AI4SEAICE: COMPARING LOSS REPRESENTATIONS FOR SAR SEA ICE CONCENTRATION CHARTING

Andrzej S. Kucik* andrzej.kucik@esa.int
Andreas Stokholm* stokholm@space.dtu.dk

ABSTRACT

Sea ice charts, an important tool for navigation in the Arctic, are to this day manually drawn by professional ice analysts. The primary descriptor of the charts – the Sea Ice Concentration (SIC) – indicates the ratio of ice to open-water in an area. Automating the SIC chart production is desired but the optimal representation of the corresponding machine learning task is ambivalent. Here, we explore it with either regressional or classification objective, each with two different (weighted) loss functions: Mean Square Error and Binary Cross-Entropy, and Categorical Cross-Entropy and the Earth Mover's Distance, respectively. While all perform well, the regression-based models achieve higher numerical similarity to the ground truth, whereas classification results in more visually pleasing and consistent charts. Weighting the loss functions improves the performance for intermediate classes at expense of open-water and fully-covered sea ice areas.

1 INTRODUCTION

The Arctic oceans are experiencing diminishing sea ice covers due to global warming (Perovich et al., 2020), rapidly making them more accessible to new commercial opportunities and resource extraction (Funk, 2009). The opening of the Northern Sea Route linking the far East to Europe and the coasts of North America across the Arctic could disrupt global supply chains by decreasing the shipping time and costs substantially (Bekkers et al., 2017). Thus, high-resolution sea ice charts outlining the ever-changing sea ice conditions are indispensable.

Synthetic Aperture Radar (SAR) images offer high-resolution, versatile measurements independent of sun illumination and clouds. However, they are difficult to interpret for untrained eyes; the radar backscatter is dependent on surface properties, and ambiguities between open water and sea ice are common. Professional sea ice analysts at the Danish Meteorological Institute (DMI) interpret these signatures to manually draw charts of the sea ice conditions around Greenland (Saldo et al., 2020). This is a resource- and time-consuming endeavour, inspiring the need for automation.

Using Convolutional Neural Networks (CNN) and SAR for automatic sea ice charting was originally published by Wang et al. (2016). Newer advancements apply SAR and Passive Microwave Radiometer CNN data fusion models in the Automatic Sea Ice Products (ASIP) project (Malmgren-Hansen et al., 2020; 2021). de Gélis et al. (2021) is the first paper to adopt the U-Net CNN architecture (Ronneberger et al., 2015) to map sea ice using downscaled SAR data. Finally, Stokholm et al. (2022) demonstrated that increasing the receptive field of the U-Net improves predictions. In this study, we analyse different representations of the model's objective to better capture the intuition of human-labelled sea ice charts.

The Sea Ice Concentration (SIC) is the percentage ratio of sea ice to open-water for an area, discretised into 11 10%-bin classes ranging from 0% (open-water) to 100% (fully-covered sea ice). Training the model with classification loss functions, such as categorical Cross-Entropy (CE), may not reflect the inter-class relationship, penalizing the model disproportionately, e.g. if 60% is predicted instead of 50%, the penalty is equal to a prediction of 0%. Indeed, Stokholm et al. (2022) showed that (weighted) CE excels at predicting open-water, but it often remained inadequate for the intermediate SIC classes (10-90%). Therefore, here we compare CE to Mean Squared Error (MSE), Binary Cross-Entropy (BCE) and squared Earth Mover's Distance (EMD^2) losses. These four loss representations allow us to express the learning objective as either classification, regression (linear or logistic), or transportation problems, respectively, and wage between their relative advantages.

*Equal contribution, alphabetically ordered.

2 DATA

The experiments are conducted with the European Space Agency's (ESA) AI Ready Earth Observation (AIREO) sea ice dataset, AI4Arctic / ASIP v2 (ASID-v2) (Saldo et al., 2020), compiled by DMI, the Technical University of Denmark (DTU), and Nansen Environmental and Remote Sensing Center (NERSC). It includes 461 co-located and georeferenced scenes distributed across the Greenland coast, between March 14, 2018, to May 25, 2019. We utilise the two layers; the Sentinel-1 dual polarised HH and HV SAR images and the corresponding professionally drawn SIC chart (see Fig. 1 for illustration).

2.1 SENTINEL-1 SAR

The Sentinel-1 satellites use a C-band SAR with a 5.405 GHz frequency (Geudtner & Torres, 2012). We use the medium resolution level 1 ground range detected data product, measured in the Extra-Wide operational mode with a pixel spacing of 40 m and resolution of 93×87 m (range $\times$ azimuth). Following the findings of (Stokholm et al., 2022), we applied NERSC SAR noise correction to the data. This denoising technique is described in (Park et al., 2018; 2019).

2.2 SEA ICE CHARTS

Sea ice charts are snapshots of the ice conditions at acquisition time, drawn as polygons of fairly homogeneous sea ice areas, and based on the professional interpretation of the subsequent SAR images. Charts follow the SIGRID3 (Sea Ice GeoReferenced Information and Data) created by the World Meteorological Organization. The process of shaping polygons and assigning SIC, albeit steered by common guidelines, is a creative and individual interpretation – studies suggest that ice analysts assign concentrations that can vary on average 20% and up to 60% (Karvonen et al., 2015). Intermediate classes are particularly difficult to assess, and regions with potential for high maritime activity, such as the edge of the sea ice cover, receive more attention. Nevertheless, we treat the human-made SIC labels as ground truth.

