Estimating Floating Ice Coverage in Tidewater Glacier Bays Automatically from Aerial Imagery

Hari Vishnu*, Lin Tianyue*†, Mandar Chitre*†, Bharath Kalyan*, Emily J. Venables+
*Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore
†Department of Electrical and Computer Engineering, National University of Singapore
+UiT, The Arctic University of Norway

***Abstract*—This study focuses on the estimation of floating ice distribution on the sea surface from aerial imagery at tidewater glaciers. The presence of ice-mélange in the glacial bay affects ice-loss at the glacier terminus via buttressing of the glacier face, insulation from ocean heat and breaking of oceanic currents. Thus, estimation of its formation, evolution, distribution and dissipation will aid an understanding of climate-change mechanisms at glacial bays. Given the challenges stemming from lack of data and labeled training samples, we employ a pre-trained vision model with zero-shot performance, the Segment Anything Model (SAM). SAM is fine-tuned to perform automatic estimation of floating-ice distribution on a dataset of images collected in Svalbard at various locations and through various modalities, by segmenting the images into floating ice and ocean surface not covered by ice. The Low-rank adaptation technique is employed on the image encoder structure to reduce the training data and hardware resources needed for model fine-tuning. This approach opens new avenues for real-time analysis of glacial dynamics and their implications on global climate patterns.**

***Index Terms*—glaciers, ice-mélange, Arctic, tidewater, climate-change, segmentation, SAM**

I. Introduction

The rapid changes occurring in polar regions due to climate-change are of global significance, affecting not only local ecosystems but also global sea-levels and weather patterns. A significant component of global sea-level rise is attributed to melting glaciers. In tidewater glacial bays, calving and submarine-melting account for a large percentage of the ice loss via frontal ablation. Understanding these ice loss mechanisms and the dissipation of the ice formed is useful to understand climate-change mechanisms and their impact at a global scale, including via sea-level-rise.

The calving at tidewater glaciers leads to floating ice in the bay, often in the form of ice mélange — a mix of floating icebergs and smaller ice fragments floating around in the bay [1]. The presence of this mélange in the glacial bays in turn affects the ice-loss via a feedback mechanism involving buttressing of the glacier face [2], insulation from ocean heat and breaking of oceanic currents and tidal mixing. Furthermore, the mélange also affects the biological habitat of local flora and fauna, and changes the underwater soundscape of the bay due to the sound generated by its melting [3] and changes in sound propagation in the underwater channel [4]. Thus, understanding the distribution and dissipation rate of floating ice in the bay is important to assess and model climate-change impacts in these regions.

Visual methods are a good way to assess the distribution of floating ice. Recent technological advances, particularly in machine-learning (ML) aided computer vision techniques and camera-based imagery using drones or land or ship-mounted systems, open up new means for assessing the floating ice. This paper focuses on employing computer vision techniques to analyze the distribution of floating ice in the glacial bay, via semantic segmentation. Semantic image segmentation involves classifying each pixel in an image into a predefined category. Modern image segmentation techniques often leverage deep learning methods, particularly Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs).

In order to compile a dataset from glacial bays for visual estimation of floating ice, aerial photographs were taken during field campaigns to Hornsund fjord and Kongsfjorden in Svalbard in June to August, 2023. Photos were taken through different modalities, including through drone-surveys and ship top-mounted cameras. This paper outlines the development of an automatic ML vision-based technique using this data to segment and estimate floating ice coverage in glacial bays from aerial imagery. We use an adaptation of a foundational model developed by Meta named the Segment-Anything-Model (SAM) [5], and fine-tune the Vit-L model from Meta using a Low-Rank Adaptation (LoRA) approach. This segmentation can then be used to estimate the coverage percentage of floating ice in the bay.

The technique developed in this work is shown to perform well on the aerial imagery data acquired from the field, and outperform a benchmark segmentation model available in the literature. The technique may also be used on permanent land-mounted camera systems for long-term ice-distribution monitoring. The goal is to estimate the floating ice distribution of ice mélange, providing methods and insights to refine models of ice dynamics. Section II details the field campaigns for data collection in Svalbard. Section III details the literature review on visual estimation and outlines the development of the proposed method for floating ice segmentation and coverage estimation, including fine-tuning techniques. In Section IV we discuss results using the proposed technique and compare it to another benchmark technique available in the literature, and in Section V we conclude the paper.

