POAC'25
St. John's, Newfoundland and Labrador, Canada

Proceedings of the 28th International Conference on Port and Ocean Engineering under Arctic Conditions
July 13–17, 2025
St. John's, Newfoundland and Labrador
Canada

Autonomous Labeling and Supervised Learning of a Deep Neural Network for Sea Ice Segmentation

Robert N. Norfleet, Oskar G. Veggeland, Ekaterina Kim, Nabil Panchi

Norwegian University of Science and Technology, NTNU, Trondheim, Norway

ABSTRACT

The application of Deep Neural Networks (DNNs) for Arctic sea ice segmentation and classification shows promise but requires large, labeled datasets. Fully annotated Arctic Sea Ice datasets are scarce, demanding domain expertise and significant manual effort. This study presents an automated approach to generating labeled datasets by integrating LiDAR data and RGB imagery, focusing on binary ice-water classification. Data from the GoNorth23 expedition aboard the RV Kronprins Haakon in July 2023 were used to generate sea ice labels by projecting LiDAR point clouds onto images, assuming all LiDAR returns originated from ice features. Three preprocessing methods–Thresholding, Morphological operations, and Otsu-Hybrid–were applied to refine the sparse point cloud data before training a neural network with these datasets as the ground truth. The preprocessing methods were evaluated against a 361-image manually labeled dataset, with the Morphological and Otsu-Hybrid methods achieving the highest resemblance (0.60 IoU). While this study serves as a proof of concept rather than optimization exercise, a U-Net model trained on the Morphological dataset for 30 epochs achieved an IoU score of 0.76 on the manually labeled test set. In general, the Morphological-trained U-Net exhibited strong recall, whereas the Otsu-Hybrid-trained model excelled in precision. These results suggest that large-scale labeled datasets can be generated from different weather conditions with minimal manual effort.

KEY WORDS: Deep neural networks; binary segmentation; LiDAR point cloud; sea-ice; image processing; automated labeling

I. INTRODUCTION

Deep Neural Networks (DNNs) have demonstrated sucess in various computer vision tasks and are being increasingly applied to sea ice mapping and monitoring. Several studies have utilized neural networks for segmenting and classifying sea ice features using optical imagery from icebreakers (Dowden et al. 2021, Li et al. 2024, Ma et al. 2024, MacMillan et al. 2024, Panchi et al. 2021). Additionally, methods addressing optical distortions caused by Arctic weather conditions have been investigated (Panchi & Kim 2024, Pedersen & Kim 2020).

Despite the potential of DNNs for sea ice monitoring, these models require extensive labeled datasets, which are time-consuming and require domain expertise to create. While large, publicly available datasets exist for applications such as autonomous driving (Caesar et al. 2019), pedestrian tracking (Leal-Taixé et al. 2015), and object recognition (Lin et al. 2015), comprehensive labeled datasets for sea ice segmentation remain sparse (Dowden et al. 2021, Li et al. 2024). The lack of annotated datasets is not due to a shortage of optical imagery, as studies frequently utilize publicly available images and video streams from icebreakers such as the *Nathaniel B. Palmer*, *Tian'en*, *USCG Healy*, and *50 Let Pobedy* (Dowden et al. 2021, Kim

et al. 2019, Li et al. 2024, Zhang et al. 2022). Rather, the challenge lies in the labor-intensive process of labeling images.

This study aims to automate the generation of labeled datasets by leveraging sparse LiDAR data. While the sparsity of this ground truth information presents challenges, similar issues exist in remote sensing (Maggiolo et al. 2022), medical imaging (Li et al. 2020), and coral segmentation (Alonso et al. 2017). Using synchronized RGB images and LiDAR point clouds from the research vessel (Veggeland et al. 2025), three datasets were generated automatically. These datasets were then used to train three U-Net models, whose predictions were evaluated against a manually labeled subset.

II. EXPERIMENTAL SETUP AND DATA COLLECTION

In 2023, RV Kronprins Haakon was equipped with an instrumentation system containing an optical camera and a Light-Detecting and Ranging (LiDAR) sensor during a summer research campaign in the Fram Strait. The system was mounted on the side of the bridge, roughly 17 meters above sea level, facing downwards to acquire images and point clouds of the ice conditions on the vessel's starboard side. Images of the instrument and its mounting position are provided in Fig. 1.

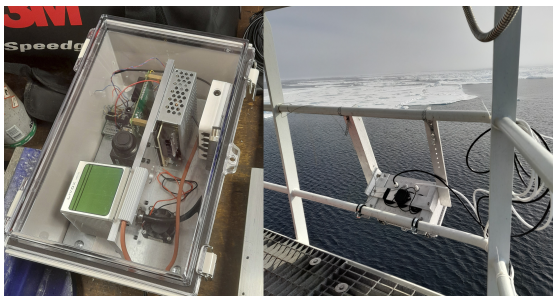


Fig. 1: Left: image of the sensor system used for dataset generation. Right: Mounting position on RV Kronprins Haakon.