For data preprocessing, the procedure in Stokholm et al. (2022) is applied. 306 training and 23 test scenes are selected (9:1 split), the latter deemed particularly difficult by DMI and mirror the training set class distribution. Classes: 0 (open-water), 10 (100% sea ice), and 11 (masked pixels) are by far the most represented in the dataset. The remaining ones are relatively equally distributed.

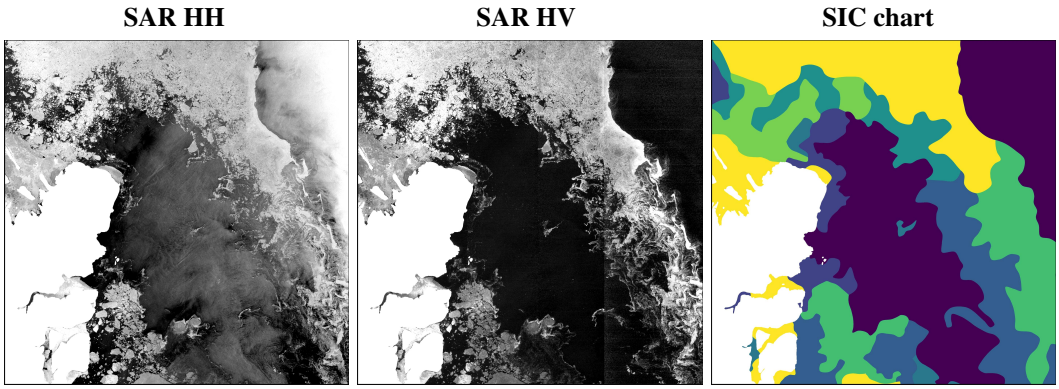

Figure 1: Sample scene – Fram Strait, Northeast Greenland. Scene acquired August 22, 2018. **a-b**) HH and HV SAR images, respectively. **c**) corresponding human-drawn SIC chart (ground truth).

3 MODEL IMPLEMENTATION

We train a U-Net (Ronneberger et al., 2015) with 8 encoder-decoder blocks (16 and 32 filters in the first two and 64 filters in the remaining) for 100 epochs, each with 500 training steps, using the Adam optimiser. A batch consists of 32 patches of 768×768 pixels, randomly cropped from training scenes. The crop sampling and augmentation procedure from Stokholm et al. (2022) are applied.

Models are evaluated using accuracy, average class accuracy (excluding the masked pixel class), and the R^2 metric – often better in capturing the continuity of the distance between predictions and the ground truth. All the experiments were carried out on two Nvidia TeslaV100 SXM2 32 GB GPUs using PyTorch 1.8 library. The code for the experiments is available at (Stokholm & Kucik, 2021).

4 LOSS REPRESENTATION

In this paper, we study how the choice of loss function representation affects the performance of the model with respect to the given objective. Predicting the percentage concentration of ice in seawater is a seemingly regressional problem, demanding the distance between the predicted and expected scalars minimised in a geometric sense using, for instance, the MSE.

Alternatively, assuming that the sea surface may only consist of water or ice particles, we may view SIC as the (percentage) probability of sampling an ice molecule from a given sea surface region. From this perspective, it is more natural to tackle the problem as a logistic regression one, treating the SIC values as confidence levels of sampling ice rather than water. This can be executed with BCE loss, for example.

The main issue with the two approaches above is that SIC does not represent a continuous value. Furthermore, the discrete classes are not derived by quantifying well-defined physical properties. Instead, they reflect the intuition and experience of professional ice analysts. Therefore, sea regions with a fixed ice concentration can still be assigned to varying SIC classes, and also the size or shape of the region itself can vary. To illustrate this, let us consider the continuous value of human age and the idea of someone being either young or old. A person of age 30 can be classified as both young or old, depending on their characteristics and on who is judging. Analogously, it may be more appropriate to represent SIC segmentation as a classification problem, using CE loss, for example.

4.1 EARTH MOVER'S DISTANCE

CE classification has a perceptible flaw in this setting – it assumes the lack of correlation between class predictions, which is not the case here. Let us recall the human age example: if we classify people as babies, kids, adolescents, adults, or elderly, then clearly the elderly group must be more correlated with the adults rather than the babies category. Therefore, a loss function that leverages the inter-class relationship is recommended in this instance.