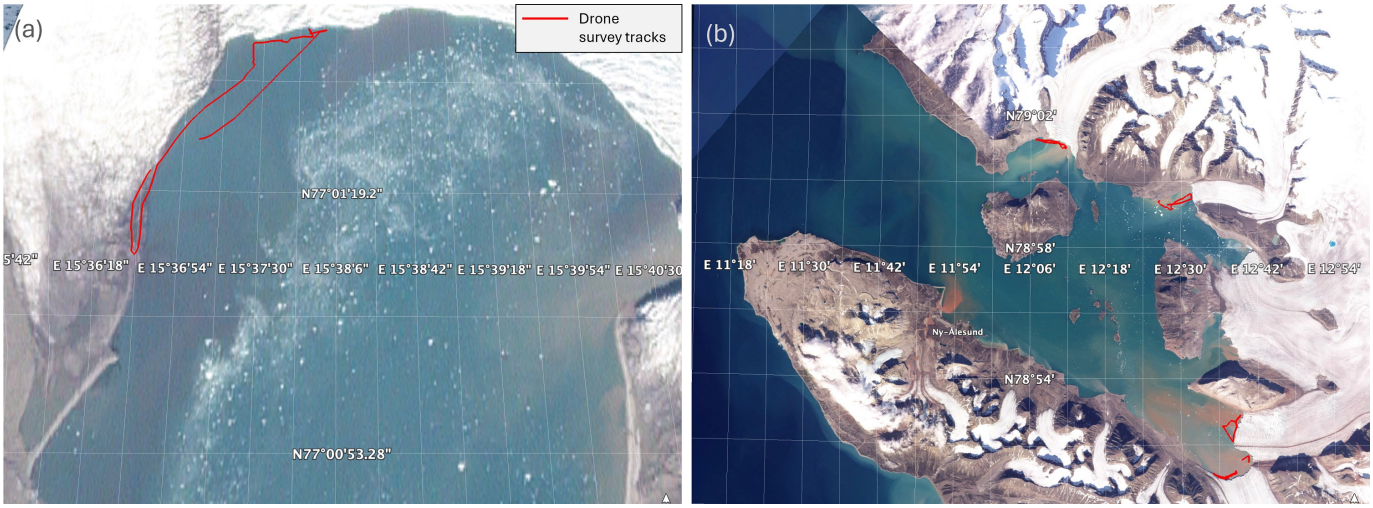

Fig. 1. Locations in Svalbard and tracks of drone-based video surveys undertaken during the field campaign at (a) Hansbreen glacier at Hornsund fjord and (b) four tidewater glaciers at Kongsfjorden. The background maps were obtained from www.planet.com, Planet Labs PBC, under a CC BY-NC-SA 2.0 license

II. Data collection

Field campaigns were undertaken in 2023 in two regions in Svalbard – (1) Hornsund fjord and (2) Kongsfjorden, shown in Fig. 1, to record passive acoustics, visual data, active acoustics, and conduct robotic surveys near the glacier termini. The Hornsund campaign was staged from the Polish Polar Research station, whereas the Kongsfjorden campaign was staged from the Norwegian Polar Institute research station in Ny-Ålesund. We analyze data from videos and photos taken during this field campaign, specifically from 5 glacier termini - (1) Hansbreen at Hornsund fjord, and (2) Kronebreen, (3) Kongsvegen, (4) Conwaybreen and (5) Blomstrandbreen at Kongsfjorden, whose locations are shown in Figs. 1(a) and (b). Our dataset comprises image collections taken from different locations and modalities, which includes drone-mounted camera (from Hansbreen, Kronebreen, Kongsvegen, Blomstrandbreen and Conwaybreen) and ship-mounted camera photographs (Kronebreen). The drone-transects of the surveys are shown in Fig. 1. Drone footage in Hornsund was acquired by walking on foot to a spot within safe distance from the glacier terminus, and flying the drone to survey the terminus. Drone footage in Kongsfjorden was acquired by using a Polarcirkel boat from the Norwegian Polar Institute station to get to within ∼500 m of the glacier termini, and flying the drone from the boat. Ship-mounted camera photographs were taken by mounting the camera on the R/V MS Teisten operated by Kingsbay AS at the Ny-Ålesund station.