The dataset generated for this paper is based on the methodology described in Veggeland et al. (2025), where a more detailed description of the sensor system can also be found. In short, using integrated navigation sensors, the raw LiDAR data is converted to a global 3D point cloud of the ice fields surrounding the ship. These point clouds are then projected onto the image frame of the camera, enhancing the optical images with sparse information from the LiDAR. Since the point cloud mostly consists of signals returned from ice features, these sparse projections can be considered positive labels in a binary ice/water segmentation problem. This process is depicted in Fig. 2.

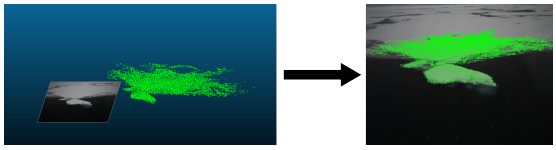


Fig. 2: Illustration of dataset generation. 3D point clouds are projected onto the optical images to provide sparse ice labels.

III. METHODS

A. LiDAR Field-of-View (FOV) Mask

LiDAR systems emit beams of light and measure the time taken to return. However, beyond a certain range, no meaningful data is returned. This limitation occurs near the horizon, necessitating a masking strategy to prevent the misclassification of the sky as open water. During training, loss was back-propagated only within the mask, where the LiDAR data was valid.

Figure 3 displays two examples of optical images with areas outside the mask highlighted. Typically, the mask removes distant regions and portions of the sky, though occasionally clipping occurs at the start and end of recordings.

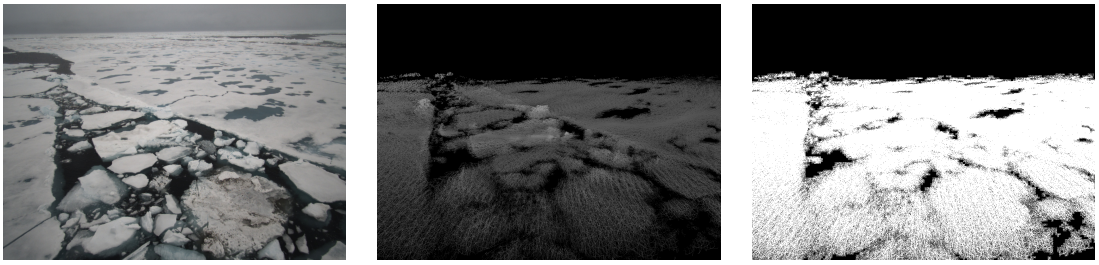
Fig. 3: Optical Images with areas outside the LiDAR FOV Mask Highlighted in Red

B. Dataset Preprocessing

The LiDAR points were processed in three different ways to address the sparsity of the point clouds. The following methods were designed to minimize this sparsity and create a more accurate ground truth representation.

1) Thresholding Dataset

This dataset represents the most basic preprocessing that should be done to create an ice mask from the LiDAR point cloud. This process, illustrated in Figure 4, forms the basis for the following two methods and is based on the assumption that if the LiDAR receives a point in the back-scatter, there is ice at that point. Therefore, if there is a LiDAR point (no matter the intensity) it is converted to an 'ice' pixel for the binary ground truth mask.



(a) RGB Image (b) Image of Raw Point Cloud (c) Thresholded Point Cloud

Fig. 4: Threshold Mask Example

2) Morphological Dataset

The second dataset is based on the binary threshold implemented in Section III-B1 with the addition of one iteration of morphological closing using a 3x3 pixel kernel. This step was taken to fill the holes that are present in the LiDAR point cloud and involves two steps and is demonstrated in Figure 5. Figure 5c presents a visual representation of the pixels added during this process, highlighted in blue.

1. *Dilation*: The binary thresholded image's white regions are expanded by checking each pixel's 3x3 neighborhood. If any pixel in the neighborhood is white (1), the center pixel is set to white as well.

2. *Erosion*: After dilation, the white regions are shrunk by again checking each pixel's 3x3 neighborhood. Now, the center pixel is set to white only if all pixels in the neighborhood have a value of 1.