The *Earth Mover's* or *Wasserstein's* EMD distance between two distributions P and Q may be defined as

$$\mathrm{EMD}(P,\, Q) := \inf_{\gamma \in \Pi(P,Q)} \mathbb{E}_{(\mathrm{x},\mathrm{y})\sim\gamma} \|\mathrm{x} - \mathrm{y}\|, \tag{1}$$

where $\Pi(P, Q)$ is the set of all joint distributions whose marginals are P and Q, and $\| \cdot \|$ is some norm on the space on which P and Q are defined (Arjovsky et al., 2017). The reason behind the EMD name is that we can view it as the cost of optimally transporting "earth mass" between x and y (Monge, 1781). An obstacle in using the EMD as the objective function is the infimum in equation 1 is often intractable, so certain assumptions are made to simplify it. Frogner et al. (2015) and Martinez et al. (2016) approximate EMD for supervised multi-class multi-label learning, Arjovsky et al. (2017) use weight clipping and Gulrajani et al. (2017) use gradient penalty EMD-based generative adversarial networks. Under certain conditions (Levina & Bickel, 2001), EMD is equivalent (up to a normalisation constant) to the Mallows distance:

$$\mathrm{EMD}(P,\, Q) = \| \,\mathrm{CDF}(P) - \mathrm{CDF}(Q)\|, \tag{2}$$

where CDF is the Cumulative Density Function. Those conditions are satisfied for ordered-, single-class learning (Hou et al., 2016). The authors recommend using the squared EMD (EMD^2) for faster convergence. Thus, in the Euclidean setting, we can represent the SIC (per-pixel) EMD^2 loss as

$$\mathrm{EMD}^2(\boldsymbol{y},\, \hat{\boldsymbol{y}}) = \sum_i \left(\sum_{j=1}^{i} y_j - \sum_{j=1}^{i} \hat{y}_j \right)^2, \tag{3}$$

where $\boldsymbol{y}_j$ and $\hat{\boldsymbol{y}}_j$ are the j^{th} elements of true and predicted label distributions, $\boldsymbol{y}$, $\hat{\boldsymbol{y}}$ respectively, and the outer sum is taken over all available classes i.

4.2 LOSS AND THE TRAINING DATA IMBALANCE

As mentioned, the dataset is inherently imbalanced. This can be mitigated by associating a penalty with each class, i.e. multiplying individual pixel losses by a weight factor w. We choose a weighting factor as the inverse class frequency and normalise it (i.e. $\sum w = 1$).

Another issue is the variable number of valid (i.e. not masked pixels) in the training samples, discarded during the total loss computation. If we view each sample as a batch of pixels, this is equivalent to having a variable batch size at each iteration. To harmonise them, we discount each example's loss by the inverse of the number of its valid pixels. Alternatively, we may picture this as associating a weight 1 with each valid and 0 with each masked pixel, and then calculate the per example loss as a weighted average of per-pixel losses.

To summarise, if we have an example $\mathbf{X}$, associated labels $\boldsymbol{Y} = (Y_{i,j})_{i,j}$, and model predictions $\hat{\boldsymbol{Y}} = (\hat{Y}_{i,j})_{i,j}$, then we define the per example loss function as

$$\text{Loss}(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) := \frac{\sum_{i,j} w(Y_{i,j}) \, \text{loss}(Y_{i,j}, \hat{Y}_{i,j})}{v(\mathbf{X})}, \tag{4}$$

where $w(Y_{i,j}$ is the weight associated with class $Y_{i,j}$, "loss" is per-pixel loss function (in this case either MSE, BCE, CE, or EMD^2), $v(\mathbf{X})$ is the number of valid pixels of $\mathbf{X}$, and summed across all valid pixels (i, j). Without the loss of generality, we assume that the unweighted loss is obtained simply by setting w to be uniformly equal to 1 for valid classes.

Weighting the BCE and EMD^2 reduces their standard value range by several orders of magnitude, degrading the effectiveness of gradient descent, which results in the model's performance being reduced by a double-digit percentage. To perform a more adequate comparison, we keep them in line with other loss choices by multiplying them by factors of 100 and 1000, respectively.

5 RESULTS AND DISCUSSION

The R^2-scores and the accuracy values for each of the loss functions are summarised in Table 1. Since R^2 values are similar for all the losses, we repeated the experiments 5 times to reduce the dependence on weights' initial values. The distributions of the resulting R^2 scores and accuracies are shown in Figs. 2-3) in Appendix A. In each loss function category, we select the model with the best R^2 score for more detailed analysis. The accuracy is further split into the: overall-, average per class-, open-water-, intermediate SIC average-, and 100% SIC accuracy. The accuracy levels for each of the individual classes is fully-catalogued in Table 2 in Appendix A.

Table 1: Test R^2 scores and accuracy values (in %). w indicates a weighted loss. Best results in **bold**. ACA = Average Class Accuracy; AICA = Average Intermediate Class Accuracy.

Loss	R^2	Acc	ACA	0% acc	AICA	100% acc
MSE	**93.12**	65.81	31.90	83.60	27.18	80.41
BCE	92.64	68.92	36.25	90.30	24.68	83.76
CE	90.31	**74.08**	33.54	**97.49**	21.85	**95.06**
EMD^2	91.37	73.34	33.26	96.23	22.22	93.92
w MSE	89.93	51.86	34.84	68.25	29.26	51.05
w BCE	90.10	56.30	34.83	73.64	29.97	56.10
w CE	87.71	62.92	**38.56**	85.93	**32.18**	57.29
w EMD^2	89.78	61.53	36.89	80.79	29.34	65.07

Generally, the unweighted loss functions result in a higher R^2 score, overall-, open-water-, and 100% SIC accuracy. This is expected given the class imbalance in the dataset (skewed towards 0% or 100% SIC). On the other hand, in the weighted setting, we observe a better average per class- and intermediate SIC accuracy, but with lower 0% and particularly the 100% accuracies. Regardless of whether the loss weighting is applied, the models optimised with respect to MSE and BCE obtain the highest R^2-score, while the accuracy is superior when CE or EMD^2 are used.

