1 Datasets

This study combines different datasets consisting of fields from the Arome Arctic forecasting product and ice charts created by the Tromsø division of MET Norway.

1.1 Sea Ice Charts

The Sea Ice charts is an operational Sea Ice Concentration product provided by MET Norway. The product is manually drawn by a Sea Ice Specialist, and is distributed every workday at 15:00 UTC. The Sea Ice specialist assesses available SAR data from Sentinel 1 and Radarsat 2. However, due to the spatial variability in daily SAR coverage, visual, infrared and low resolution passive microwave observations are supplied to achieve a consistent spatial coverage [2]. The Sea Ice charts are drawn in an ArcGIS production environment, and is as such intrinsically not projected onto a defined grid. Yet, the operational product available for download on Copernicus is provided as mean values on a 1km grid.

From the description of the Sea Ice charts given above, it is worth addressing the spatial inconsistency following the projection onto a uniformly sized grid. As the Sea Ice specialist draws polygons based on data from different satellite sources with a wide range of spatial resolution (80m from SAR, 1000m from visible / infrared and even lower resolution for passive microwave), the underlying uncertainty and detailed structures in the Sea Ice chart varies [2]. Furthermore, I was made aware by one of the Sea Ice Analysts that time constraints also limits the hours different sections of the Ice chart is alloted. Moreover, the Sea Ice charts is an operational product aimed at end users in industries such as fishing, tourism, shipping or other maritime operations. This influences the decision-making when creating the final operational product. . As a consequence, the Sea Ice analyst spends approximately half of the total time draw polygons around the Svalbard archipelago.

ask Trond on email

In conclusion, concerning the limited resources available both with regards to data availability as well as total hours available, the Sea Ice charts represents a dataset with a spatial uncertainty that is non-uniform across a single sample, and that changes in time. In spite of that, the involvement of a Sea Ice specialist which manually assures each Sea Ice charts, the temporal consistency as well as their high resolution has led us to believe that the Sea Ice charts is the overall best Sea Ice Concentration product available for the current study region.

Endring

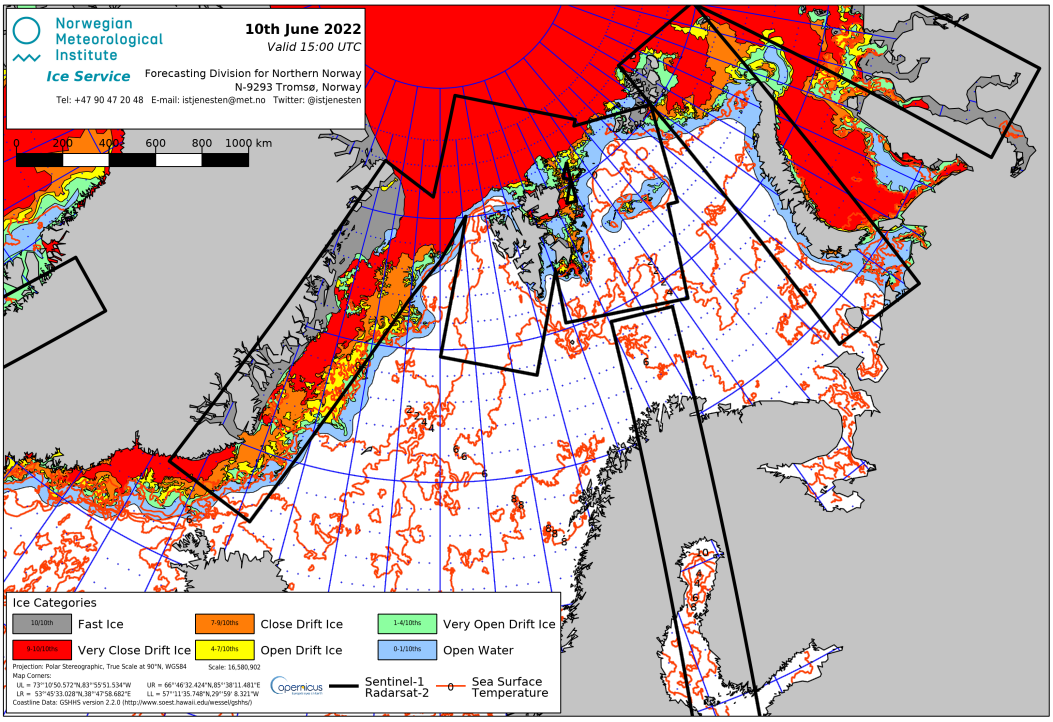
2 Data pipeline

The main motivation for writing this section is to outline the production pipeline which collects and merges different sources of data into the finished data-generator used to train the Deep Learning model. Ideally, this section will be somewhat exhausting such that all steps taken are presented clearly with the intent of providing a source for the inspiration as well as the necessity of the step. However, source code will be presented sparingly.