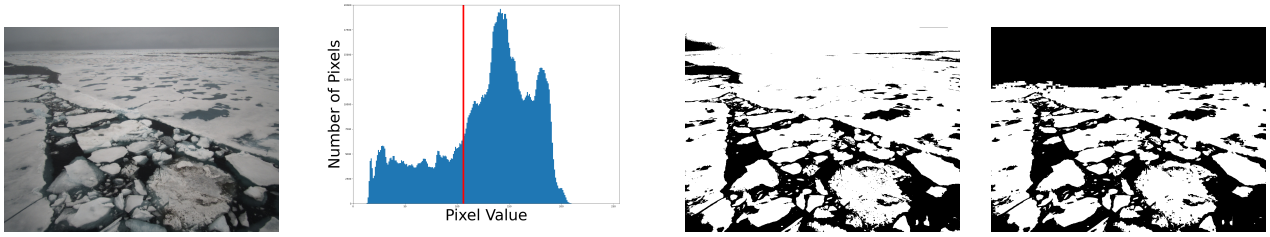
(a) Threshold Mask (b) Morphological Mask (c) Pixels Added by Morphological Closing Highlighted in Blue

Fig. 5: Morphological Mask Example

3) Otsu-Hybrid Dataset

The third and most complex of the datasets was created to improve the quality of the ground truth label. The optical image is first converted to grayscale and then a threshold value is found using Otsu's

binarization as presented in Otsu (1979). This mask is then overlaid with the thresholded LiDAR point cloud mask presented in Section III-B1 (Figure 4c). If the combined mask contains an ice pixel in both masks, it is kept as ice. Finally, the same closing process presented in the previous section is applied to the hybrid mask. The process is visualized in Figure 6, with Figure 6b displaying the histogram of the grayscale image and the resulting threshold value.



(a) RGB Image

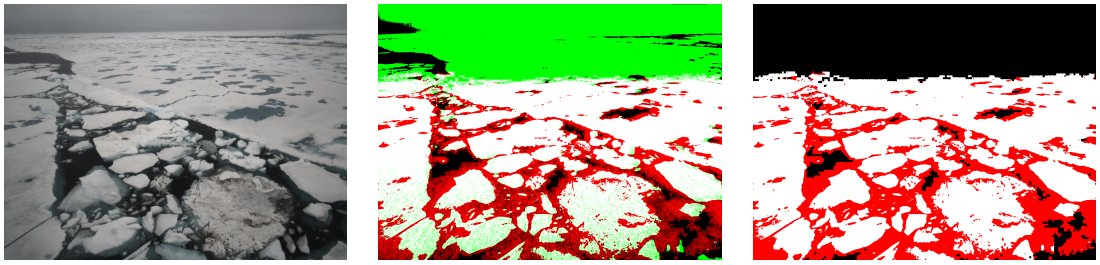
(b) Histogram of Grayscale Image with Otsu's Threshold Highlighted in Red

(c) Result of Otsu's Binarization

(d) Final Otsu-Hybrid Mask

Fig. 6: Otsu-Hybrid Mask Example

The reasons for the development of the Otsu-Hybrid preprocessing are illustrated in Figure 7. Figure 7b shows the benefits of the hybrid nature of the mask: The green pixels represent ice only present in the Otsu mask, while the red pixels illustrate ice in the Threshold mask. By combining only the overlapping areas, the ice floe boundaries are better defined and there is no misclassification of the sky as sea ice. Figure 7c presents a comparison to the Morphological mask, where pixels only present in the Morphological mask are highlighted in red. Images with multiple smaller ice floes benefit from the image thresholding methods in the Otsu-Hybrid method.



(a) RGB Image

(b) Hybrid Mask Classification
Green: ice pixels only present in Otsu mask (Figure 6c). Red: ice pixels only present in Threshold mask (Figure 4c)

(c) Morphological and Otsu-Hybrid Mask Comparison
Pixels only present in Morphological Mask (Figure 5b) highlighted in red

Fig. 7: Otsu-Hybrid Mask Attributes

C. Training

The U-Net model was implemented using the *Segmentation Models Pytorch* package (version 0.3.4) (Iakubovskii 2019), employing a ResNet101 encoder (He et al. 2015) pretrained on ImageNet (Deng et al. 2009). Training was conducted using PyTorch (version 2.0.1+cu117) on an NVIDIA RTX 3060 Laptop GPU with CUDA 11.2. The Stochastic Gradient Descent (SGD) optimizer was utilized along with a learning rate scheduler, *ReduceLROnPlateau*. All models were trained with for 30 epochs with a batch size of 16.

1) Loss Function

The Binary Cross Entropy (BCE) Loss with logits, formulated below, was used to train each model.

$$\text{BCEWithLogits} = \frac{1}{N}\sum_{i=1}^{N}\left(\log\left(1+e^{-z_i}\right) - y_i z_i\right) \tag{1}$$

where N is the number of samples, z_i is the logit from the model, and y_i is the true label (0 or 1).