It appears that using the regressional losses results in more uniform performance across the individual classes, whereas the classification objectives seem to prioritise particular classes, e.g. EMD^2 has accuracy close to 0 for 3 of the classes but up to 38% in other intermediate classes. Weighting the loss functions seem to somewhat mitigate this tendency. Additionally, Figs. 2-3 illustrate that the R^2 and accuracy performance variability are largest for the weighted functions, indicating difficulties in learning the underrepresented classes despite equal class weighting.

Figs. 4-7 in Appendix B illustrate four selected tests scenes, including the HH and HV SAR images, the professional labelled SIC chart (ground truth), and the corresponding model outputs with associated R^2-scores. The scenes in Figs. 4-5 are from Scoresbysund, East Greenland, and Figs. 6-7 are from the Fram Strait in Northeast Greenland. Fig. 4 exhibits a large 100% SIC area, with landfast sea ice in the fjord, which often appears darker than typical sea ice and is thus difficult to distinguish from open water. The scene also contains a sea ice edge with abundant and detailed intermediate SIC, as well as a bright near-range field in the right side portion of the SAR image. The detailed sea ice edge is predicted well except for the weighted MSE and BCE. The regressional models are deficient in the landfast ice area, while MSE objectives produce noise in the bright near-range field. Fig. 5 also has landfast ice in the fjord but lower SIC at the mouth. Here the weighted losses perform well on the low SIC while the unweighted ones correctly classify the 100% SIC in the fjord, with the MSE- and EMD^2-based models performing best. CE-optimised models struggle with the landfast ice, substantially lowering their scores. The scene in Fig. 6 contains large variations in SIC with multiple transitions between sea ice and open water. The unweighted setting performs better in both: along the ice edge with high detail level and in the 100% SIC areas. Weighted losses tend to make blurry predictions along low SIC edges. The final scene in Fig. 7 exhibits sea ice close to a meandering coast with a variety of islands and SICs. Here, the unweighted losses over-predict 100% SIC, whereas the weighted versions score higher with more adequate SICs. Again, the MSE-based models are impacted by the bright near-range field.

6 Conclusion and Future Work

This study presents the comparison of how different loss function representations affect model performances with respect to predicting the discrete percentage concentration of ice in seawater. Four analogues: Mean Square Error and Binary Cross-Entropy for regressional representation and Categorical Cross-Entropy and the squared Earth Mover's Distance for classification, are considered, with an optional class weighting scheme.

Generally, the models score well on the metrics and generate ice charts with a strong resemblance to the ground truth. The unweighted-loss models' performance is superior but they tend to over-predict 100%. Their weighted equivalents take precedence for intermediate SICs, but they create less detailed, more blurry charts. Given the modest improvement in the intermediate classes, weighting the loss functions may not justify sacrificing the performance in the 0% and 100% classes.

The loss functions can be quantitatively summarised as follows. For the R^2-score, from best to worst: MSE, BCE, EMD^2 and CE. For accuracy the order is reversed: CE, EMD^2, BCE and MSE. Despite regressional models achieving higher R^2-scores, predictions are inferior in predicting 100% SIC in areas with landfast ice and produce charts with large SIC variations, not observed in the ground truth. In addition, the MSE-based models have noisy outputs. Although the SIC representation is naturally continuous, classification objectives outperform the regressional models in terms of qualitatively assessing the predictions, which is most important for ice chart users. In this setting, EMD^2 achieves higher R^2 scores than CE, and is more reliable with landfast ice, but with otherwise no improvements in accuracy. The more numerically similar predictions of EMD^2 could form the basis for further model exploration.

Future work could involve identifying an approach to improve intermediate SICs performance while retaining high accuracies in 0% and 100%. Balancing the dataset, sampling more intermediate SICs or decoupling them from the 0% and 100% classes during the training, could provide improvements, regardless of the choice of the loss function.

Acknowledgments

The authors would like to thank DTU Space and ESA Φ-lab for their support on this research.

REFERENCES

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. arXiv:1701.07875, Jan 2017.

Eddy Bekkers, Joseph F. Francois, and Hugo Rojas-Romagosa. Melting Ice Caps and the Economic Impact of Opening the Northern Sea Route. *The Economic Journal*, 128(610):1095–1127, Aug 2017. doi:10.1111/ecoj.12460.

Iris de Gélis, Aurélien Colin, and Nicolas Longépé. Prediction of Categorized Sea Ice Concentration From Sentinel-1 SAR Images Based on a Fully Convolutional Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5831–5841, 2021. doi:10.1109/jstars.2021.3074068.

Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya, and Tomaso A Poggio. Learning with a Wasserstein Loss. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper/2015/file/a9eb812238f753132652ae09963a05e9-Paper.pdf.

McKenzie Funk. Arctic Landgrab. *National Geographic*, May 2009. URL https://www.nationalgeographic.com/magazine/article/healy.

Dirk Geudtner and Ramon Torres. Sentinel-1 system overview and performance. In *2012 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Jul 2012. doi:10.1109/igarss.2012.6351191.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved Training of Wasserstein GANs. arXiv:1704.00028, Mar 2017.

