Corresponding author:
Siddharth Shankar;
Email: sid.shank07@gmail.com

cambridge.org/jog

Semantic segmentation of glaciological features across multiple remote sensing platforms with the Segment Anything Model (SAM)

Siddharth Shankar[1] ⬤, Leigh A. Stearns[1,2] and C. J. van der Veen[3]

[1]Center for Remote Sensing and Integrated Systems, The University of Kansas, Lawrence, KS, USA; [2]Department of Geology, The University of Kansas, Lawrence, KS, USA and [3]Department of Geography & Atmospheric Science, The University of Kansas, Lawrence, KS, USA

Abstract

Semantic segmentation is a critical part of observation-driven research in glaciology. Using remote sensing to quantify how features change (e.g. glacier termini, supraglacial lakes, icebergs, crevasses) is particularly important in polar regions, where glaciological features may be spatially small but reflect important shifts in boundary conditions. In this study, we assess the utility of the Segment Anything Model (SAM), released by Meta AI Research, for cryosphere research. SAM is a foundational AI model that generates segmentation masks without additional training data. This is highly beneficial in polar science because pre-existing training data rarely exist. Widely-used conventional deep learning models such as UNet require tens of thousands of training labels to perform effectively. We show that the Segment Anything Model performs well for different features (icebergs, glacier termini, supra-glacial lakes, crevasses), in different environmental settings (open water, mélange, and sea ice), with different sensors (Sentinel-1, Sentinel-2, Planet, timelapse photographs) and different spatial resolutions. Due to the performance, versatility, and cross-platform adaptability of SAM, we conclude that it is a powerful and robust model for cryosphere research.

1. Introduction

Since the advent of satellite remote sensing platforms in the 1970s, observational data has grown exponentially. In the span of decades, the polar community went from having on average one image a year of a polar study site, to having (potentially) multiple images a day. Accompanying this increase in observations is the need for efficient feature analysis.

Segmentation techniques are needed to quantify the frequency, size, and location of many different features in glaciology – for example, icebergs, terminus position, crevasses, surface water – particularly in rapidly changing polar regions. Traditional remote sensing indices (Normalized Difference Water Index, NDWI) or object detection algorithms have been successfully used to delineate features such as surface water (Chudley and others, 2021), terminus position (Liu and Jezek, 2004; Seale and others, 2011) and icebergs (Sulak and others, 2017; Moyer and others, 2019). However, these techniques rely heavily on image pre-processing, sensor stability and homogeneous environments (e.g. seasonally variable snow or melt or sea ice in the background will impact the classification results). To take advantage of the range of satellite sensors imaging polar regions, a segmentation algorithm that is agnostic of sensor type, or seasonal shifts in the environment, is needed.

The resurgence of artificial intelligence in 2006 (Hinton and others, 2006), followed by the success of AlexNet in 2012 (Krizhevsky and others, 2012), helped to jump-start machine learning and deep learning algorithms. Convolutional Neural Networks (CNNs) that focus on object detection, semantic segmentation, and instance segmentation provide a methodology inspired by the visual cortex to understand various scenes and identify specific objects. As a result, CNNs have recently been used to segment surface lakes (Yuan and others, 2020), crevasses (Lai and others, 2020; Zhao and others, 2022) icebergs (Bentes and others, 2016; Rezvanbehbahani and others, 2020), glacier termini (Krieger and Floricioiu, 2017; Baumhoer and others, 2019; Mohajerani and others, 2019; Zhang and others, 2019), and other features. However, a major roadblock in using CNNs is that they require large training datasets; a robust custom trained segmentation model may require 10 000 training labels. The absence of good training data greatly impacts the performance, and thus utility, of deep learning models in the earth sciences.

