

COMPARISON OF CROSS-ENTROPY, DICE, AND FOCAL LOSS FOR SEA ICE TYPE SEGMENTATION

Rafael Pires de Lima¹, Behzad Vahedi¹, Morteza Karimzadeh¹

¹ Department of Geography, University of Colorado Boulder

ABSTRACT

Up-to-date sea ice charts are crucial for safer navigation in ice-infested waters. Recently, Convolutional Neural Network (CNN) models show the potential to accelerate the generation of ice maps for large regions. However, results from CNN models still need to undergo scrutiny as higher metrics performance not always translate to adequate outputs. Sea ice type classes are imbalanced, requiring special treatment during training. We evaluate how three different loss functions, some developed for imbalanced class problems, affect the performance of CNN models trained to predict the dominant ice type in Sentinel-1 images. Despite the fact that Dice and Focal loss produce higher metrics, results from cross-entropy seem generally more physically consistent.

Index Terms— Convolutional Neural Networks, loss function, sea ice, semantic segmentation, Synthetic Aperture Radar (SAR)

1. INTRODUCTION

The recent reduction in sea ice volume has led to increased potential for shipping in the Arctic (e.g., [1, 2]), which requires improved marine information services to ensure both safe transit and minimal environmental impact. Several national ice centers periodically produce sea ice charts, used by navigators to find the best and safest shipping routes. The generation of ice charts continues heavily dependent on sea ice experts that manually interpret remotely sensed data in conjunction with other information to create maps that indicate ocean conditions. Machine learning was identified as a methodology with significant potential for accelerating the generation of sea ice charts. Convolutional Neural Networks (CNNs), in particular, showed promising results for sea ice characterization using remotely sensed data as input, in both image classification and semantic segmentation frameworks. In classification, subsampled patches of remotely sensed images are labeled as water or other sea ice attribute, and then patches are mosaiced into a classified scene (e.g., [3, 4]). Most studies approaching sea ice mapping as a semantic segmentation task have used U-Net [5] variations. Ren et al.

[6] added attention components into U-Net that further captured the spatial correlation of features in individual feature maps and the interrelation of pixel values between different feature maps. Adding attention components improved accuracy when compared to the original architecture for a binary segmentation of open water and sea ice. Wang and Li [7] combined five U-Nets outputs to perform a binary segmentation of Sentinel-1 scenes into open water and sea ice. Stokholm et al. [8] explored expanding the number of convolutional blocks in a U-Net architecture to perform semantic segmentation of Sentinel-1 images to segment sea ice concentration. In [3, 4, 6, 7, 8], the models were trained using categorical or binary cross-entropy (CE) loss. Wang and Li [7] combined binary CE and Dice loss [9] to train their models. Despite these improvements, automated ice type classification remains unreliable for operational purposes, which can be attributed to multiple potential reasons, including the subjectivity of sea ice expert interpretation (labels) or loss functions used to train models. Sea ice type is generally an imbalanced problem, and adopting different losses has the potential to improve the performance of the models without requiring significant changes in model development and training. Recently, Kucic and Stokholm [10] evaluated the effect of different losses for sea ice concentration mapping. Here, we use a semantic segmentation framework and rely on the ExtremeEarth V2 dataset [11] to perform a set of experiments and evaluate what is the effect in performance when we train the models using different loss functions for ice type mapping.

2. MODEL ARCHITECTURE AND TRAINING STRATEGY

We approach sea ice type charting as a semantic segmentation task. To do so, we design a model architecture that uses the first three blocks of ResNet18 [12] as encoder, and a decoder based on the Atrous Spatial Pyramid Pooling (ASPP) module [13]. The encoder and decoder sum 4 M trainable parameters. We initialize the encoder with ImageNet [14] weights that are updated in training, while the decoder is randomly initialized. Our model takes as input Horizontal-Horizontal (HH), Horizontal-Vertical (HV), and incidence angle from Sentinel-1 images, and outputs ice type. We train the models using

This research is supported by the National Science Foundation under Grant No. 2026962.