Le Hou, Chen-Ping Yu, and Dimitris Samaras. Squared Earth Mover's Distance-based Loss for Training Deep Neural Networks. arXiv:1611.05916, Nov 2016.

Juha Karvonen, Jouni Vainio, Marika Marnela, Patrick Eriksson, and Tuomas Niskanen. A Comparison Between High-Resolution EO-Based and Ice Analyst-Assigned Sea Ice Concentrations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(4):1799–1807, Apr 2015. doi:10.1109/jstars.2015.2426414.

E. Levina and P. Bickel. The Earth Mover's distance is the Mallows distance: some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. IEEE Comput. Soc, 2001. doi:10.1109/iccv.2001.937632.

David Malmgren-Hansen, Leif Toudal Pedersen, Allan Aasbjerg Nielsen, Henning Skriver, Roberto Saldo, Matilde Brandt Kreiner, and Jørgen Buus-Hinkler. ASIP Sea Ice Dataset - version 1. doi:10.11583/DTU.11920416.v1, Mar 2020.

David Malmgren-Hansen, Leif Toudal Pedersen, Allan Aasbjerg Nielsen, Matilde Brandt Kreiner, Roberto Saldo, Henning Skriver, John Lavelle, Jøorgen Buus-Hinkler, and Klaus Harnvig Krane. A Convolutional Neural Network Architecture for Sentinel-1 and AMSR2 Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):1890–1902, Mar 2021. doi:10.1109/tgrs.2020.3004539.

Manuel Martinez, Monica Haurilet, Ziad Al-Halah, Makarand Tapaswi, and Rainer Stiefelhagen. Relaxed Earth Mover's Distances for Chain- and Tree-connected Spaces and their use as a Loss Function in Deep Learning. arXiv:1611.07573, Nov 2016.

Gaspard Monge. *Mémoire sur la théorie des déblais et des remblais*. 1781. ISBN OCLC:51928110.

Jeong-Won Park, Anton A. Korosov, Mohamed Babiker, Stein Sandven, and Joong-Sun Won. Efficient Thermal Noise Removal for Sentinel-1 TOPSAR Cross-Polarization Channel. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3):1555–1565, Mar 2018. doi:10.1109/tgrs.2017.2765248.

Jeong-Won Park, Joong-Sun Won, Anton A. Korosov, Mohamed Babiker, and Nuno Miranda. Textural Noise Correction for Sentinel-1 TOPSAR Cross-Polarization Channel Images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6):4040–4049, Jun 2019. doi:10.1109/tgrs.2018.2889381.

Donald Perovich, W. Meier, M. Tschudi, S. Hendricks, A. A. Petty, D. Divine, S. Farrell, S. Gerland, C. Haas, L. Kaleschke, O. Pavlova, R. Ricker, X. Tian-Kunze, M. Webster, and K. Wood. Arctic Report Card 2020: Sea Ice. Technical report, 2020. doi:10.25923/n170-9h57.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241. Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-24574-4_28.

Roberto Saldo, Matilde Brandt Kreiner, Jørgen Buus-Hinkler, Leif Toudal Pedersen, David Malmgren-Hansen, Allan Aasbjerg Nielsen, and Henning Skriver. AI4Arctic / ASIP Sea Ice Dataset - version 2. doi:10.11583/DTU.13011134, Oct 2020.

Andreas Stokholm and Andrzej Kucik. AI4SeaIce, 2021. URL `https://github.com/astokholm/AI4SeaIce.git`.

Andreas Stokholm, Tore Wulf, Andrzej Kucik, Roberto Saldo, Jørgen Buus-Hinkler, and Sine Munk Hvidegaard. AI4SeaIce: Towards solving ambiguous SAR textures in convolutional neural networks for automatic sea ice concentration charting. *IEEE Transactions on Geoscience and Remote Sensing*, 2022. doi:10.1109/tgrs.2022.3149323.

Lei Wang, K. Andrea Scott, Linlin Xu, and David A. Clausi. Sea Ice Concentration Estimation During Melt From Dual-Pol SAR Scenes Using Deep Convolutional Neural Networks: A Case Study. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4524–4533, Aug 2016. doi:10.1109/tgrs.2016.2543660.

A APPENDIX - MODEL'S PERFORMANCE DISTRIBUTION

Table 2: Test accuracy values for individual classes. w indicates a weighted loss.