The recently-released Segment Anything Model (SAM) by Meta AI Research is a foundational model in the field of artificial intelligence (Fig. 1). Foundational models are deep learning models that are built using a large amount of unlabeled training data through self-supervised learning (Schneider, 2022). As a result, foundational models perform efficiently for instance and semantic segmentation, object classification and detection purposes. Since their inception in 2018, several versions of these large-scale models have been released, such as Dall-E 2 (Ramesh and others, 2022) and GPT-3 (Brown and others, 2020). The key
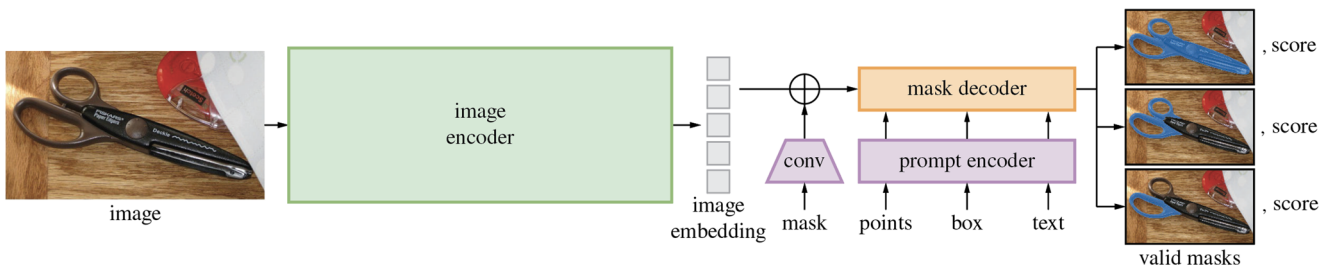
Figure 1. Description of the Segment Anything Model (SAM), which is an image encoder that outputs masks in real-time. Masks are produced for every instance identified, along with the corresponding confidence score (image, with permission from Kirillov and others, 2023).
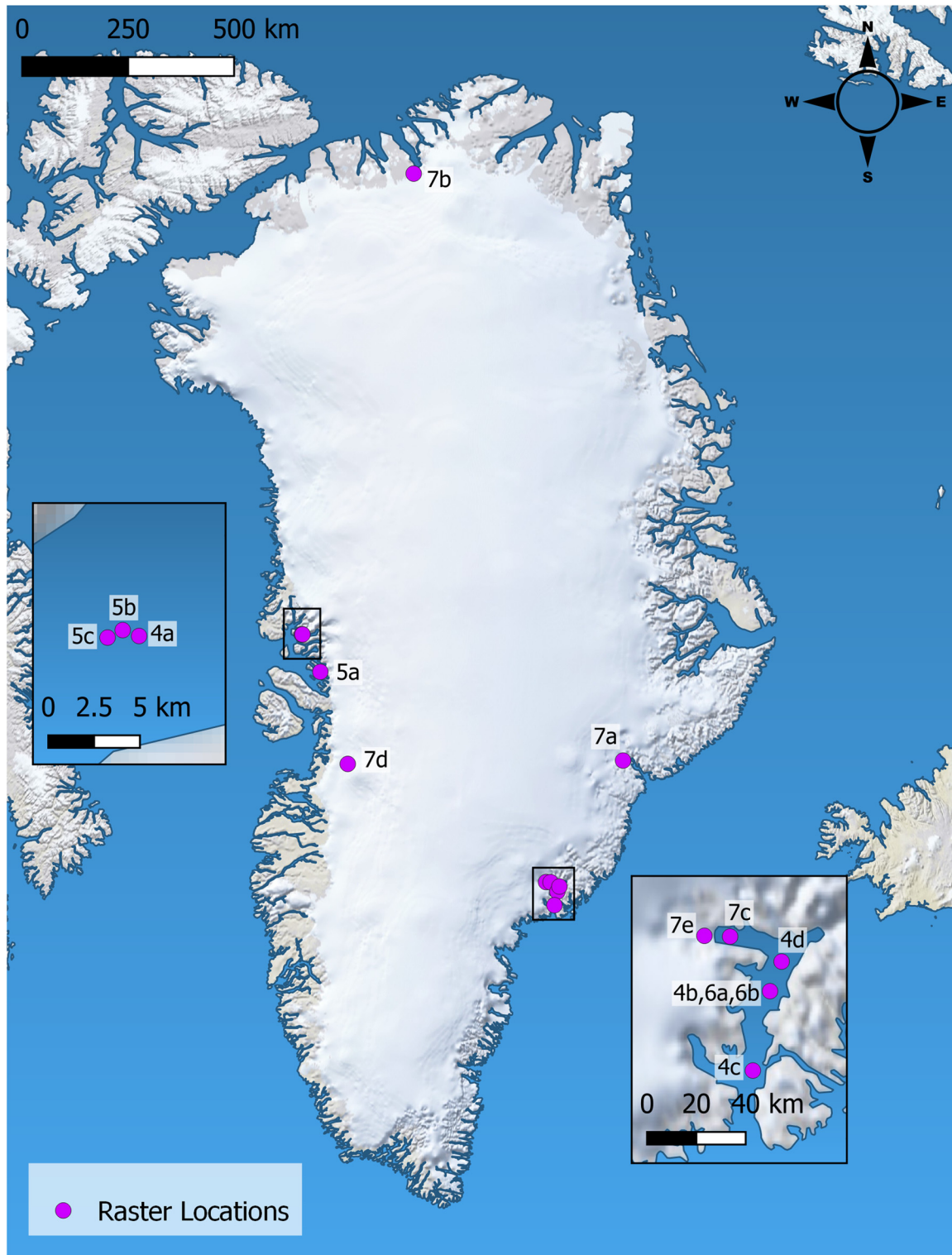