Sea Ice Charts

The Sea Ice Charts used are a derived dataset of the Sea Ice Charts presented in a previous section . The present Ice Chart dataset has been postprocessed by Nick Hughes of the National Ice Service,

label sections


Norwegian Meteorological Institute
Ice Service
10th June 2022
Valid 15:00 UTC
Forecasting Division for Northern Norway
N-9293 Tromsø, Norway
Tel: +47 90 47 20 48 E-mail: istjenesten@met.no Twitter: @istjenesten
0 200 400 600 800 1000 km
Ice Categories
10/10th Fast Ice
9-10/10ths Very Close Drift Ice
7-9/10ths Close Drift Ice
4-7/10ths Open Drift Ice
1-4/10ths Very Open Drift Ice
0-1/10ths Open Water
Projection: Polar Stereographic, True Scale at 90°N. WGS84
Scale: 16,580,902
Map Corners:
UL = 73°10'50.572"N,83°55'51.534"W
UR = 66°16'32.424"N,85°38'11.481"E
LR = 53°45'33.028"N,38°47'58.682"E
LL = 57°11'35.748"N,29°59' 8.321"W
Coastline Data: GSHHS version 2.2.0 (http://www.soest.hawaii.edu/wessel/gshhs/)
Copernicus
Sentinel-1
Radarsat-2
Sea Surface Temperature

such that they are presented on a 1km Arome Arctic grid. Furthermore, the Ice Charts does not feature a land-mask, which has been replaced with interpolated values resulting in a spatially consistent dataset where all values present are according to the WMO Sea Ice Concentration intervals [8].

3 Forecast verification metrics

A robust verification scheme is essential to gain insight into how the developed forecasting product performs. Both from the point of view of a developer which aim to increase the skill of the prediction but also from the user which may utilize the verification score to assess the quality of a given forecast [1]. In the context of Sea Ice forecasting, a spatial field of continuos or discrete sea ice concentration is predicted, the latter being the case for the current work. Given the uneven distribution intra sea ice concentration classes as well as sea ice compared to ice free open water, simply comparing pixels for correctness would be biased by the large portion of open water and result in difficult to interpret values devoid of physical reasoning. Furthermore, as the rate of maritime activity such as commercial shipping increases in the Arctic due to the sea ice decline [7], having user relevant metrics can aid and alleviate the risks surrounding Arctic navigation. As such, several studies have proposed calculating the position of the ice edge as a user relevant metric which also provides information of the distribution of the Sea Ice Concentration [3, 5, 4]. However, there is no agreement with how to best calculate the position of the Ice Edge, with the currently available metrics posing different advantages/disadvantages [14, 11]. For the purpose of this thesis, The ice edge position and length will be calculated according to [11, Melsom 2019 et.al], whereas the IIEE originally proposed by Goessling. H. [5] will also be utilized.

Syk udokumentert påstand, må modereres

3.1 Defining the Ice Edge

The ice edge for a given Sea Ice Concentration product is derived on a per pixel basis, and defined as the grid cells which meet the condition

$$c[i,j] \geq c_q \wedge \min(c[i-1,j], c[i+1,j], c[i,j-1], c[i,j+1]) < c_e \tag{1}$$

i.e. a pixel is marked as a ice edge pixel if the current pixel itself is larger than some given concentration threshold c_e and the minimum of the pixel's 4-neighbors is less than the same threshold. Moreover, the marked grid cells each contribute to the total length of the ice edge, with each pixel's length contribution determined based on the number neighbors also marked as an ice edge pixel. Consequently, a neighborless pixel is assumed to yield a contribution the length of the diagonal to the ice-edge ($l = \sqrt{2}s$) where s is the side length of the pixel. A pixel with one neighbor a contributes a mixed horizontal - diagonasl length $l = \frac{s+\sqrt{2}s}{2}$. Finally a pixel with two or more neighbors contributes with a pixel side-length $l = s$.