Loss	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
MSE	83.60	30.50	30.96	28.95	16.85	16.32	31.19	24.43	18.26	37.98	80.41
BCE	90.30	32.78	16.41	22.58	12.08	20.63	44.04	30.15	14.60	31.44	83.76
CE	97.49	15.06	4.47	33.87	27.85	12.47	4.10	32.38	15.20	30.98	95.06
EMD^2	96.23	0	29.83	38.00	33.92	0.10	1.18	28.45	12.42	31.77	93.92
w MSE	68.25	38.22	35.45	28.26	23.58	19.92	42.70	25.92	15.33	34.57	51.05
w BCE	73.64	32.59	33.27	35.34	21.82	22.21	30.21	19.79	16.54	41.66	56.10
w CE	85.93	64.91	21.63	22.98	25.76	26.35	13.07	35.13	24.81	46.36	57.29
w EMD^2	80.79	66.74	17.36	9.36	30.9	23.50	7.17	43.63	14.83	46.4	65.07

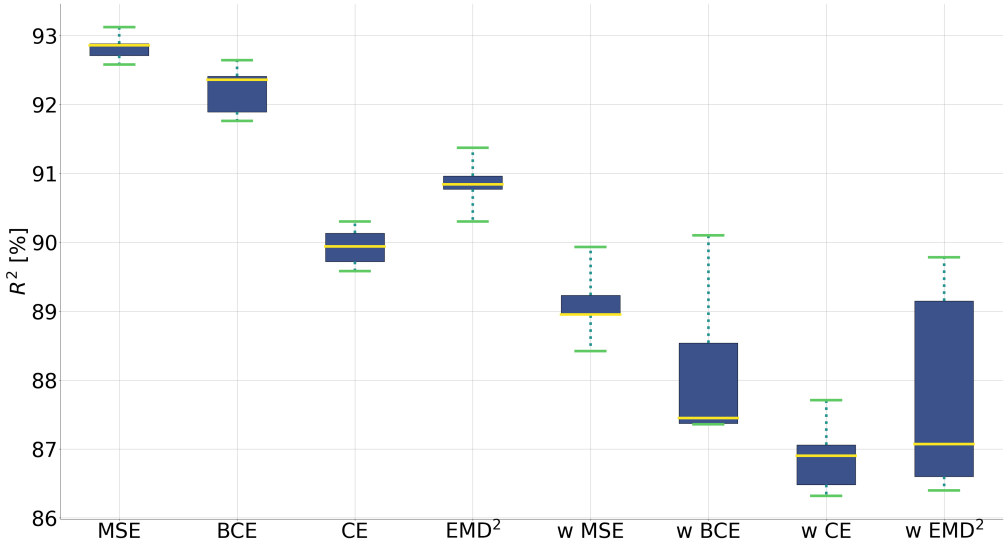

Figure 2: Test R^2 score distribution. w indicates a weighted loss.

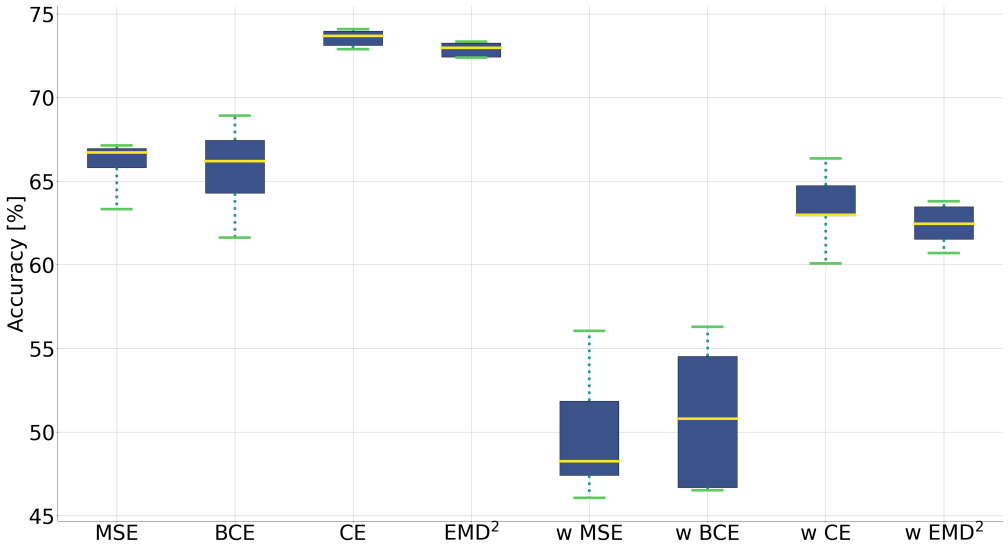

Figure 3: Test accuracy distribution. w indicates a weighted loss.