Figure 2. Location of the remote sensing data for SAM analysis, with highlighted figure numbers from this manuscript.

advantage of foundational models is that they allow for generalization of the model through self-supervised learning and a minimum amount of training labels as compared to CNN models.

2. Methods

While SAM does not require training data, model performance can be improved by adding "prompts" (see Supplementary Figs. S1 & S2). Prompts allow the user to identify features of interest (and features that are not of interest) and can be either points or boxes. We quantify the performance of SAM using no-prompts and with-prompts by calculating the F1 score for each image. Our dataset, like most real-world datasets, is imbalanced (the number of features being detected is not evenly balanced by the background). As a result, the F1 metric most accurately represents model performance. The F1 score ranges between 0 and 1; segmentation results with an F1 score close to 1 are good. To prepare ground truth data for validation of the model, we created manual annotations using the V7 labs Darwin (V7Labs, 2023) application and the iPad Pro. The V7 Labs Darwin annotation tool along with the iPad Pro stylus improves the speed, accuracy, and control on labeling significantly (V7Labs, 2023).

Semantic segmentation is an important form of data extraction heavily used within cryosphere research. However, the complexity of segmenting glaciological features makes it difficult to create an automated segmentation approach. With SAM we do not incorporate any additional training data, or pre-process any imagery. There are currently three SAM encoders – ViT-B (Vision Transformer Base), ViT-L (Vision Transformer Large), ViT-H (Vision Transformer Huge) – which have varying numbers of parameter counts. We found that the ViT-L encoder for SAM model performs most consistently for our datasets, so all results are generated with the ViT-L encoder (see Supplementary Fig. S3).

2.1 Data

We acquire Sentinel-1 and Sentinel-2 imagery from Google Earth Engine (Gorelick and others, 2017). The Sentinel-1 SAR image used in this study is obtained in Interferometric Wide Swath (IW) mode at a spatial resolution of 20×22 m (pixel spacing of 10 meters) and with HH, HV, and HHxHV polarization bands. The Planet imagery is obtained from the PlanetScope sensors accessed via the Planet data portal. PlanetScope images are at 3 m spatial resolution. The timelapse imagery is obtained from Stardot Technologies CAM-SEC5-B that has a Standard 4.5–10 mm Varifocal Lens (LEN-MV4510CS). Landsat-4, Landsat-5, and Landsat-8 images are downloaded through the USGS Earth Explorer. We chose these remote sensing platforms as they are commonly used datasets in glaciology research. All the optical images have an RGB band combination. To create a diverse dataset for this analysis, we use images from different regions of the Greenland Ice Sheet as shown in Figure 2.