III. Visual estimation of floating ice

A. Prior work

The photogrammetric assessment of ice distribution has received attention in the literature due to potential applications in hydrology, geography, and environmental sciences. Research in this area has increasingly adopted ML techniques. For example, Li et al. explored a semi-supervised approach for ice-water classification [8], and Evans et al. has applied unsupervised ML methods to detect and analyze iceberg populations [9]. These primarily focused on analyzing satellite imagery. Recent advances have seen the development of methods tailored for aerial photography. Ansari et al introduced a CNN-based framework for river ice classification [10]. Panchi et al utilized CNNs to classify close-range optical images of ice, focusing on textural and pattern recognition [11].

These works and the approaches employed therein, while insightful, rely on large datasets for training, which is challenging in the current scenario given the smaller scale and specific environmental settings of our data. The data available from our field campaign are representative of the locations surveyed, but not necessarily enough to train a full segmentation model from scratch. Hence, we use adaptable techniques, namely, pre-trained ViTs models for this task, with the aim of training a generalized end-to-end semantic segmentation model. The design objective is that this model must be capable of identifying ice mélange, including large icebergs and smaller ice fragments, from images taken in the glacial bay. Additionally, it should be able to distinguish sea surface that is not covered by ice, as well as other regions (such as glacier cliffs or rocky shores), so that we can use this to compute the percentage cover of floating ice in the bay by area.

B. Segment Anything Model

SAM is a foundational model for visual segmentation launched by Meta [5], [6]. Primarily, this model is trained by utilizing prompt engineering to facilitate segmentation based on prompts provided to a pretrained large-scale model. It has the potential to be applied in downstream segmentation tasks and can be integrated with other visual tasks to develop solutions for visual analyses. The model exhibits zero-shot predictive capabilities, and is able to perform semantic segmentation, instance segmentation, and panoramic segmentation

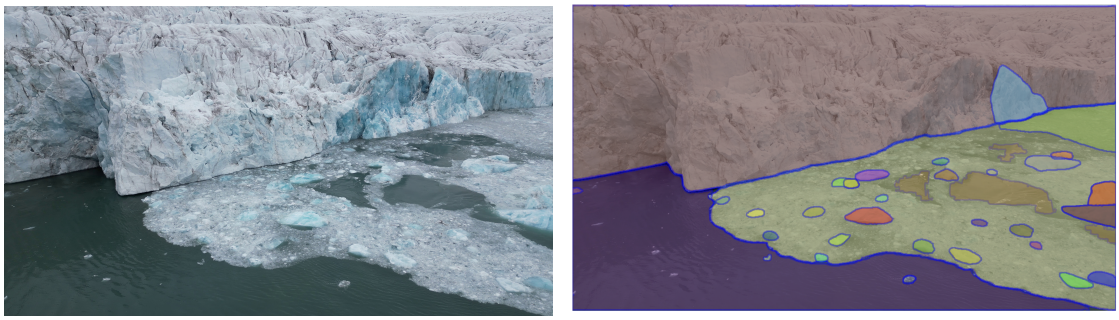

Fig. 2. SAM segmentation output with a sample image using the demo provided on the website [6]

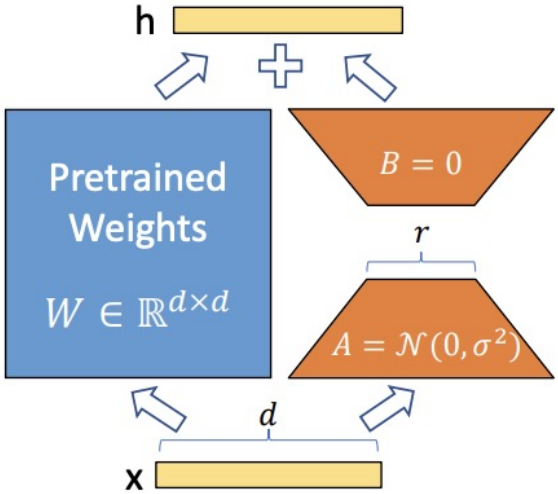

Fig. 3. Illustration adopted from [7] demonstrating LoRA fine-tuning procedure using matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, for pretrained weights $\boldsymbol{W}$, input $\mathbf{x}$ and output $\mathbf{h}$. $\boldsymbol{A}$ and $\boldsymbol{B}$ are initialized with a random normal distribution and zeroes, respectively.

concurrently. Moreover, it allows for real-time adjustments to the outcomes based on user prompts [5]. Based on these considerations, we adopt SAM for our task.