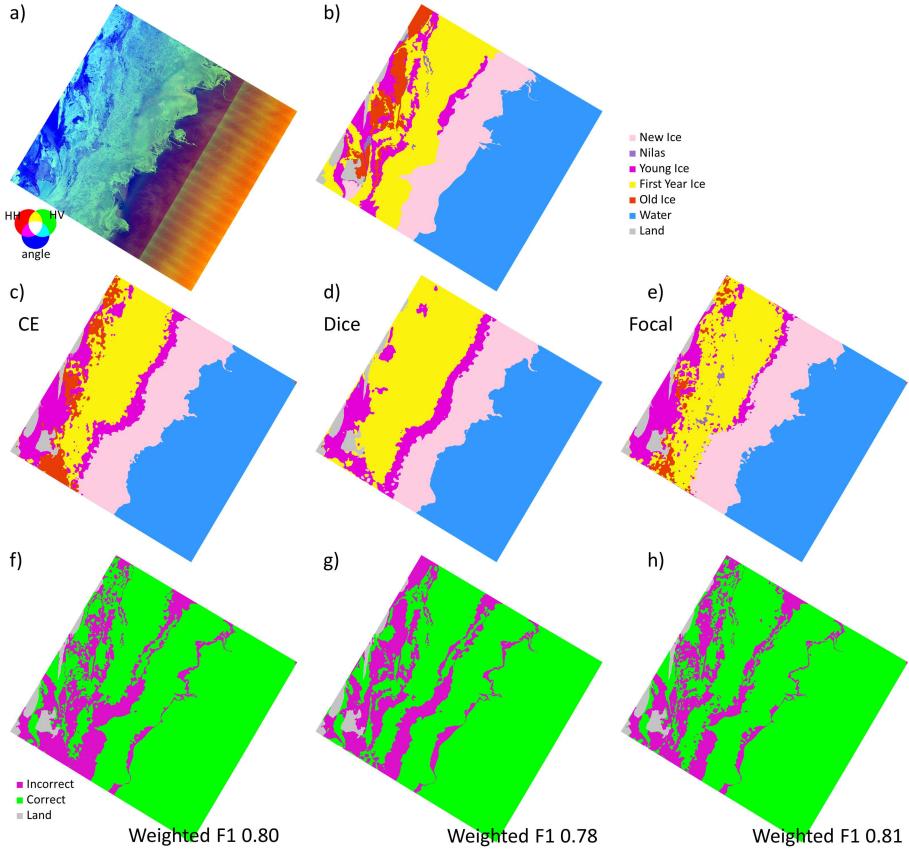


Fig. 1. Example of model predictions with highest scores for January. a) Input to the models, a Sentinel-1 image. b) Rasterized ExtremeEarth V2 labels. c), d), and e) share the same legend as b) and show predictions with highest weighted F1 for models trained with CE, Dice, and Focal loss, respectively. f), g), and h) share the same legend and show errors of losses compared to the labels in b). Although there is a larger mismatch between Dice results and ExtremeEarth V2 labels, the First Year Ice shows fewer blobs of other ice types in comparison to CE and Focal loss results.

a batch size of 24, Adam optimizer [15] with a learning rate starting at $1e-5$. The learning rate is multiplied by 0.1 if the validation loss does not decrease in five epochs, to a minimum of $1e-8$ and training stops if the validation loss does not decrease in 20 epochs. The weights with smallest validation loss are used for testing. These hyperparameters are kept the same, and we train models using three different loss functions with their default parameters: CE, Dice [9], and Focal [16]. We repeat each experiment three times to account for stochastic variation.

3. DATASET PRE-PROCESSING

The ExtremeEarth V2 dataset [11] is a collection of labels in the form of high-resolution ice charts for twelve Sentinel-1 images in Extra Wide (EW) mode acquired over the East Coast of Greenland. The twelve images are roughly one month apart, one for each month of 2018, and show several types of sea ice under different weather conditions. Ex-

tremeEarth V2 labels are polygons representing sea ice or open water with different characteristics. Each polygon contains attributes characterizing oldest sea ice, second oldest sea ice, as well as their respective concentrations. We derive the “dominant ice type” by computing the sea ice with the highest concentration in a polygon and use that as the target for our CNN models. We separate two scenes, January and July for testing. We clip half of February, June, August, and December for validation. Test and validation outputs are generated on a single pass for full images. To generate training samples, we extract 100 randomly placed patches of size 80 km^2 , which correspond to 1000×1000 pixels of Sentinel-1 images that were resampled to 80×80 m pixel size.