2.2 Mask generation

For the no-prompt approach, individual instance segmentation results are generated by SAM. SAM can detect multiple feature instances within a single image such glacier termini, icebergs, fjord walls, snow, and water. However, the model is not an object detection model. The model does not recognize that these features are icebergs, sea-ice, terminus, land, water, or something else. The user needs to provide the context of what is present within the image, similar to how most deep-learning models operate. A potential future enhancement of SAM or a derivative of SAM
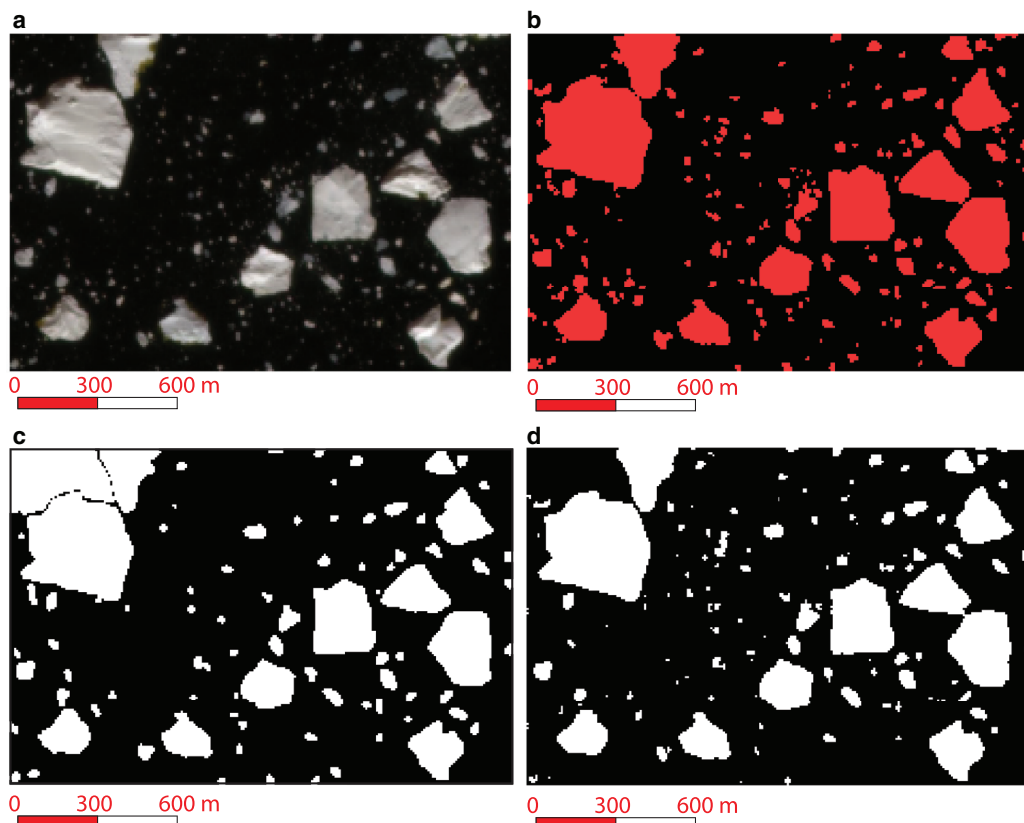

Figure 3. SAM segmentation process, showing (a) the raw satellite image, in this case from Planet, (b) manual iceberg labels, (c) no-prompt SAM segmentation, (d) with-prompt SAM segmentation.