SAM comprises of 3 components- an image encoder, a prompt decoder, and a mask decoder [5]. The *Image encoder* is a pre-trained ViT which processes each image once to extract embeddings for segmentation. The *Prompt encoder* can encode diverse user inputs, including points, bounding boxes, and text, into prompt embeddings. The points and boxes given through prompts are represented in the network using positional encoding. Dense prompts like masks can also be embedded into the model. The *Mask decoder* employs a modification of a transformer decoder block followed by a dynamic mask prediction head. It computes the mask probability for each image location. The prompt encoder and mask decoder are computationally light-weight, relatively, ensuring rapid response to user prompts upon the pre-computed image embeddings. This architecture enables SAM to perform accurate and efficient interactive annotation processes.

When applied on the image datasets described in Section II, the vanilla SAM performs well, broadly speaking, in terms of accuracy of distinguishing different components of the image, as demonstrated in the example in Fig. 2. It successfully separates the glacier cliff, sea water, and floating ice, using natural color and textural differences in the image with the capabilities endowed during its pre-training. However, as far as our application is concerned, the segmentation is incomplete - eg. SAM treats some of the large ice pieces and floating ice as separate objects, which need to be combined into a single category. The final model is intended to perform a three-way segmentation of the scene into three categories:

1) floating ice,
2) the ocean surface that is not covered by ice, and
3) other elements in the scene.

Thus, we fine-tune the SAM to achieve this three-way segmentation required of our system.

C. Fine-tuning

While large models pre-trained on extensive datasets demonstrate substantial capabilities, their general training process and the diverse dataset used often limit initial effectiveness on specific, nuanced downstream tasks. Fine-tuning, a form of transfer learning, addresses this by training the large models using the datasets for the specific downstream task. Common fine-tuning strategies include altering the model's architecture or freezing a subset of the model's parameters, significantly reducing the number of trainable parameters, therefore the training is feasible even with limited computational resources. Leveraging the emergent abilities of large pre-trained model, the necessity for large datasets also diminishes, and the model can perform well with small amount of training data. Most importantly, performance in downstream tasks can be improved.

We fine-tune the ViT-L model provided by Meta [5], [6], using the LoRA approach [7], to achieve a three-way segmentation into the categories mentioned previously. LoRA is an efficient way to adapt pre-trained models to downstream tasks, inspired by the idea that the change in weights during model adaptation has a low "intrinsic rank". This approach introduces rank decomposition matrices to dense layers, in order to train these layers in the pre-trained model indirectly by freezing the

ML architecture	MIOU
SAM-based model fine-tuned with training dataset	0.89
UNet model trained on training dataset	0.54

TABLE I
PERFORMANCE COMPARISON BETWEEN SAM AND UNET-BASED MODELS

original weights $\boldsymbol{W}$ and optimizing the weights in the matrices alone. As shown in the illustration in Fig. 3, only the low-rank matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are trainable, and the inference result computed by $\boldsymbol{A}$ and $\boldsymbol{B}$ are added to the original result output by the pre-trained weights. LoRA is computationally efficient, as only small low-rank matrices need to be trained [7]. At the same time, the simple linear design introduces minimal inference latency compared to adapters.

In order to fine-tune SAM to develop a segmentation model to meet our design objectives, we create a training set by manually annotating a number of images from the acquired dataset using MobileSAM [12], a lightweight variant of SAM that accepts user prompts. These are used to fine-tune the ViT-L model [5]. First, the image encoder component is fine-tuned, as this is the stage at which the information from the image at the input is encoded in the first stage. This fine-tuning is done using the LoRA approach because the encoder has a large number of parameters. LoRA matrices are added to every 24 attention layers and the rank is set to 512 during fine-tuning with the annotated images. The loss function used during training is a combination of Dice loss [13] with cross-entropy (DiceCE loss) [14], and the Adam optimizer is used.

While fine-tuning the image encoder leads to an improvement in the model performance, the test performance does not yet achieve the desired specifications. This is likely because modifying the image encoder only adjusts the model's capability in *perceiving* images of our task, but the model's *interpretation* of these images which generates the segmented mask outputs remains unchanged. In order to tackle this, we next fine-tune the mask decoder component, which takes the image embedding and the output from the prompt encoder, and outputs the mask and score as segmentation results. The prompt encoder endows SAM with potent interactive capabilities. However, for our application, we require a fully automatic segmentation model to segment floating ice. Hence, in our case, it's necessary for the mask decoder to be able to produce the desired outcomes even without prompts and the prompt encoder component is not used.

