likely arises from the general movement-pattern of the sea ice, where the intermediate classes are much less likely to appear than the edge-most classes. Furthermore, the sea ice is much more likely to represent a wider range of concentration classes in the intermediate ice edge region over time, making it more difficult for the network to confidently predict those classes compared to the more probable classes. As can be seen by the network immediately predicting class 4 after class 1, creating an artificial cut-off region. However, to what extent the intermediate classes are predicted has not been inspected directly, though it is likely to assume that they are predicted though with a lower confidence than that of class 4 (which is consequently why it is visualized, as the most probable class is chosen regardless).

This may have a source

They should be

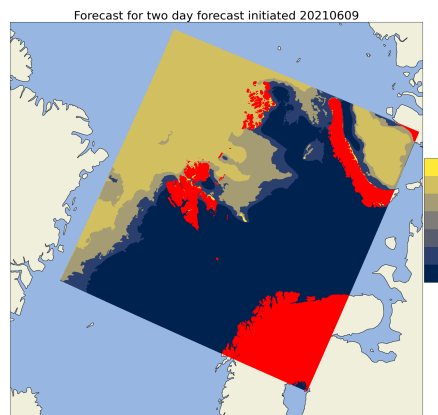
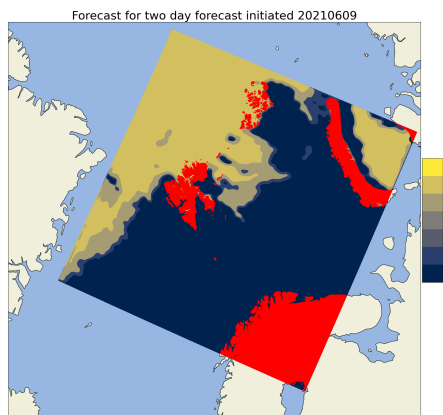
1.4 FocalLoss

The focal loss is derived as a generalization of the Cross Entropy Loss listed in Equation (1). The intent of the loss function is to downweight the easy to predict samples, while focusing on the hard to predict samples by allowing their gradient to have a higher impact on the network [6]. Mathematically, focal loss is defined as

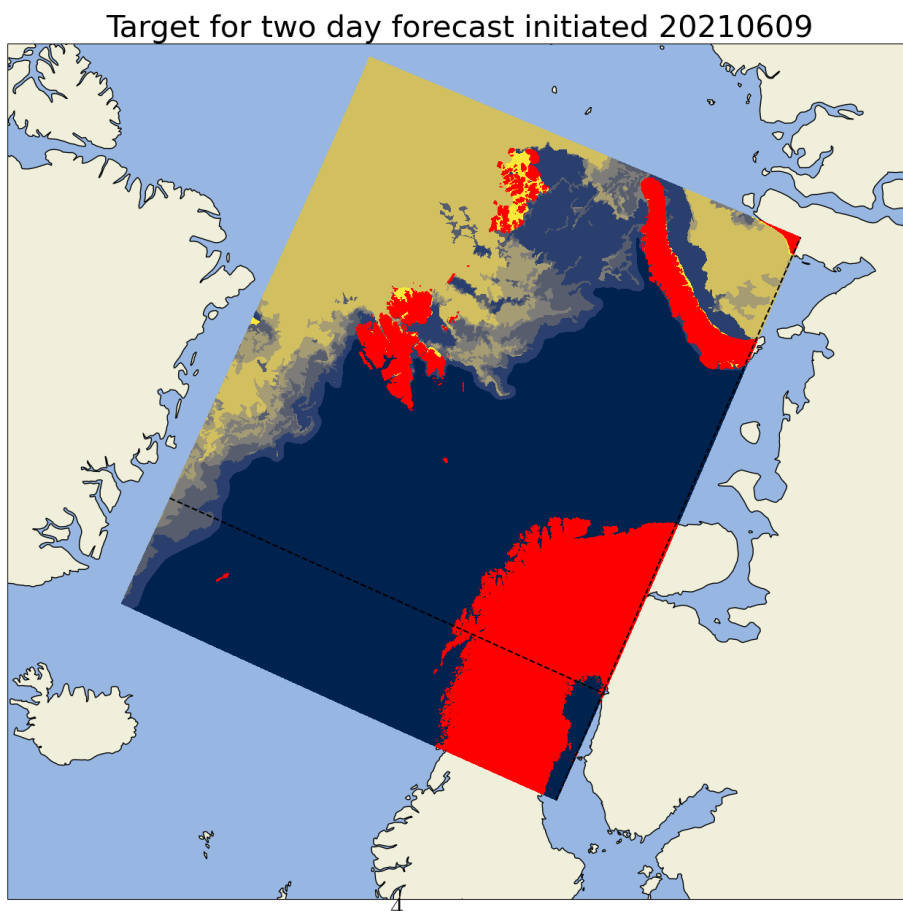
$$FL = - \sum_i^C \alpha_i (1 - \hat{y}_i)^\gamma y_i \log(\hat{y}_i) \quad (2)$$

where α is a balancing parameter, γ is the focusing parameter ($\gamma = 0 \rightarrow CE$), with the rest similar as Equation (1).

Include figure showing focal loss output, discuss implications of using this loss function



(a) Forecast with two day lead time with model_512 architecture (b) Forecast with two day lead time with model_1024 architecture



(c) Target for forecast with two day lead time

Figure 2: Example forecast attempt made by model_512 and model_1024 09-06-2021

By inspecting Equation (2), it can be seen that predictions that the model is quite confident in making, i.e. $\hat{y}_i \rightarrow 1$ send the Focal Loss towards zero. For the current application, the assumptive motivation is that this affects (by reducing) the contribution made by the Ice Free Open Water pixels as well as the Very Close Drift Ice (class 6), which are the most represented classes in the CE loss model seen in Figure (2). Consequently, as the loss contributions of the most likely (and most represented classes) is reduced, the harder to predict (both due to being less represented and due to sea ice movement) have a larger impact on the overall loss propagating backwards throughout the model. As a result, these intermediate classes should be predicted as the most likely class, resulting in a less sharp ice edge which closer represent the Ice Charts.

1.5 Cumulative probability distribution model

1.5.1 Separate convolutional layers as output

1.5.2 Single convolutional layer for multiple outputs

References

- [1] Frode Dinessen, Bruce Hackett, and Matilde Brandt Kreiner. Arctic ocean - sea ice concentration charts - svalbard and greenland, 2015.
- [2] Malte Müller, Yurii Batrak, Jørn Kristiansen, Morten A. Ø. Køltzow, Gunnar Noer, and Anton Korosov. Characteristics of a convective-scale weather forecasting system for the european arctic. *Monthly Weather Review*, 145(12):4771–4787, dec 2017.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. May 2015.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015.
- [5] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, pages 807–814, 2010.
- [6] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. August 2017.

Discuss difference in dataloader, same dataset is used differently

Data exists, start writing

Not started on