can be to build an object detection model (like the object detection model "You Only Look Once" or YOLO), that identifies if something is an iceberg or glacier terminus. In our extraction of mask instances, we add all the instances detected to a new 2D array of the same shape as the original image. The 2D array will add foreground values (1 s) at location indices of every foreground instance detected. During the testing of the model, there are certain instances where the entire scene is detected as an object. To overcome this, we put in a condition to exclude such instances. We remove any instance larger than 25% of the original image as that suggests a background detection that is too large to be a feature of interest (e.g. an iceberg or supraglacial lake). For other glaciology features such as a glacier terminus, all the instances are saved as an image and the potential feature of interest (glacier terminus) is extracted from the stack of instances that

were identified by SAM. The number of instances in such scenes are small, and is therefore a quick selection process. The terminus for the no-prompt segmentation classifies the glacier and land features on the edge of the image under a single class while the with-prompt segmentation classifies the mélange and land together.

For the with-prompt approach, we create two different shapefiles. One shapefile consists of prompt points representing the foreground or object of interest (value of 1) and the other consists of prompt points representing the background (value of 0). The location coordinates (the row and column values of each prompt) are added to the point shapefiles. Each shapefile is then read and coordinates are extracted as an array with corresponding labels of 1 and 0 s (ones and zeros).

Prompts are selected based on the foreground, or objects that will be segmented (1 s), and background (0 s). The necessity and
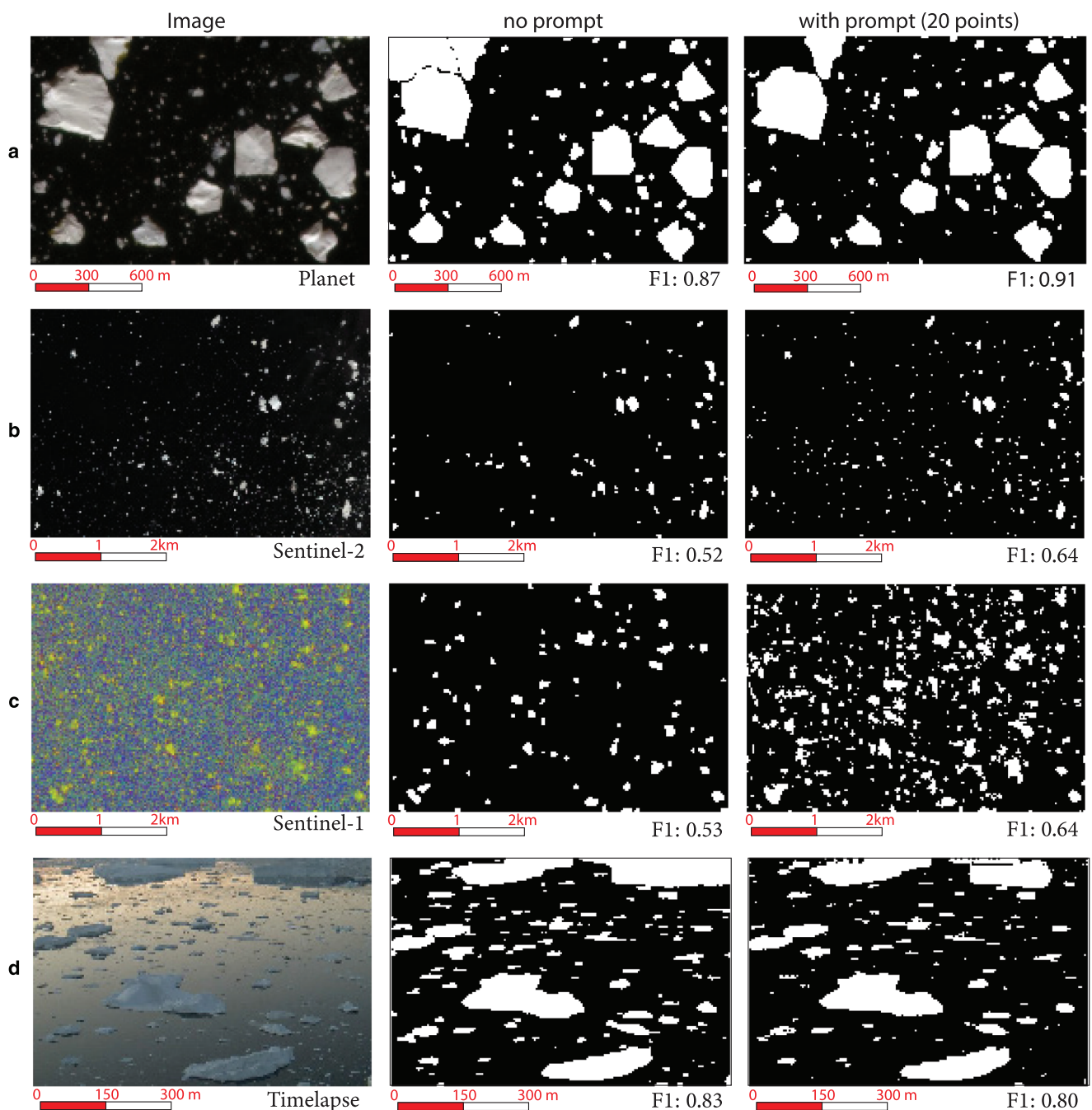