An important part of this loss function was the addition of the LiDAR FOV mask. Adding the mask allows the model to be trained only on areas of the image where the LiDAR is 'active' and avoid class imbalance and incorrect labels. For implementation in code, the Binary Cross-Entropy loss was calculated and then masked prior to back-propagation. Implementation is displayed below, where $\mathbf{l}$ represents the logits (predicted values), $\mathbf{y}$ represents the true labels, $\mathcal{L}(\mathbf{l}, \mathbf{y})$ signifies the loss function (binary cross-entropy), and $\mathbf{m} \in \{0, 1\}$ the mask.

First, the loss is computed to give the loss for each element of the batch:

$$\mathbf{L} = \mathcal{L}(\mathbf{l}, \mathbf{y}) \tag{2}$$

Then the LiDAR FOV mask is applied to the loss to ensure that only elements where the mask $\mathbf{m} = 1$ contribute to the loss:

$$\mathbf{L}_{\text{masked}} = \mathbf{L} \cdot \mathbf{m} \tag{3}$$

Finally, the mean loss is calculated then back-propagated, ensuring that the sum of the loss is normalized by the number of valid mask elements:

$$\text{mean\_loss} = \frac{\sum \mathbf{L}_{\text{masked}}}{\sum \mathbf{m}} \tag{4}$$

After the loss was back-propagated, a sigmoid function was applied to the output logits. Then a threshold of 0.5 was applied to the output of the sigmoid to compute the predicted ice mask.

D. Evaluation

To evaluate the training datasets, each model was trained on each preprocessed dataset and then evaluated on the manually labeled ground truth. There are two subsets of the manually labeled ground truth: 175 image dataset of non-cropped images referred to as 'Roboflow' and a 186 image dataset of images with the horizon cropped out known as 'GoNorth'. Ice features such as meltponds, flooded ice, undersea ice, and sea ice rubble were all labeled as 'ice'.

$$\text{IoU} = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN} \tag{5}$$

$$\text{DICE} = \frac{2 * \mid A \cap B \mid}{\mid A \mid + \mid B \mid} = \frac{2 * TP}{2 * TP + FP + FN} \tag{6}$$

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{SIPP} = \frac{N_{\text{ice}}}{N_{\text{total}}} \tag{10}$$

Intersection-Over-Union (IoU) (Eq. 5), Dice score (Eq. 6), Pixel Accuracy (Eq. 7), Precision (Eq. 8), Recall (Eq. 9), and Sea Ice Pixel Proportion (SIPP) (Eq. 10) were used to assess performance. SIPP is defined as the ratio of ice pixels within an image. Only the pixels within the LiDAR FOV mask are considered. N_{ice} denotes the number of pixels classified as ice and N_{total} is the total number of pixels within the LiDAR FOV mask.

IV. RESULTS

A. Evaluation of Preprocessing Methods

1) Comparison to Manually Labeled Datasets

Subset	Dataset	IOU	DICE	Pix Acc	Precision	Recall
GoNorth	Threshold	0.54	0.68	0.77	0.83	0.62
	Morphological	0.66	0.78	0.83	0.81	**0.79**
	Otsu-Hybrid	**0.68**	**0.79**	**0.85**	**0.97**	0.69
Roboflow	Threshold	0.29	0.42	0.58	0.79	0.30
	Morphological	**0.53**	**0.66**	**0.75**	0.78	**0.59**
	Otsu-Hybrid	0.52	0.65	**0.75**	**0.90**	0.53
Combined	Threshold	0.42	0.55	0.68	0.81	0.47
	Morphological	**0.60**	**0.72**	0.79	0.80	**0.69**
	Otsu-Hybrid	**0.60**	**0.72**	**0.80**	**0.93**	0.61

TABLE I: Manually Labeled Dataset Evaluation

Bold numbers represent the best performances for each subset between the three preprocessed datasets

Table I summarizes the results of the comparisons between the preprocessing methods and manual labels. To compare against the manually labeled GoNorth dataset, the preprocessed datasets were cropped accordingly. While none fully replicate manual labels, the Morphological and Otsu-Hybrid datasets provide closer approximations. The IoU scores for these datasets are similar, yet their performance characteristics differ. Figure 8 illustrates that the Morphological dataset reduces false negatives at the cost of false positives, while the Otsu-Hybrid dataset does the opposite.