B APPENDIX - SAMPLE SCENES

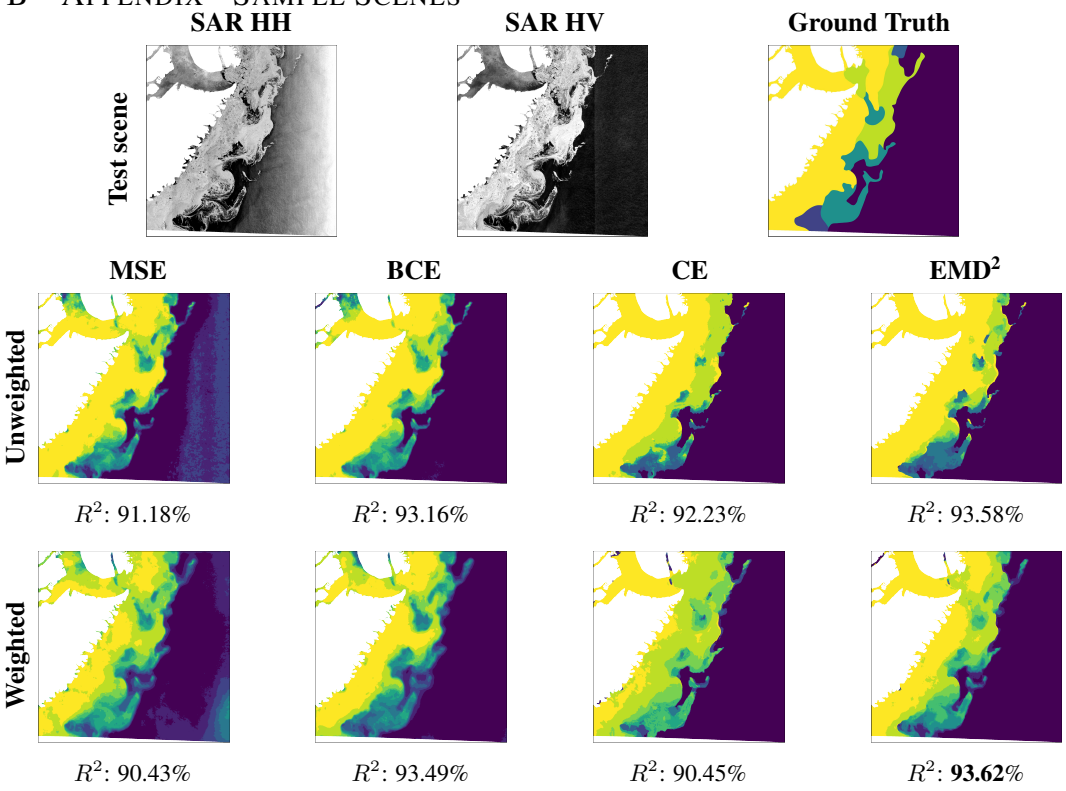

Figure 4: Scoresbysund, East Greenland. Scene acquired May 9, 2019.

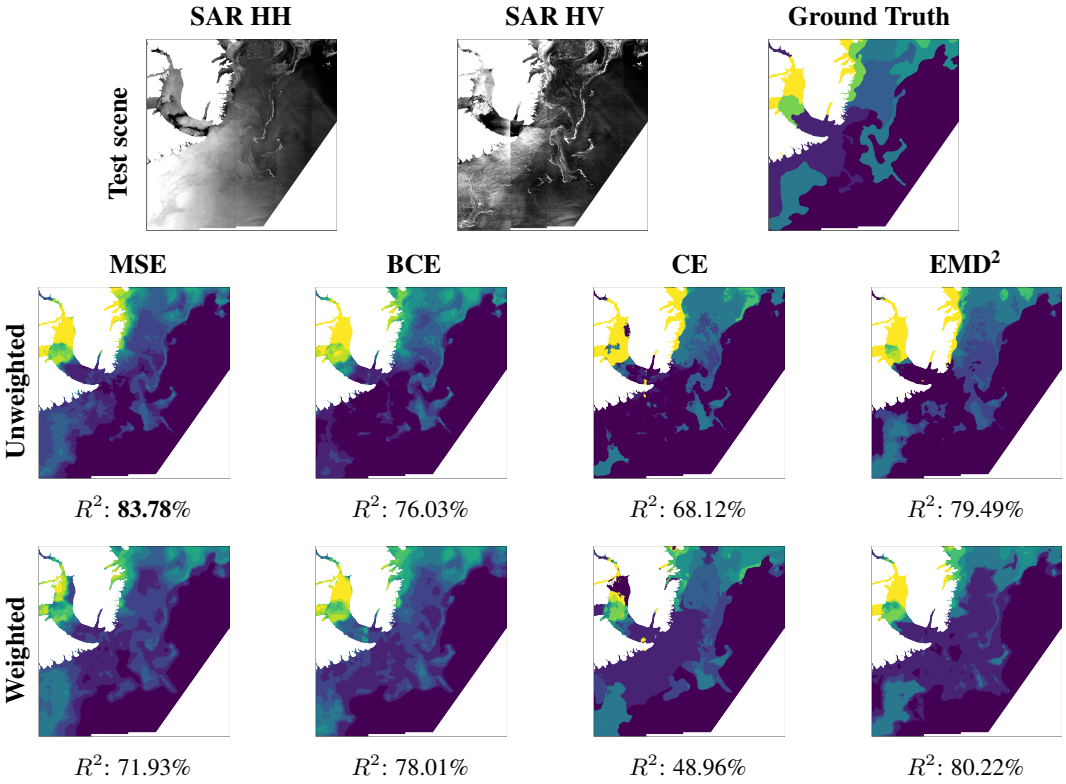

Figure 5: Scoresbysund, East Greenland. Scene acquired June 26, 2018.