4. RESULTS AND DISCUSSION

Results show similar performance metrics for all losses investigated, with a marginal improvement for Focal loss. The average test weighted F1 score (average across January and July

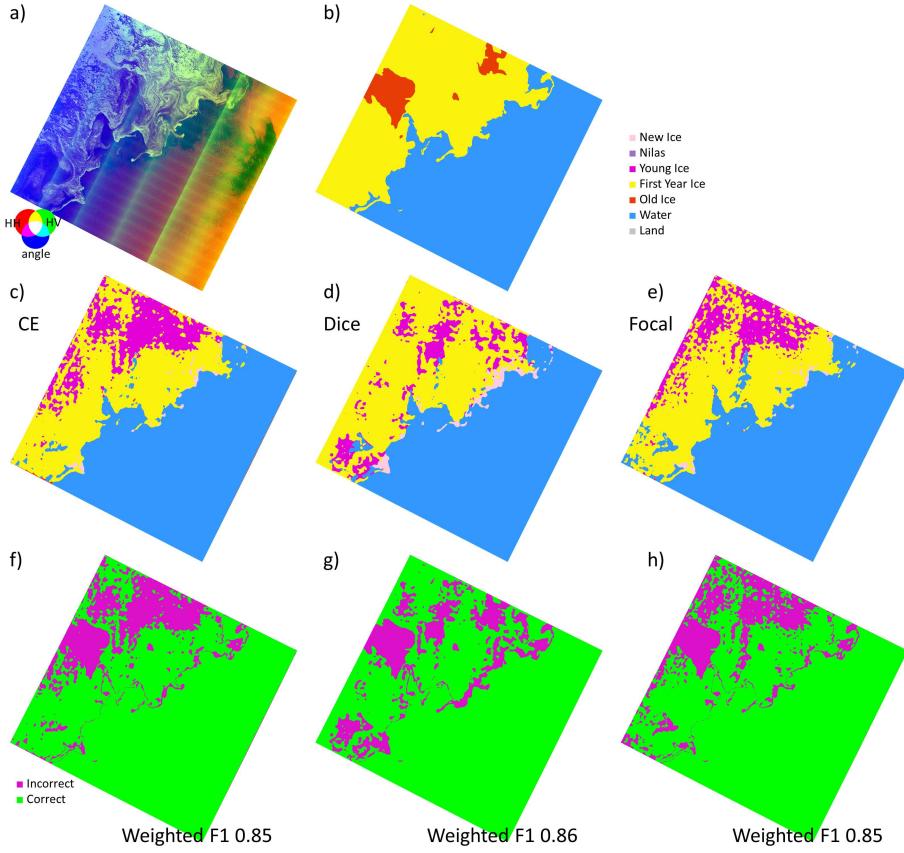


Fig. 2. Example of model predictions with highest scores for July. a) Input to the models, a Sentinel-1 image. b) Rasterized ExtremeEarth V2 labels. c), d), and e) share the same legend as b) and show predictions with highest weighted F1 for models trained with CE, Dice, and Focal loss, respectively. f), g), and h) share the same legend and show errors of losses compared to the labels in b).

for all three experiments) for CE is 0.815 (minimum of 0.805, maximum of 0.825). Dice results are 0.810 (0.80, 0.82), and Focal results are 0.822 (0.81, 0.83). Although Dice and Focal loss were developed for imbalanced datasets, our results show comparable performance for the more common CE loss for the ExtremeEarth V2 dataset. This potentially hints at other reasons that might have contributed to a relatively low performance in ice type classification, including lack of sufficient training samples, subjective labels at the polygon level as interpreted by the sea ice analyst, and ambiguity of SAR backscatter over different types. However, evaluating the results solely based on metrics performance can be problematic. Fig. 1 shows the results for the models with the highest weighted F1 for January. The model trained with Dice loss misses all old ice and classifies those regions as first-year ice. Due to the difference in thickness, incorrectly identifying old ice as first-year ice can have serious consequences for shipping and other marine operations. Of the three proposed methods, CE results appear to best capture the old ice areas. All models struggle to correctly identify the ice edge, which

is attributed to a combination of the model's architecture resolution and the fact that the new ice at the edge appears darker in the SAR image. Fig. 2 shows results for the models with the highest weighted F1 for July, during Summer when ice is generally melting and there is no ice growth. All models miss the old ice regions that are classified either as young or first-year ice. The weighted F1 for the model trained with Dice is the highest, however, the model predicts relatively large new ice regions when compared to CE or Focal, an undesirable outcome.