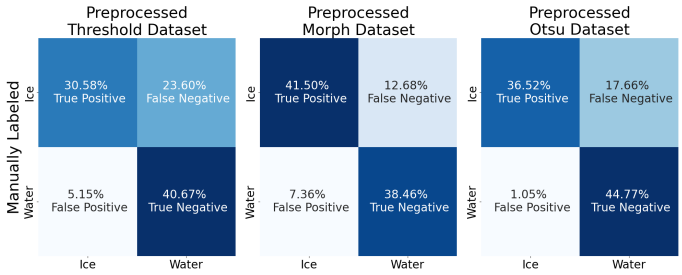
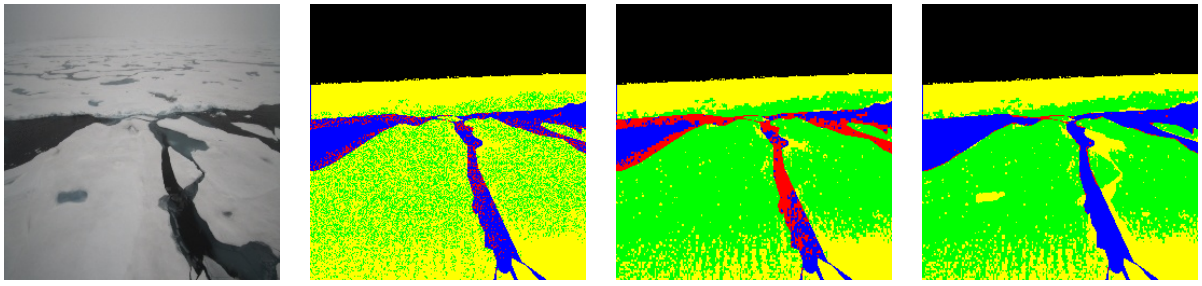


Fig. 8: Preprocessed vs. Combined Manually Labeled Dataset



(a) RGB Image (b) Threshold Dataset (c) Morphological Dataset (d) Otsu-Hybrid Dataset

Fig. 9: Pixel Classification of Preprocessing Methods Compared to Manual Labels

Green: True Positive (TP), Red: False Positive (FP), Blue: True Negative (TN), Yellow: False Negative (FN)

Figure 9 displays the pixel classifications of the three preprocessing methods when compared to a manual label and provides a visual confirmation of the statistics in Figure 8. Both the Threshold and Otsu-Hybrid masks (Figures 9b & 9d) have excessive false negatives but for different reasons. The Threshold dataset suffers from sparsity while the Otsu-Hybrid dataset misclassifies submerged ice and meltponds as open water. The Morphological mask has the highest percentage of false positives, as shown by the red pixels in Figure 9c. Close to the horizon, each preprocessing method produces false negatives due to the definition

of the LiDAR FOV mask. The edge of the mask is placed too far from the extent of the LiDAR point cloud, resulting in areas that are manually labeled ice but contain no LiDAR points.

B. Model Training

Each U-Net model was trained for 30 epochs with a batch size of 16. Loss curves, illustrated in Figure 10, reveal that validation loss, while initially lower, stabilizes as training progresses. The trend is likely due to the data augmentation that is applied to training samples but not to validation sets. Data augmentation consisted of resizing, flipping, rotating, cropping, and alterations to brightness and contrast. This was intended to create a more robust model and attempt to simulate adverse weather conditions in the Arctic. Overfitting is most evident in the Morphological dataset, where the training loss decreases even as the validation loss remains stable. Training loss on the Threshold dataset was the highest, confirming the challenges of learning from sparse ground truth data.

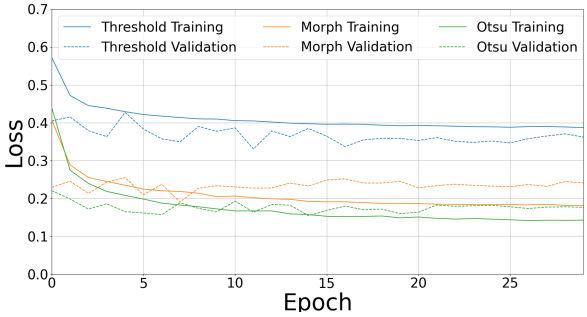


Fig. 10: Training Losses for each Dataset

The Morphological and Otsu-Hybrid datasets facilitated faster learning, likely because their preprocessing better represents sea ice features. With some of the sparsity of the point cloud filled in, the models are increasingly able to distinguish between sea ice and water. In simple terms, this binary segmentation task is finding a correlation between the 'lighter' sea ice pixels in the images and the preprocessed ground truth. Thus, the closer the preprocessed datasets emulate the 'perfect' ice mask, the quicker these models will learn to properly define the sea ice objects. Table II shows that while the Morphological dataset achieved the best training IoU, the Otsu-Hybrid dataset scored highest in validation IoU, corroborating the overfitting seen in Figure 10.