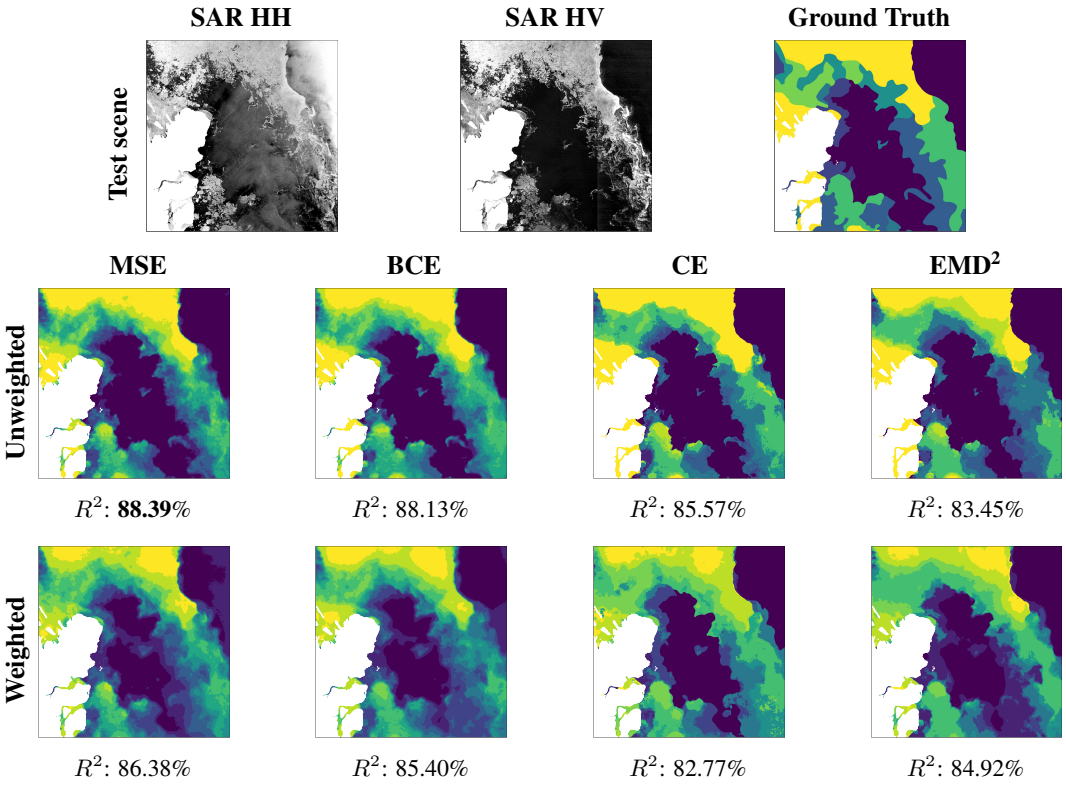

Figure 6: Fram Strait, Northeast Greenland. Scene acquired August 22, 2018.

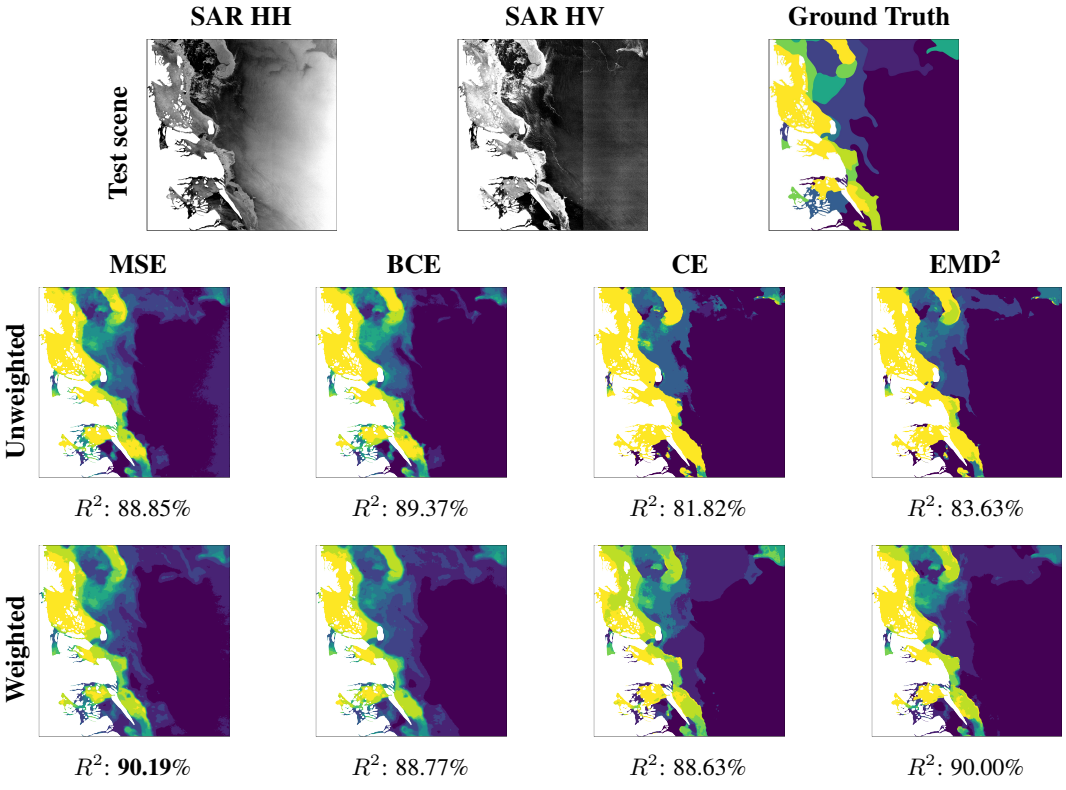

Figure 7: Fram Strait, Northeast Greenland. Scene acquired September 3, 2018