3.2 Integrated Ice Edge Error

The Integrated Ice Edge Length (IIEE) is an error metric which compares the forecast to some ground truth target [5]. The metric is defined as

$$\text{IIEE} = \text{O} + \text{U} \tag{2}$$

where

$$\mathrm{O} = \int_A \max(c_f - c_t, 0) dA \tag{3}$$

and

$$\mathrm{U} = \int_A \max(c_t - c_f, 0) dA \tag{4}$$

with c_t and c_f being the target and forecast concentration respectively, attaining a value of 1 if the concentration for a given pixel i above a set threshold, and 0 elsewhere. From the definition of the metric, it can be seen that the IIEE is a sum of the forecast overshoot and undershoot compared to the ground truth target. For the current work, the IIEE is an easily interpreted metric as it quantifies the total forecast error and reports on the error spatially.

Furthermore, the IIEE can be combined with the length of the Ice Edge which was derived in the previous section 3.1. Thus, the metric is seasonally normalized, assuming that the IIEE and Ice Edge Length is seasonally correlated.

4 Impact of increased resolution on the IIEE

From both the definition of Equation (1) and (2), it can be seen that there is a dependance on the number of pixels which constitutes the ice edge. However, what effect would a change of resolution, i.e. change in number of pixels, have on the IIEE / ice edge length ratio? To answer this question, the "two day" ice edge targets where compared against persistance on all valid two day forecasts samples for the period 2019 - 2021. Furthermore, the original target resolution of 1km will be assessed, as well as a regrided product downscaled onto a 10km resolution.

The Pearson correlation coefficient was computed directly from the DataFrames containing the metrics for both 1km and 10km resolution, using the statistical Python package Pandas [16, 10]. For clarity, the correlation between mean_length and IIEE for 1km and 10km was calculated. From the resulting computations, the correlation coefficient gets reported as $r_{mean\_length} = 0.9458$ and $r_{IIEE} = 0.9995$. Furthermore, the IIEE divided by the mean length correlation is reported as $r_{normalized\_IIEE} = 0.9747$. Finally, Figure (1) display the mean monthly IIEE computed along the inspected three year period.

By inspecting Figure (1) in conjunction with the reported correlation coefficients, it can be seen that increasing the working resolution of the dataset from 1km to 10km has a negligible impact on the reported metrics. Though the 1km ice edge is about 13 times longer than the 10km, coarser resolution ice edge, it does not impact the overall stability of the normalized IIEE metric. However, the normalized IIEE is reduced for the 1km grid, as a consequence of the increased ice edge length due to resolution. Note that the from the definition of the IIEE given in Equation (2), the unnormalized metric is dependent on pixel spatial size. Thus, the number of pixels constituting the sum is inverse proportional with the size of each pixel, hence keeping the stability of the values. However, a discrepancy may arise due to the coarser resolution pixels covering a larger area, thus losing out on the fine details. However, for the current work, the ration of 1km IIEE and 10km IIEE is 0.9986, indicating that they are close to equal.