Dataset	Loss		IOU		DICE	
	Train	Val	Train	Val	Train	Val
Threshold	0.49	0.37	0.51	0.30	0.67	0.42
Morphological	0.50	0.32	**0.77**	0.62	**0.87**	0.74
Otsu-Hybrid	**0.34**	**0.19**	0.70	**0.67**	0.83	**0.79**

TABLE II: Training Statistics by Model

Bold numbers represent the best performances for each model between the three preprocessed datasets

Split	# of Images	SIPP			
		Manual Label	Threshold	Morphological	Otsu-Hybrid
Train	1464	-	0.49	0.69	0.45
Validation	317	-	0.34	0.39	0.29
Test	361	0.55	-	-	-

Fig. 11: Dataset Split Statistics

Table 11 presents the characteristics of the dataset split. The rational behind the split was mostly focused on minimizing dataset leakage– images within the same recording were placed within the same dataset split. The Sea Ice Pixel Proportion (SIPP) values for the training and test splits are roughly equal, but are significantly higher than the Validation set. For future studies, focus should be placed on balancing the amount of sea ice in each split, minimizing bias.

C. Evaluation of Model Predictions

The trained models were tested against two manually labeled datasets, with the results summarized in Table III and confusion matrices presented in Figure 12. The Morphological-trained U-Net achieved the

Subset	Dataset	IOU	DICE	Pix Acc	Precision	Recall
Roboflow	Threshold	0.48	0.59	0.74	0.79	0.51
	Morphological	**0.73**	**0.82**	**0.86**	0.86	**0.83**
	Otsu-Hybrid	0.61	0.72	0.82	**0.90**	0.62
GoNorth	Threshold	0.64	0.73	0.82	0.87	0.72
	Morphological	**0.78**	**0.87**	0.88	0.82	**0.95**
	Otsu-Hybrid	0.76	0.85	**0.89**	**0.96**	0.78
Combined	Threshold	0.56	0.66	0.78	0.83	0.62
	Morphological	**0.76**	**0.85**	**0.87**	0.84	**0.89**
	Otsu-Hybrid	0.69	0.78	0.86	**0.93**	0.70

TABLE III: Trained U-Net Performance on Manually Labeled Datasets

Bold numbers represent the best performances for each subset

highest IoU scores, reducing false negatives and improving recall, while the Otsu-Hybrid-trained model minimized false positives, yielding higher precision.