5. CONCLUSIONS

We evaluated how different loss functions affect the semantic segmentation of CNN models trained to classify the dominant ice type of ExtremeEarth V2 labels. Although Dice and Focal loss sometimes show better performance metrics, visual inspection indicates that predictions from models trained with CE loss are generally more physically reasonable. Further research for proper classification of minority classes might help

in the loss function choice.

6. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2026962. We thank the Extreme Earth project and MET Norway for making the ExtremeEarth dataset available to the sea ice community. The code used for this research is available at <https://github.com/geohai/sea-ice-segment>.

7. REFERENCES

- [1] N. Melia, K. Haines, and E. Hawkins, “Sea ice decline and 21st century trans-Arctic shipping routes,” *Geophysical Research Letters*, vol. 43, no. 18, pp. 9720–9728, 2016.
- [2] L. Pizzolato, S. E. L. Howell, J. Dawson, F. Laliberté, and L. Copland, “The influence of declining sea ice on shipping activity in the Canadian Arctic,” *Geophysical Research Letters*, vol. 43, no. 23, pp. 12,146–12,154, 2016.
- [3] H. Boulze, A. Korosov, and J. Brajard, “Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks,” *Remote Sensing*, vol. 12, no. 13, p. 2165, Jan. 2020, number: 13 Publisher: Multidisciplinary Digital Publishing Institute.
- [4] S. Khaleghian, H. Ullah, T. Krämer, N. Hughes, T. Eltoft, and A. Marinoni, “Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks,” *Remote Sensing*, vol. 13, no. 9, p. 1734, Jan. 2021, number: 9 Publisher: Multidisciplinary Digital Publishing Institute.
- [5] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [6] Y. Ren, X. Li, X. Yang, and H. Xu, “Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021, conference Name: IEEE Geoscience and Remote Sensing Letters.
- [7] Y.-R. Wang and X.-M. Li, “Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning,” *Earth System Science Data*, vol. 13, no. 6, pp. 2723–2742, Jun. 2021, publisher: Copernicus GmbH.
- [8] A. Stokholm, T. Wulf, A. Kucik, R. Saldo, J. Buus-Hinkler, and S. M. Hvidegaard, “AI4SeaIce: Toward Solving Ambiguous SAR Textures in Convolutional Neural Networks for Automatic Sea Ice Concentration Charting,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [9] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, “Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, ser. Lecture Notes in Computer Science, M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu, Eds. Cham: Springer International Publishing, 2017, pp. 240–248.
- [10] A. Kucik and A. Stokholm, “AI4seacie: selecting loss functions for automated SAR sea ice concentration charting,” *Scientific Reports*, vol. 13, no. 1, Apr. 2023. [Online]. Available: <https://doi.org/10.1038/s41598-023-32467-x>
- [11] N. Hughes and F. Amdal, “ExtremeEarth Polar Use Case Training Data,” 2021, publisher: Zenodo Version Number: 1.0.0. [Online]. Available: <https://doi.org/10.5281/zenodo.4683174>
- [12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2016, pp. 770–778. [Online]. Available: <http://ieeexplore.ieee.org/document/7780459/>
- [13] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking Atrous Convolution for Semantic Image Segmentation,” *arXiv e-prints*, p. arXiv:1706.05587, Jun. 2017, _eprint: 1706.05587.
- [14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015, publisher: Springer US.
- [15] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *ICLR*. San Diego, CA, USA: ICLR, Dec. 2014, p. arXiv:1412.6980.
- [16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, Oct. 2017, pp. 2999–3007, iSSN: 2380-7504.