IV. RESULTS

The final model obtained after sequential fine-tuning of the encoder and decoder components using only 10 images, spread across different glaciers and acquired using drone and ship-mounted camera, performs well on images obtained from multiple locations, as seen on results with images from Hansbreen (Fig. 4), Kronebreen (Fig. 5), and Conwaybreen and from images acquired from ship-mounted camera (Fig. 6). Specifically, we also observe the following about the model's performance:

1) As seen by comparing across multiple figures, the model is robust to changes in color of the water, spread of the ice-mélange, and size of ice pieces.
2) As seen in Fig. 5, the model is robust to variation in lighting, shadow and color, even when these occur within the same image.
3) The model is robust to reflections on the water surface, as seen in Fig. 5(b) and (c), and Fig. 6(c).
4) The model is robust to the modality based on which the image was acquired (aerial/ship-mounted) - in Fig. 6(a) and (b), the model is able to identify the water and ice correctly, and also delineate the ship as part of category 3 correctly.

Overall, the segmentation model exhibits satisfactory robustness in performance to several possible variations in the input images. Minor flaws can be spotted in some cases - eg. in Fig. 6(c), the extreme glare of the sky causes the model to wrongly classify a small portion near the terminus as category 3 instead of 2 towards the center portion of the image. Likewise, in Fig. 5(b), a few pixels in category 2 are misclassified as category 3. However, such errors are observed to be rare, and the performance is reasonable enough for usage on field data.

Based on the three-way segmentation by the fine-tuned SAM, we compute a naive metric of the percentage cover of ice over the ocean surface within each image assessed, the ice coverage percentage C, computed as

$$C = 100 \times \frac{P_1}{P_1 + P_2} \tag{1}$$

where P_1 and P_2 are the number of image pixels segmented under categories 1 (floating ice) and 2 (uncovered ocean surface) respectively. We note that this metric does not exactly correspond to the percentage cover of ice over the ocean surface in terms of *area*, but is a simple proxy for it that is easily computable from the segmented outputs.

The whole training process can be done with limited computational resources. Our training was conducted on a NVidia L40 GPU with 48 GB of RAM, over 40 epochs. Given the small training dataset and the reduction of the number of parameters by using methods like LoRA, the whole training process required only 630 seconds to complete. Deploying the SAM requires about 1385 MB of memory, which would be a consideration to take into account when deploying it on low-power edge-computational devices for real-time monitoring and processing.

In order to benchmark the performance, this technique was compared against a standard CNN-based segmentation model based on the UNet architecture [15]. The UNet architecture consists of an initial contracting path of neural network weights (with a bottleneck) to capture context in the image, and a symmetric expanding path following that for precise localization. We train the UNet model for 50 epochs using the same training set we used to fine-tune the SAM.

We also quantify the performance of segmentation using the Mask-intersection-over-union (MIOU) metric, as shown in