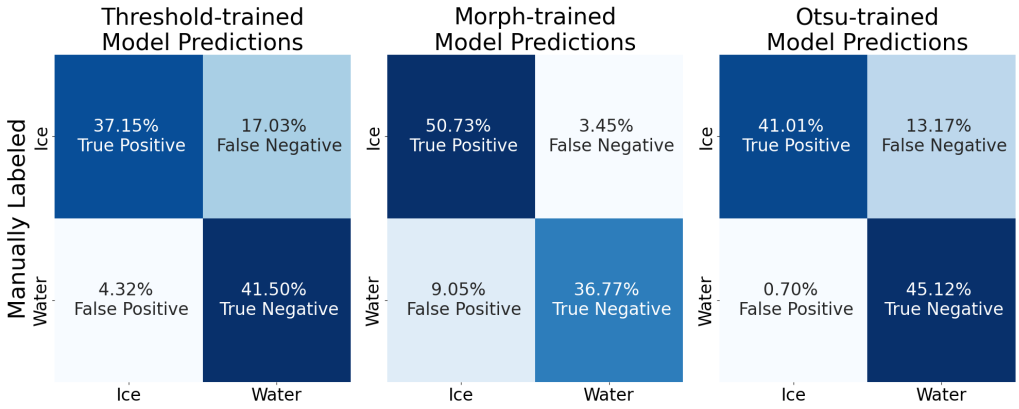


Fig. 12: U-Net Predictions vs. Combined Manually Labeled Dataset

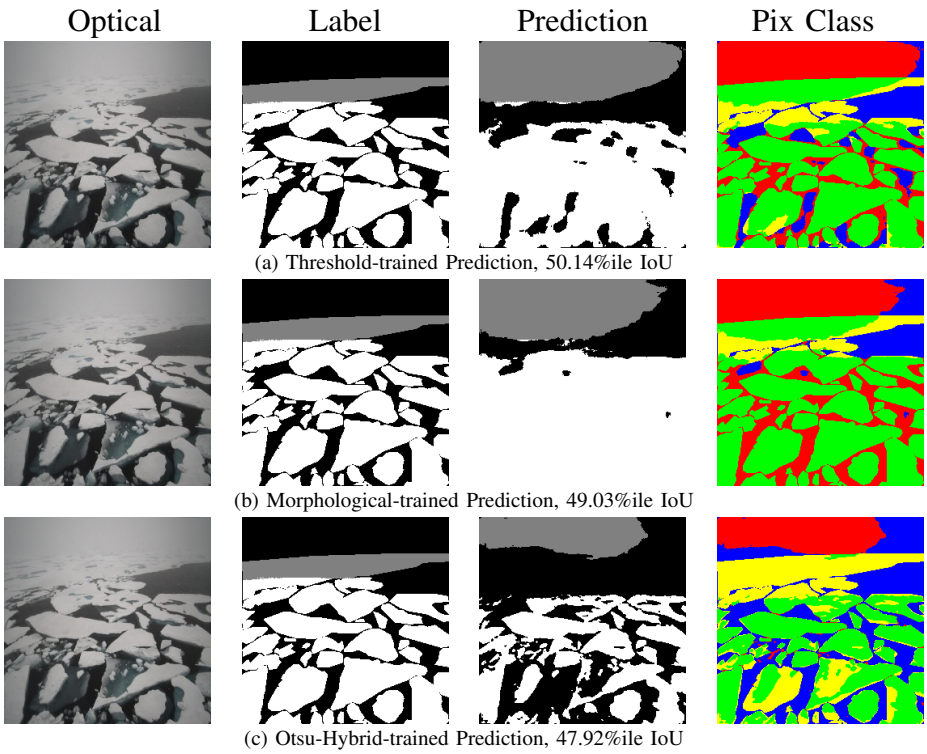
Figure 13 display a roughly average prediction for each of three different U-Net models given the same input image. The Threshold-trained model struggles due to the sparsity of the point cloud and motion blur. The Morphological-trained model overpredicts ice, merging multiple ice floes together. The Otsu-Hybrid-trained model better delineates floe boundaries but misclassifies submerged floes and meltponds. Despite masking regions of the sky during training, models still predict ice in bright sky areas, likely due to the high pixel intensity. Performance in these regions was expected to be suboptimal, as they were absent from training.

It is important to mention that the work above focuses on the strengths and drawbacks of the automatic labeling algorithms, but does account for LiDAR limitations. For instance, perfectly flat sea ice may fail to reflect the emitted signal, resulting in misclassification as open water.

V. CONCLUSION

This study presents three automated methods for labeling Arctic sea ice using LiDAR-derived information. A U-Net model trained on each of these datasets was evaluated against manually labeled ground truth images. While performance did not match state-of-the-art segmentation models, results demonstrate the feasibility of generating labeled datasets with minimal manual effort. After 30 epochs, a U-Net trained on the Morphological dataset achieved an IoU score of 0.76 on a 361-image test set.

Quantitative comparisons reveal that the Threshold dataset's excessive sparsity limits its usefulness. The Morphological dataset-trained model achieved the highest IoU and recall but tended to overpredict ice due to motion blur. The Otsu-Hybrid dataset-trained model excelled in defining ice boundaries with high precision but misclassified meltponds and shadows as open water.



(a) Threshold-trained Prediction, 50.14%ile IoU

(b) Morphological-trained Prediction, 49.03%ile IoU

(c) Otsu-Hybrid-trained Prediction, 47.92%ile IoU

Fig. 13: Average Image Prediction, ∼50%ile IoU

Gray shading represents predictions outside the LiDAR FOV mask. The last column colors the prediction based on pixel classification: Green: True Positive (TP), Red: False Positive (FP), Blue: True Negative (TN), Yellow: False Negative (FN)

Dataset selection should align with application needs and goals. For scenarios prioritizing high recall (e.g. Arctic navigation requiring ice avoidance), the Morphological dataset is preferable. For precise ice floe boundary deliniation, the Otsu-Hybrid dataset is more suitable. If meltpond segmentation is of interest– given its correlation with September sea ice extent (Schröder et al. 2014)–Otsu-Hybrid trained models show potential for their detection and classification.

A. Recommendations and Future Work

For future studies in this field, efforts should be made to increase the accuracy of the automated labeling system. For example, a devignetting algorithm similar to the one applied in (Sandru et al. 2020) could greatly increase the accuracy of the Otsu-Hybrid dataset. Other image processing methods, such as superpixel segmentation as demonstrated in (Alonso et al. 2017), could better emulate a manually labeled image of sea ice. Furthermore, testing with different ice and weather conditions (fog, snow, rain, sundogs, etc.) could prove valuable given the variable weather conditions present in the Arctic.

To improve training and inference scores, model architectures can be modified and hyperparameters tuned. Although this study was designed as a proof of concept, modifying model architectures for sea ice segmentation and classification has shown promise in the literature (Li et al. 2024, Panchi et al. 2021, Zhang et al. 2020).

