Semantic Segmentation of Arctic Sea Ice Concentration Using SegFormer Architecture

Florian Frick, Jen MacDonald

University of Colorado, Boulder

Abstract. Creating sea ice maps of the Arctic and other polar regions has remained a challenge for both research and industry; currently this remains a manual interpretation task by experts in the field. Attempts have been made to automate this process using neural networks like CNNs. We expanded on previous work to automate this process by utilizing a colorized image dataset of the Hudson Bay in Canada and a SegFormer architecture to perform semantic segmentation prediction. Our results indicate that the model performs significantly better than a former attempt: we achieved an overall accuracy of 89% and a mean IoU of 0.50 compared to a previous result using a standard U-Net model with 83% and 0.44, respectively. While using this method is not accurate enough to reliably implement in industry applications, it is a positive step forward towards automating sea ice mapping.

1 Introduction

Accurately mapping sea ice from satellite imagery has consistently been a difficult task for scientists. With the increasing availability of high-resolution remote sensing data, sea ice charts can be created for both academic and industrial purposes. However, the process of developing these charts has historically been a laborious process: experts in the field look carefully at satellite images and map the different features indicated in the representation. There are many different metrics that are often mapped [1], such as sea ice concentration ("the fraction of sea surface covered by ice"), stage of development ("age and structural characteristics of the ice that may be inferred from specific visible features and knowledge of regional conditions prior to the observation"), and floe size [2] (the "cohesive sheet of ice floating in the water"). Because these maps are rendered manually, it can be difficult to generate up-to-date maps. These maps are particularly useful to both environmental scientists tracking markers of climate change as well as navigators in the shipping industry looking to plan out maritime routes. Creating a method to reliably automate this process will increase the availability of these maps to the industries that utilize them.

Many attempts have been made to computerize the process of detecting and classifying sea ice, including the use of deep learning models. Traditionally, various CNN models have been used successfully: U-net architectures, for example, have been used to classify images with above 90% accuracy [3]. However, there is a gap in the literature: not much research has been performed

on colorized satellite imagery using transformers. Traditionally, satellite data is processed into dual-polarization data, like horizontal-horizontal (HH) and horizontal-vertical (HV) bands. While this creates highly-detailed images, individual images can be incredibly large and difficult to process for researchers with limited resources. The sea image images that are in a smaller format (e.g. colorized jpg/png) are usually not utilized via satellite [4].

By processing satellite images with color channels, a model could be developed that is more accessible to a wider audience. Additionally, convolutional neural networks still dominate the research domain despite the increasing popularity of transformers for computer vision tasks. By utilizing a newer architecture on a satellite dataset that is processed in a more data friendly way, we will attempt to improve semantic segmentation accuracy over current methods. This paper proposed a novel application of SegFormer [5], a "semantic segmentation framework which unifies Transformers with lightweight multilayer perception (MLP) decoders," on an underutilized dataset of red/green/IR formatted sea ice charts of the Hudson Bay in Canada. By utilizing this combination of model architecture and dataset, we hope to establish a new methodology to increase the accuracy of automated sea ice mapping.

2 Related Work

2.1 Transformers for semantic segmentation

Transformers have been used for semantic segmentation successfully in the past. In the landmark SegFormer paper by Xie et al. (2021) [5], a transformer is able to achieve 84% mean IoU accuracy on an aerial image dataset of Christchurch in New Zealand. SegFormer works by taking in patches of images and uses positional embeddings and self attention to make predictions. SegFormer differs from traditional transformer architecture in that it allows for flexible patch sizes, uses a decoder along with the traditional encoder, and has a modified output head for semantic segmentation. In another paper, a pure transformer called SEgmentation TRansformer (SETR) was used to show a mIoU accuracy of 50.28% and 55.83% on the ADE20K and Pascal Context datasets [6]. These results strongly support the use of transformers for semantic segmentation tasks in computer vision challenges. Despite convolutional neural network (CNN) models still being the more popular architecture for visual tasks, especially in the domain of sea ice mapping, there is evidence that transformers can perform well on image datasets. A key difference in how we are using a semantic segmentation transformer is in its application: previous studies have applied the architecture to non-satellite images, which could have more distinctive classes than those found in sea ice satellite pictures.