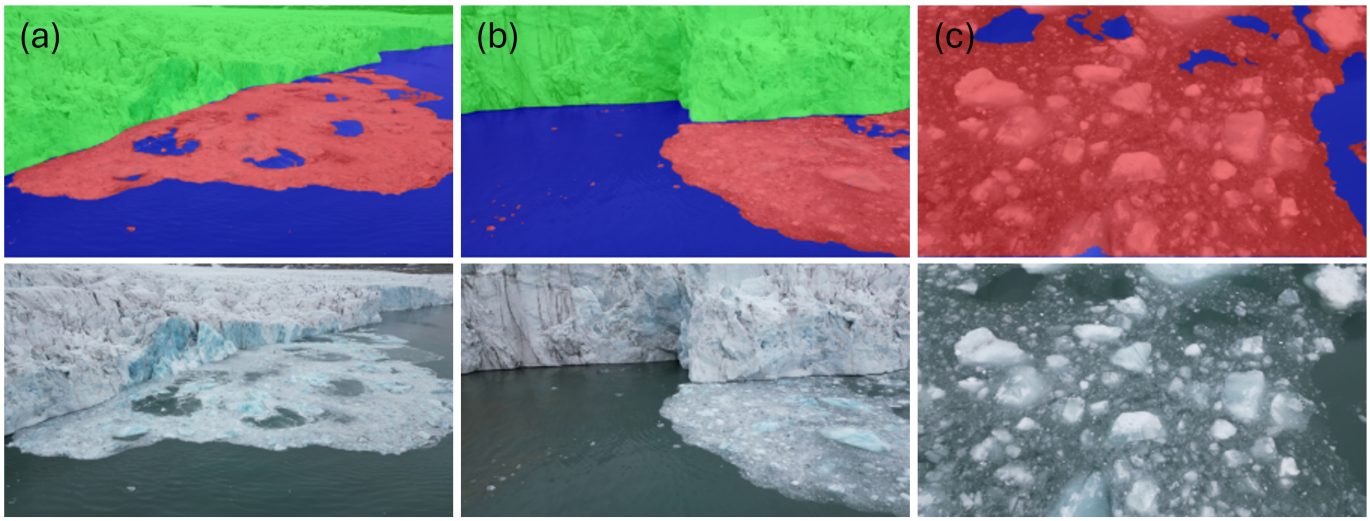

Fig. 4. Ice-segmentation results with drone images from Hansbreen - row 2 shows the original images, and row 1 shows the segmented outputs. Red semi-transparent mask indicates the areas labeled as "floating ice" (category 1), blue mask indicates "ocean surface not covered by ice" (category 2), and green indicates "other areas" (category 3). The pixel-based coverage percentage C is estimated as 46.4%, 38.9% and 90.6% respectively for the three images.)

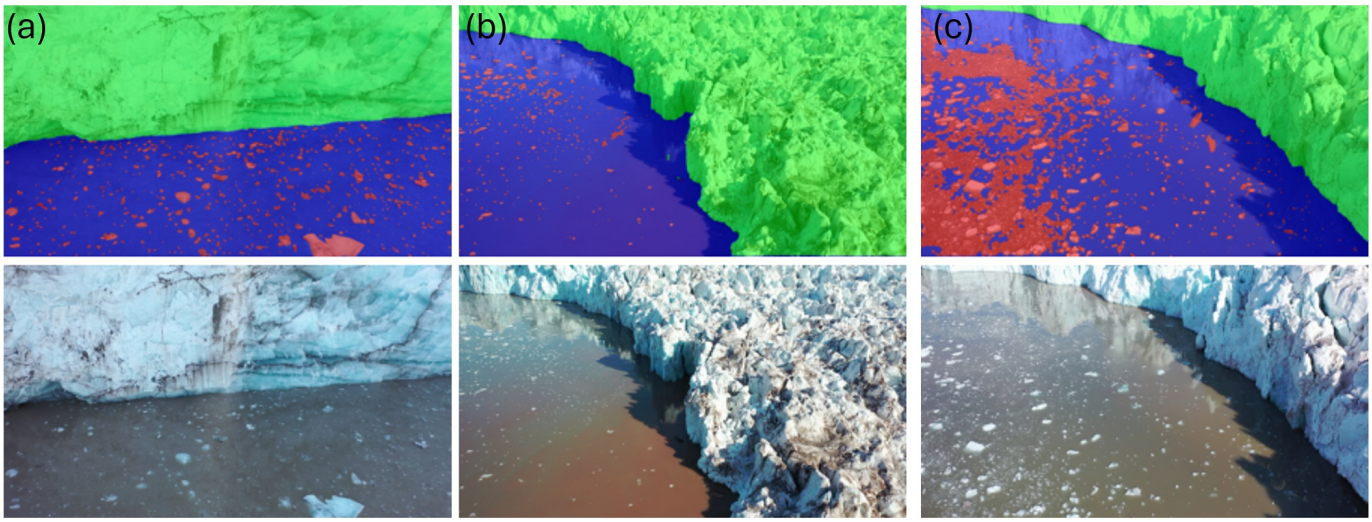

Fig. 5. Ice-segmentation results with drone images obtained at Kronebreen - row 2 shows the original images, and row 1 shows the segmented outputs. The segmentation color convention is the same as Fig. 4. The pixel-based coverage percentage C is estimated as 10.5%, 3.67% and 31.0% respectively for the three images.