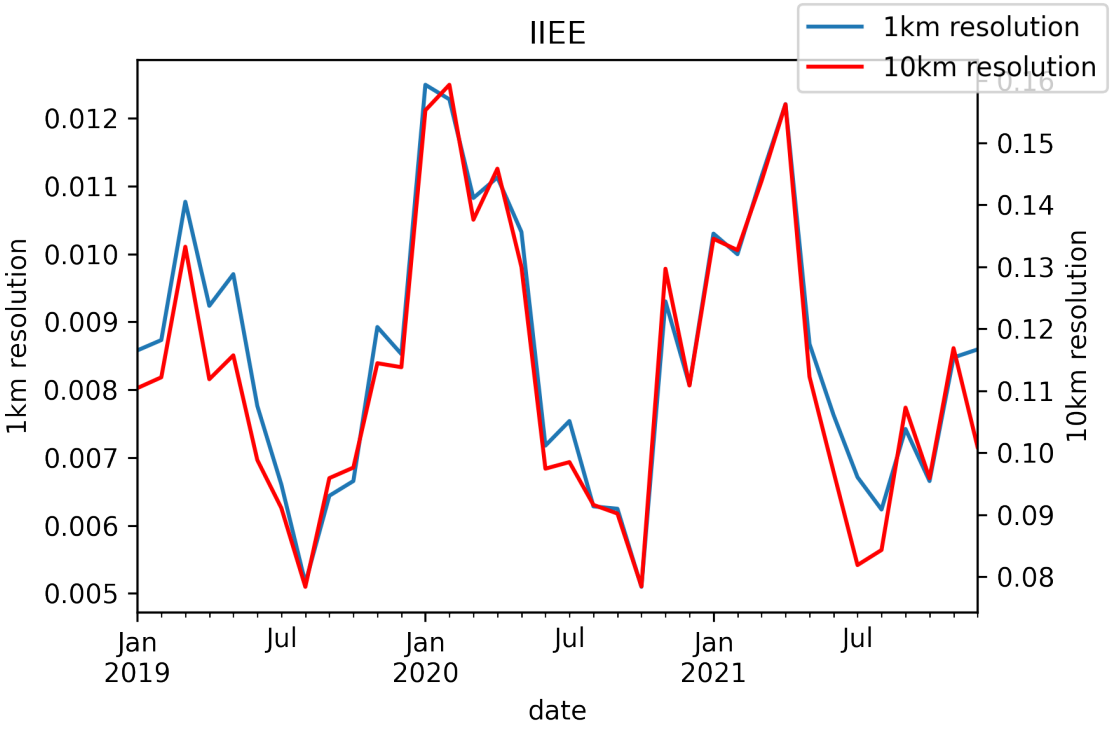


Figure 1: Monthly mean of Normalized IIEE spanning 2019 - 2021

5 Model development log and notes

5.1 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 [2, 12]. 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 (2). 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.

5.2 Model Architecture

The model architecture follows an encoder - decoder structure, commonly referred to as a U-NET [15] 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 [6], as the activation function used throughout the network is the ReLU function [13]. The final output of the model is a (1920, 1840, 7) tensor containing softmaxed probabilities along its final axis.

5.3 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

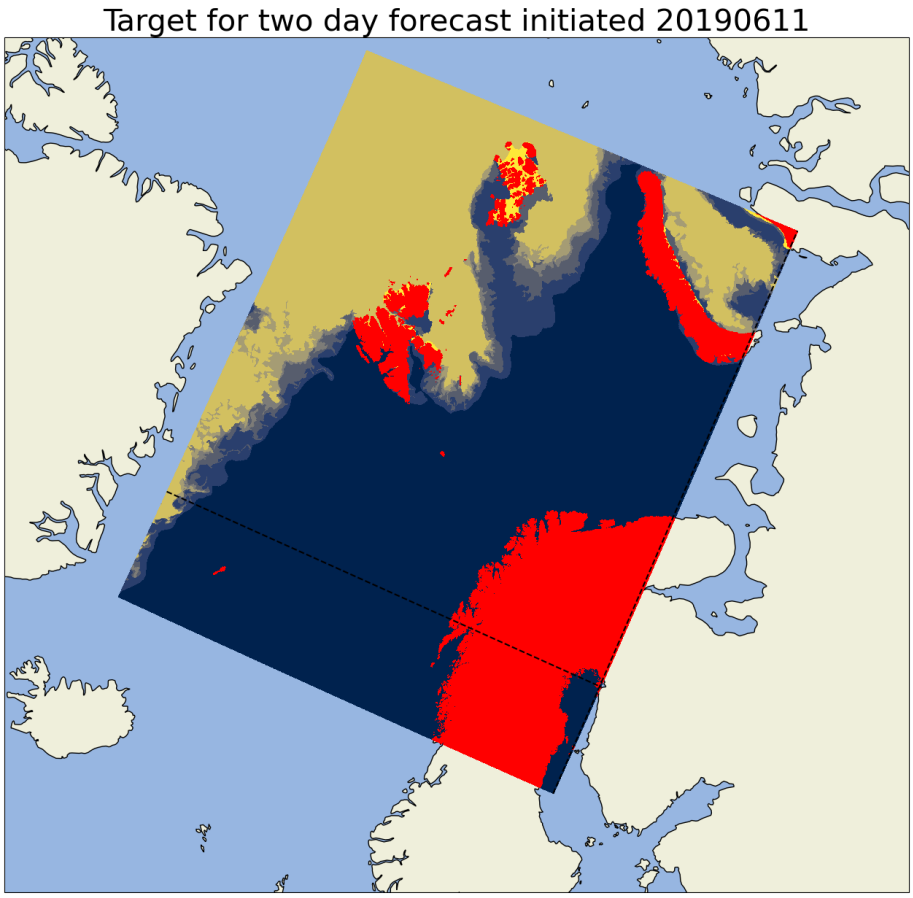


Figure 2: Example sample displaying an Ice Chart on a 1km Arome Arctic projection. Note the horizontal and vertical dashed black line which indicate the domain subsection used by the UNET

$$CE = -\sum_{i}^{C} y_i \, log \, (\hat{y}_i) \tag{5}$$

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 (5). 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 (3).