2.2 Deep learning models for sea ice satellite data

The application of CNN architecture on sea ice images has been used frequently in research over the last decade. An extremely popular model is the

U-net architecture proposed for biomedical image segmentation [7], which uses an encoder to contract images and a decoder to expand and resize images to the original input. This architecture has been adapted to various image segmentation tasks in other domains. Ren et al. (2021) [8] adapted the U-net model to classify sea ice and open water, achieving a mean IoU score of 94.66%, 89.60%, and 91.61% on three different sea ice datasets with vertical-vertical (VV) and vertical-horizontal (VH) polarizations. Other CNNs modified specifically for sea ice classification, like Ice-Deeplab [4], have also been utilized with success. Ice-Deeplab modifies the Deeplab architecture [9] (a spatial-feature extraction model) to train a model on a 320-image dataset, resulting in a 90.5% overall accuracy. While applying deep learning to the task of sea ice segmentation has been successful, these images are in ultra-high resolution polarization formats. A single image can be several gigabytes of data, which can make it difficult for researchers without large resources to train their own sea ice classification models. Some research has been done with RGB/colorized satellite datasets, like Goncalves et al. (2021)'s research with weakly supervised CNN's on sea ice segmentation [10], but this area is still underexplored. The red/green/IR dataset of the arctic ice in the Hudson Bay will expand on how deep learning models perform on these image channels.

3 Methods

The arctic sea ice dataset of the Hudson Bay [11] starts with 3,392 images, identified by location (patch) and time. This dataset was created by the Canadian Ice Service, which produces publicly-available satellite images weekly. It contains JPEG images that are represented by red, green, and infrared color bands and comes in 357x306 pixel resolution. Each image has an associated mask, with fourteen possible pixel values based on SIGRID-3's ice chart codes, representing land, fast ice (ice which forms and remains [fastened to] the coast [12]), bergy water (area of freely navigable water in which ice of land origin is present [12]), and different concentrations of ice from open water to total coverage. To match previous models [13], we reduced these down to eight values representing land, fast ice, and larger granularity of ice concentration. We split the data such that 70% is used for training, and 30% for testing and validation. Although we believe that the model would better generalize to satellite images outside of the Hudson Bay if we ensured that each image of the same location/patch goes to the same split (as the segmentation maps are similar across time), the previous model [13] we wanted to compare accuracy with did not, so we split the data without this consideration. Then, to augment the training data, before every epoch, we randomly flipped and rotated the images and masks before resizing them to 256x256 pixels and normalizing. The flips were both horizontal and vertical and the rotation was within 10 degrees to increase the diversity of the training data. The same transformations were, of course, applied to both an image and its mask. Resizing reduced the computational complexity and made training much faster. Finally, normalization of pixel in-

tensity ensured that pixels were equally important, regardless of their absolute brightness. The validation/testing dataset was also resized and normalized to match the model's expected input.

We trained six models for semantic segmentation of the sea ice concentration of satellite images, with each model increasing in size and complexity. Each model started with one of the SegFormer models (B0 to B5) from the original paper [5] pre-trained on ImageNet, ADE20K, Cityscapes and COCO-stuff, and then we fine-tuned each to the augmented sea ice concentration data. Training was performed in 30 epochs with the AdamW optimizer, an initial learning rate of 6e-5, a batch size of 15, and cross-categorical entropy loss. During training, we saved the best checkpoint for each model, and loaded it for inference afterwards. With the trained models, we performed the experiments described below.

4 Experiments

To compare the performance and inference speed of each model, we perform inference on the test dataset - predicting the sea ice concentration at each pixel based on the satellite image. We measure the performance of each model by calculating the mean and overall categorical accuracy of pixels, the mean intersection over union (IoU), and the average time to make an inference for each model. While overall accuracy weighs all classes the same and gives an overall performance of the model, the mean accuracy weighs pixel categories by their frequency, which better measures the model's ability to identify minority classes. The mean IoU lets us evaluate how good the model is at identifying the borders between different concentrations of sea ice. The inference time lets us compare the computation cost with performance of different models, and better inform the direction future experiments should go.

To determine the viability of using the SegFormer architecture for semantic segmentation of sea ice concentration, we compare the accuracy and IoU of our models described above with an existing model trained on this data that uses a U-Net architecture. The existing model reports only overall accuracy and mean IoU, which we use as a medium of comparison between CNN and SegFormer models for the purpose of semantic segmentation of sea ice concentration.

The results of Table 1 illustrate how all models have very similar overall accuracy, mean accuracy, and mean IoU, despite the models being vastly different sizes. The smallest model, in fact, has the best accuracy and IoU, but the difference between models seems negligible, especially since the b4 model diverges from the inverse relationship between model size and performance. As expected, the inference time of a model increases as the number of its parameters increases.