Table I. The metric was computed on fully annotated test data consisting of 17 images from Hansbreen. It can be seen that the fine-tuned SAM-based model obtains a reasonable MIOU of 0.89 with the test data. The fine-tuned SAM performs much better than the UNet model trained from scratch, which only obtains an MIOU of 0.54. This is clearly because the SAM has the advantage that comes with pre-training, whereas the number of images was not enough to train the UNet adequately from scratch.

V. Conclusions and Future Work

We developed and showcased the capabilities of an ML-based image segmentation model to segment floating ice and ice-melange in tidewater glacial bays. The model is based on the Segment-Anything-Model, fine-tuned from the Vit-L model using a Low-Rank Adaptation approach. It is able to work with a low amount of training data, and carries out three-way segmentation of the images into floating ice, uncovered ocean, and other regions. The model was shown to perform robustly, handling images from different glaciers in different regions, and obtained using various modalities, with different lighting and color conditions. The model was also shown to outperform another benchmark model architecture for segmentation without pretraining. This underscores the superior performance and nuanced recognition abilities of advanced pretrained models like SAM in complex image segmentation tasks. Field campaigns in polar regions are challenging and expensive, and collecting a large amount of data from such

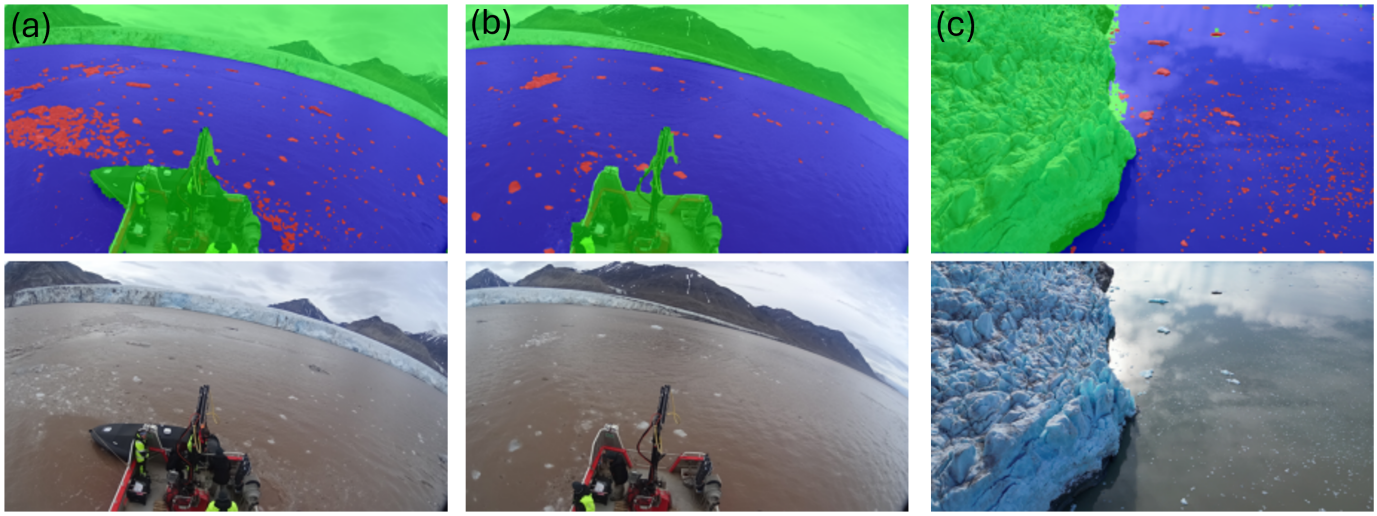

Fig. 6. Ice-segmentation results with (a) and (b) images obtained from ship-mounted camera near Kronebreen and (c) drone images obtained at Conwaybreen. Row 2 shows the original images, and row 1 shows the segmented outputs. The segmentation color convention is the same as Fig. 4. The pixel-based coverage percentage C is estimated as 8.89%, 2.43% and 4.76% respectively for the three images.

regions is not easy. Given this, the approach outlined herein based on the minimal available data to develop a segmentation model, would help efficiently utilize the valuable data collected from such campaigns to develop robust systems for vision-based analysis.

This model can be deployed on land- or aerial- based camera systems to estimate the formation, movement, and dissipation of floating ice in glacial bays resulting from climate-change mechanisms and hydrological circulation in the bay due to forces such as tidal pumping, subglacial discharge plumes [16] or submarine melt-induced factors. In the long run, a vision-based system using this model could help study these factors effectively to better understand climate-change mechanisms in glacial regions.