By inspecting Figure (3), two observations can be made. The first observation is regarding how the model complexity affects how it fit to the data. By comparing Figure (3a) with (3b), 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

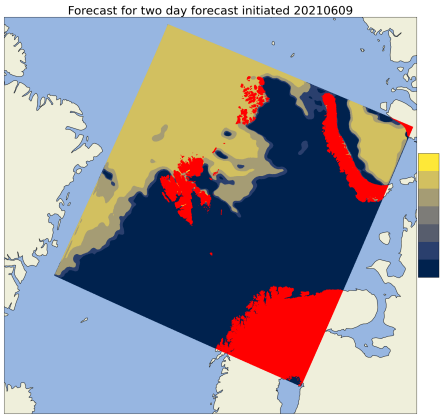
5.4 FocalLoss

The focal loss is derived as a generalization of the Cross Entropy Loss listed in Equation (5). 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 [9]. Mathematically, focal loss is defined as

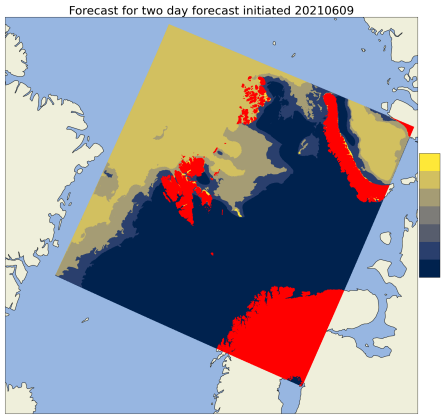
Include figure showing focal loss output, discuss implications of using this loss function

$$\mathrm{FL} = -\sum_{i}^{C} \alpha_i (1 - \hat{y}_i)^{\gamma} y_i \, log \, (\hat{y}_i) \tag{6}$$

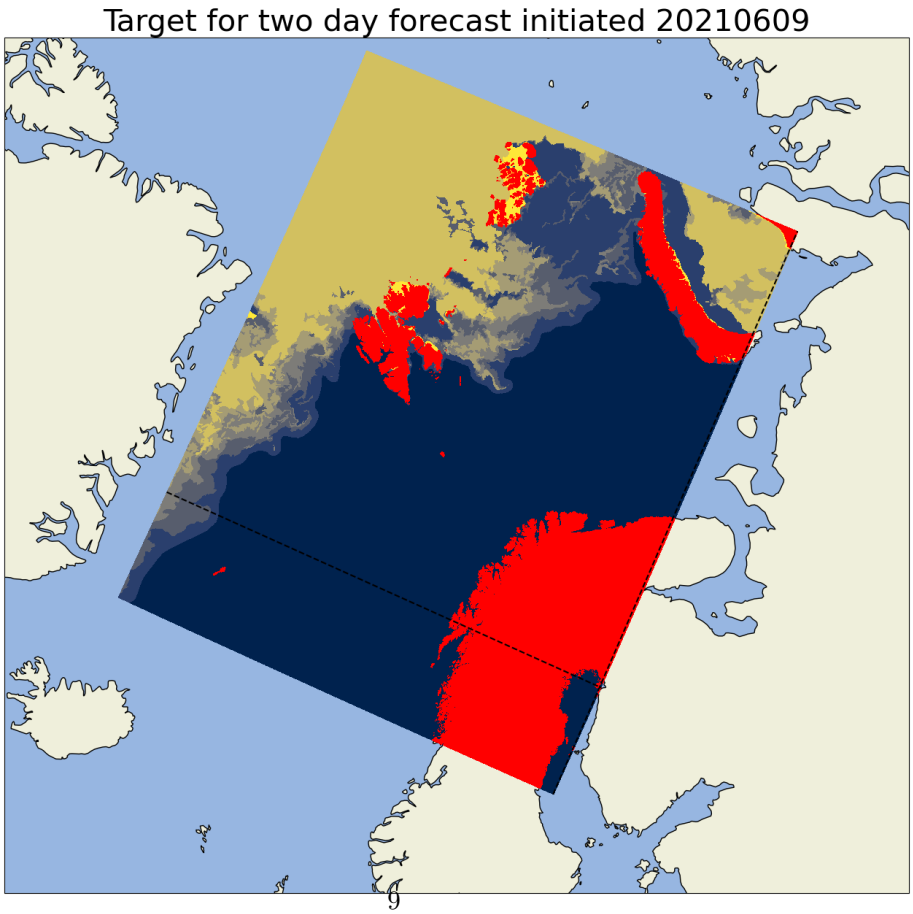
where α is a balancing parameter, γ is the focusing parameter ($\gamma = 0 \rightarrow CE$), with the rest similar as Equation (5).