The training and validation losses of Figure 1 give us some insight into these results. In the graph for b0, the smallest model, we infer that it was underfitting, as evidenced by the high loss (0.3) and small gap between the training and validation loss. We also note that the b0 model takes more epochs to achieve its best checkpoint than the larger models, presumably because a smaller model

	Overall Accuracy	Mean Accuracy	Mean IoU	Inference Time (1018 samples)	# Parameters
b0	0.886	0.605	0.504	20.3	3,716,200
b1	0.885	0.584	0.495	20.7	13,679,304
b2	0.885	0.585	0.499	29.1	27,352,776
b3	0.879	0.585	0.487	33.7	47,228,616
b4	0.882	0.608	0.501	36.3	63,999,176
b5	0.881	0.594	0.494	38.8	84,599,496

Table 1. Accuracy, IoU, and Inference Time of fine-tuned models

takes longer to learn. As the size of the model increases, already in b1, the model appears to begin overfitting, because the training loss decreases significantly while the validation loss plateaus and even rises in the biggest models. A future experiment could try training or fine-tuning a model with size between b0 and b1 (3.7M and 13.7M parameters) to see if it avoids both underfitting and overfitting.

	Open water	10-30% ice	20-50% ice	50-70% ice	70-90% ice	90-100% ice	Fast ice	Land
b0	0.927	0.445	0.0	0.358	0.521	0.962	0.710	0.918
b1	0.923	0.375	0.0	0.403	0.423	0.974	0.652	0.920
b2	0.919	0.359	0.052	0.387	0.352	0.976	0.717	0.919
b3	0.893	0.456	0.0	0.297	0.438	0.968	0.719	0.911
b4	0.932	0.347	0.009	0.537	0.374	0.955	0.802	0.909
b5	0.919	0.434	0.000	0.476	0.291	0.967	0.759	0.907

Table 2. Accuracy of individual classes

Table 1 showed that the mean accuracy is much lower than the overall accuracy in all models. This is because the mean accuracy weighs the classes according to their frequency, and the classes of land, open water, and 90-100% coverage are much more common than the other classes. This is reinforced by Table 2 and Table 3, which break down the accuracy and IoU for every class individually. The classes with the highest accuracy in all models are 90-100% ice coverage (0.955 to 0.976), open water (0.893 to 0.927), and land (0.907 to 0.920).

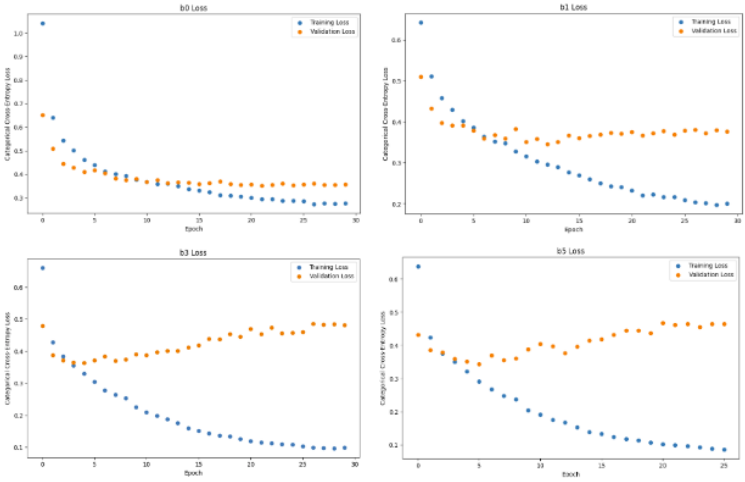

Fig. 1. Training and validation loss of increasingly large models

Similarly, the highest IoU classes are 90-100% ice coverage (0.903 to 0.915), land (0.866 to 0.871), and open water (0.815 to 0.30). This matches our expectations, because these are the most common classes in the dataset. Similarly, the lowest accuracy and IoU class, 20-50% ice is the least common in the dataset. Therefore, a future experiment could try, during data augmentation, overrepresenting the minority classes in the dataset or underrepresenting the majority classes to deal with the class imbalance.

Finally, we compare our results to the previously existing model [4], which uses a U-net architecture and achieves 0.83 overall accuracy and 0.44 mean IoU. We find that our models all achieve greater performance than the U-net model, with around 0.89 and 0.50 overall accuracy and mean IoU, respectively. Our models use different (fewer, in general) data augmentation and preprocessing techniques so a future experiment could more precisely measure the effect of a transformer architecture by applying the exact same preprocessing and data augmentation techniques to all models being compared.

Figure 2 shows examples of inference performed by our best fine-tuned model, b0, which gives further insight into our results. The predicted mask does an excellent job of matching the patterns in the input image, but sometimes differs from the true mask, revealing limitations of the model. For instance, as illustrated by the top right example, our model often predicts smoother, less granular regions than the true mask. A future experiment could try not to reduce the size of images, but instead maintain full resolution at the cost of training and inference time. These visual representations also illustrate possible limitations within the dataset itself. For example, the true mask of the bottom-right sample appears far less accurate to the input image than our predicted mask. This brings the validity of the accuracy and IoU metrics are brought into