ACKNOWLEDGMENTS

Presented research is supported by the Research Council of Norway through the POLARPROG project DigitalSeaIce (RCN no. 328960).

REFERENCES

Alonso, I., Cambra, A., Munoz, A., Treibitz, T. & Murillo, A. C. (2017), Coral-segmentation: Training dense labeling models with sparse ground truth, *in* 'Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops'.

Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G. & Beijbom, O. (2019), 'nuScenes: A multimodal dataset for autonomous driving', *arXiv e-prints* p. arXiv:1903.11027.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009), Imagenet: A large-scale hierarchical image database, *in* '2009 IEEE Conference on Computer Vision and Pattern Recognition', pp. 248–255.

Dowden, B., De Silva, O., Huang, W. & Oldford, D. (2021), 'Sea ice classification via deep neural network semantic segmentation', *IEEE Sensors Journal* **21**(10), 11879–11888.

He, K., Zhang, X., Ren, S. & Sun, J. (2015), 'Deep residual learning for image recognition', *CoRR* **abs/1512.03385**.

Iakubovskii, P. (2019), 'Segmentation models pytorch', https://github.com/qubvel/segmentation_models.pytorch.

Kim, E., Panchi, N. & Dahiya, G. S. (2019), 'Towards automated identification of ice features for surface vessels using deep learning', **1357**(1), 012042. Publisher: IOP Publishing.

Leal-Taixé, L., Milan, A., Reid, I., Roth, S. & Schindler, K. (2015), 'MOTChallenge 2015: Towards a benchmark for multi-target tracking'.

Li, J., Udupa, J. K., Tong, Y., Wang, L. & Torigian, D. A. (2020), Anatomy segmentation evaluation with sparse ground truth data, *in* B. A. Landman & I. Išgum, eds, 'Medical Imaging 2020: Image Processing', SPIE, p. 51.

Li, S., Wang, M., Wu, J., Sun, S., Shi, M. & Ma, R. (2024), 'Sea ice detection network for icebreakers in polar environments with attention-based deeplabv3+ architecture', **2718**(1), 012062. Publisher: IOP Publishing.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L. & Dollár, P. (2015), 'Microsoft COCO: Common objects in context'.

Ma, Z., Wu, J., Wang, S., Wang, Y. & Ma, D. (2024), 'A multi-task network with dynamic segmentation for sea ice classification in arctic shipping route optimization', **25**(9), 11893–11905. Conference Name: IEEE Transactions on Intelligent Transportation Systems.

MacMillan, C. G. J., Scott, K. A., Garvin, M. & Pan, Z. (2024), 'Breaking the ice: Video segmentation for close-range ice-covered waters'.

Maggiolo, L., Marcos, D., Moser, G., Serpico, S. B. & Tuia, D. (2022), 'A semisupervised CRF model for CNN-based semantic segmentation with sparse ground truth', **60**, 1–15.

Otsu, N. (1979), 'A threshold selection method from gray-level histograms', **9**(1), 62–66. Conference Name: IEEE Transactions on Systems, Man, and Cybernetics.

Panchi, N. & Kim, E. (2024), 'Deep learning strategies for analysis of weather-degraded optical sea ice images', **24**(9), 15252–15272. Conference Name: IEEE Sensors Journal.

Panchi, N., Kim, E. & Bhattacharyya, A. (2021), 'Supplementing remote sensing of ice: Deep learning-based image segmentation system for automatic detection and localization of sea-ice formations from close-range optical images', **21**(16), 18004–18019.

Pedersen, O.-M. & Kim, E. (2020), 'Arctic vision: Using neural networks for ice object classification, and controlling how they fail', **8**(10), 770. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Sandru, A., Hyyti, H., Visala, A. & Kujala, P. (2020), 'A complete process for shipborne sea-ice field analysis using machine vision', **53**(2), 14539–14545.

Schröder, D., Feltham, D. L., Flocco, D. & Tsamados, M. (2014), 'September arctic sea-ice minimum predicted by spring melt-pond fraction', **4**(5), 353–357. Publisher: Nature Publishing Group.

Veggeland, O. G., Kim, E. & Skjetne, R. (2025), 'An Autonomous System for Multimodal Mapping of

Sea Ice Fields from Ships', *IEEE Sensors Journal* pp. 1–1.

Zhang, C., Chen, X. & Ji, S. (2022), 'Semantic image segmentation for sea ice parameters recognition using deep convolutional neural networks', **112**, 102885.

Zhang, X., Jin, J., Lan, Z., Li, C., Fan, M., Wang, Y., Yu, X. & Zhang, Y. (2020), 'ICENET: A semantic segmentation deep network for river ice by fusing positional and channel-wise attentive features', **12**(2), 221.