An important future step would involve using the segmented outputs to compute the percentage cover of ice over the ocean surface in terms of surface area, which is a metric that may aid a better understanding of ice dynamics in the region. This would involve using additional information on the geometry and the camera used for the data collection. Tracking the movement of individual icebergs so as to estimate the circulation in the bay would also be a research problem with potential applications.

Acknowledgements

We acknowledge the funding support from the INTERACT III Transnational Access grant under the European Union H2020 Grant Agreement No.871120, and the Arctic Field grant Project No. 342135 under the Research Council of Norway, for the travel and logistics for the campaigns. We thank the Polish Polar research station and the Norwegian University of Science and Technology for helping with administration of the funds, as well as facilities during the campaign. We also thank Kingsbay AS and the Norwegian Polar Institute research station at Ny-Ålesund for their facilities support. We are grateful to Grant Deane, Oskar Glowacki, Dale Stokes and Hayden Johnson of the International Partnership for Acoustic Monitoring of Glaciers for discussions on ice-mélange estimation, Asgeir Sørensen, Halvar Gravråkmo and Rabea Rogge for support during the campaign at Kongsfjorden, and Elizabeth Reed-Weidner for help during acquisition of drone data at Hansbreen.

References

[1] J. C. Burton, J. M. Amundson, R. Cassotto, C. C. Kuo, and M. Dennin, "Quantifying flow and stress in ice mélange, the world's largest granular material," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 20, pp. 5105–5110, 2018.

[2] S. Xie, T. H. Dixon, D. M. Holland, D. Voytenko, and I. Vaňková, "Rapid iceberg calving following removal of tightly packed pro-glacial mélange," *Nature Communications*, vol. 10, no. 1, 2019.

[3] H. Vishnu, G. B. Deane, M. Chitre, O. Glowacki, D. Stokes, and M. Moskalik, "Vertical directionality and spatial coherence of the sound field in glacial bays in Hornsund Fjord," *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3849–3862, dec 2020. [Online]. Available: http://asa.scitation.org/doi/10.1121/10.0002868

[4] M. C. Zeh, O. Glowacki, G. B. Deane, M. S. Ballard, E. C. Pettit, and P. S. Wilson, "Model-data comparison of sound propagation in a glacierized fjord with a variable ice top-boundary layer," *The Journal of the Acoustical Society of America*, vol. 145, no. 3, pp. 1887–1887, mar 2019. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.5101839

[5] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," *arXiv:2304.02643*, 2023.

[6] "Segment Anything," 2024. [Online]. Available: https://segment-anything.com/

[7] E. J. Hu, yelong shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2106.09685

[8] F. Li, D. A. Clausi, L. Wang, and L. Xu, "A semi-supervised approach for ice-water classification using dual-polarization sar satellite imagery," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 28–35.

[9] B. Evans, A. Faul, A. Fleming, D. G. Vaughan, and J. S. Hosking, "Unsupervised machine learning detection of iceberg populations within sea ice from dual-polarisation sar imagery," *Remote Sensing of Environment*, vol. 297, p. 113780, 2023.

[10] S. Ansari, C. Rennie, S. Clark, and O. Seidou, "Icemasknet: River ice detection and characterization using deep learning algorithms applied to aerial photography," *Cold Regions Science and Technology*, vol. 189, p. 103324, 2021.

[11] N. Panchi, E. Kim, and A. Bhattacharyya, "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," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 18 004–18 019, 2021.

[12] C. Zhang, D. Han, Y. Qiao, J. U. Kim, S.-H. Bae, S. Lee, and C. S. Hong, "Faster segment anything: Towards lightweight sam for mobile applications," *arXiv preprint arXiv:2306.14289*, 2023.

[13] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, 2021.

[14] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021. [Online]. Available: http://dx.doi.org/10.1038/s41592-020-01008-z

[15] 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.

[16] D. A. Slater, F. Straneo, S. B. Das, C. G. Richards, T. J. W. Wagner, and P. W. Nienow, "Localized Plumes Drive Front-Wide Ocean Melting of A Greenlandic Tidewater Glacier," *Geophysical Research Letters*, vol. 45, no. 22, nov 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018GL080763