	Open water	10-30% ice	20-50% ice	50-70% ice	70-90% ice	90-100% ice	Fast ice	Land
b0	0.816	0.284	0.0	0.238	0.331	0.915	0.577	0.869
b1	0.820	0.253	0.0	0.262	0.296	0.909	0.546	0.871
b2	0.815	0.251	0.043	0.252	0.265	0.908	0.594	0.866
b3	0.821	0.272	0.0	0.192	0.259	0.909	0.577	0.866
b4	0.820	0.259	0.009	0.246	0.279	0.908	0.616	0.868
b5	0.830	0.292	0.0	0.237	0.228	0.903	0.593	0.869

Table 3. Mean IoU of individual classes

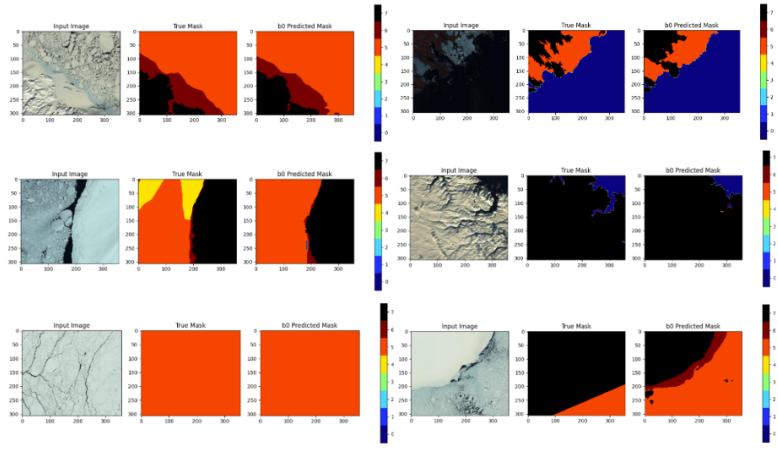

Fig. 2. Visual representation of model performance (b0 model)

question and informs our final insight into our results and future work. Due to the potential problems with this dataset's masks, and the persistent overfitting of our large models, a future experiment would highly benefit from a larger dataset of satellite images with well curated masks.

5 Conclusions

We presented a novel model architecture on a colorized satellite image dataset for sea ice segmentation. Our results show that there is potential for using a transformer architecture to automate the classification of land, sea ice, and water, even when using a lower resolution JPEG dataset. The classification of the sea ice concentration proved trickier, as the results for individual classes for sea ice were significantly lower than the class mean. Based on our results and

comparison with previous models, the Segformer model is a marked improvement on the U-net architecture. Additionally, our pretrained model can be fine-tuned in a Google Colab notebook or similar environment, and does not need advanced GPUs to perform training and evaluation. Our results show that although the model is not precise enough for dependable use in industrial applications, it signifies a small advancement in eventually automating sea ice mapping.

The ability to automate the production of sea ice charts has deep and far reaching implications for both research and industry. Swift production of these maps promises expanded datasets for climate studies, and the shipping industry can benefit significantly from the ability to quickly plan routes around areas of thick sea ice. However, widespread implementation of these maps would require extremely high accuracy, and premature adoption could create incorrect climate projections or ice-locked ships. Integration of automated maps would demand careful scrutiny and guidance from experts to ensure their accurate and responsible utilization.

References

1. Oceanic, N., Administration, A.: Ice Guide. https://response.restoration.noaa.gov/oil-and-chemical-spills/oil-spills/resources/observers-guide-sea-ice.html (2000)
2. Snow, N., Center, I.D.: Ice floe.
3. Wang, Y.R., Li, X.M.: Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning. Earth System Science Data **13**(6) (2021) 2723–2742
4. Zhang, C., Chen, X., Ji, S.: Semantic image segmentation for sea ice parameters recognition using deep convolutional neural networks. International Journal of Applied Earth Observation and Geoinformation **112** (2022) 102885
5. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems **34** (2021) 12077–12090
6. Strudel, R., Garcia, R., Laptev, I., Schmid, C.: Segmenter: Transformer for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. (2021) 7262–7272
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer (2015) 234–241
8. Ren, Y., Li, X., Yang, X., Xu, H.: Development of a dual-attention u-net model for sea ice and open water classification on sar images. IEEE Geoscience and Remote Sensing Letters **19** (2021) 1–5
9. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4) (2017) 834–848
10. Gonçalves, B.C., Lynch, H.J.: Fine-scale sea ice segmentation for high-resolution satellite imagery with weakly-supervised cnns. Remote Sensing **13**(18) (2021) 3562
11. Sylvester, A.: Arctic Sea Ice Image Masking. https://www.kaggle.com/datasets/alexandersylvester/arctic-sea-ice-image-masking/data (2021)
12. du Canada, G.: Government of Canada. Canada.ca. https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions/glossary.html (2020)
13. asylve: Sea-ice. https://github.com/asylve/Sea-Ice (2021)