(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 3: Example forecast attempt made by model_512 and model_1024 09-06-2021

By inspecting Equation (6), it can be seen that predictions that the model is quite confident in making, i.e. $\hat{y}_i \to 1$ 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 (3). 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.

5.5 Cumulative probability distribution model

5.5.1 Separate convolutionial layers as output

5.5.2 Single convolutional layer for multiple outputs

Discuss difference in dataloader, same dataset is used differently

Data exists, start writing

Not started on

6 Methodology

6.1 unet

The U-Net architecture was originally proposed by Ronneberger, Fischer, and Brox in 2015 [15] dette er en endring her endrer jeg litt til enda en endring og enda en siste endring

References

[1] B. Casati et al. "Forecast verification: current status and future directions". In: *Meteorological Applications* 15.1 (2008), pp. 3–18. DOI: 10.1002/met.52.

[2] Frode Dinessen, Bruce Hackett, and Matilde Brandt Kreiner. *Arctic Ocean - Sea Ice Concentration Charts - Svalbard and Greenland*. en. 2015. DOI: 10.48670/MOI-00128.

[3] Dmitry S. Dukhovskoy et al. "Skill metrics for evaluation and comparison of sea ice models". In: *Journal of Geophysical Research: Oceans* 120.9 (Sept. 2015), pp. 5910–5931. DOI: 10.1002/2015jc010989.

[4] H. F. Goessling and T. Jung. "A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts". In: *Quarterly Journal of the Royal Meteorological Society* 144.712 (Apr. 2018), pp. 735–743. DOI: 10.1002/qj.3242.

[5] H. F. Goessling et al. "Predictability of the Arctic sea ice edge". In: *Geophysical Research Letters* 43.4 (Feb. 2016), pp. 1642–1650. DOI: 10.1002/2015gl067232.

[6] Kaiming He et al. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. 2015. DOI: 10.48550/ARXIV.1502.01852.

[7] Joshua Ho. "The implications of Arctic sea ice decline on shipping". In: *Marine Policy* 34.3 (May 2010), pp. 713–715. DOI: 10.1016/j.marpol.2009.10.009.

[8] JCOMM Expert Team on Sea Ice. *Sea-Ice Nomenclature: snapshot of the WMO Sea Ice Nomenclature WMO No. 259, volume 1 – Terminology and Codes; Volume II – Illustrated Glossary and III – International System of Sea-Ice Symbols)* . 2014. DOI: 10.25607/OBP-1515.

[9] Tsung-Yi Lin et al. "Focal Loss for Dense Object Detection". In: (Aug. 2017). arXiv: 1708.02002 [cs.CV].

[10] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

[11] Arne Melsom, Cyril Palerme, and Malte Müller. "Validation metrics for ice edge position forecasts". In: *Ocean Science* 15.3 (May 2019), pp. 615–630. DOI: 10.5194/os-15-615-2019.

[12] Malte Müller et al. "Characteristics of a Convective-Scale Weather Forecasting System for the European Arctic". In: *Monthly Weather Review* 145.12 (Dec. 2017), pp. 4771–4787. DOI: 10.1175/mwr-d-17-0194.1.

[13] Vinod Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *ICML*. 2010, pp. 807–814. URL: https://icml.cc/Conferences/2010/papers/432.pdf.

[14] Cyril Palerme, Malte Müller, and Arne Melsom. "An Intercomparison of Verification Scores for Evaluating the Sea Ice Edge Position in Seasonal Forecasts". In: *Geophysical Research Letters* 46.9 (May 2019), pp. 4757–4763. DOI: 10.1029/2019gl082482.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: (May 2015). arXiv: 1505.04597 [cs.CV].

[16] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.