Figure 4. SAM detections of icebergs in open water across different sensors. (a) Planet, (b) Sentinel-2, (c) Sentinel-1, (d) Timelapse photograph. The second column shows segmentation results with no added prompts; the last column shows results with 20 prompt points added (10 points on the icebergs and 10 points for the background). The corresponding confusion matrices of the images can be viewed in Figure S4.

number of prompts are dictated by the density of the objects of interest, the radiometric diversity of the objects in the dataset, and the complexity of the scenes. For example, icebergs and supraglacial lakes are small features (compared to glacier termini) and are found in varied and radiometrically complex backgrounds. This makes the selection process of foreground (1 s) and background (0 s) essential. Other features, such as crevasses, are narrow and sometimes hard to differentiate from the background. In these cases, we place prompts on the background (0 s) adjacent to the objects (1 s) to help the model recognize the importance of these characteristic gradients. For larger features, such as glacier termini, prompts are placed across the terminus to get a range of radiometrically different pixels. The final binary image is created based on the features that the model detects (1 s) and the background (0 s) (Fig. 3).

Validation

To determine the validity of the SAM model, we generate the F1 score for each scene, which is a useful metric for imbalanced datasets (Jeni and others, 2013). The F1 score is quantified for each scene by comparing SAM results with our manual labels, using the following equations which account for true positives (TP), false positives (FP), and false negatives (FN).

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

The F1 score is the harmonic mean of the precision and recall. Precision focuses on minimizing the false positives in the dataset (i.e measuring the true positives in total positives predicted by the model). Recall focuses on minimizing the false negatives in the dataset. In other words, recall measures the correctly identified objects (positives) among the manually-identified objects.

The precision, recall, and F1 score are instrumental in determining the performance of a model in semantic segmentation. The F1 score is between 0 and 1. The closer it is to 1, the better the performance of the model, as it is an overall representation of a strong true positive and strong true negative score. However, it is difficult to set a threshold for an "acceptable" F1 scores, as it will depend on the study. Ideally a model aims to have a very high precision and recall to determine an overall strong F1 score. However, it is important to also assess the precision and recall values of the foreground and the background as this provides a way to determine if the performance of the model is being affected by higher false positives or higher false negatives or both. Another way of doing this, is creating a confusion matrix for each scene, which provides a visual representation of the number of true positives, true negatives, false positives, and false negatives.

3. Results

3.1 SAM performance across different sensors

Many polar remote sensing applications need to balance the acute trade-offs between consistent year-round imagery from Synthetic Aperture Radar (SAR) and long-term, high-resolution imagery from optical remote sensing sensors. Segmentation techniques

Table 1. F1 score of segmentation tests using SAM with no prompts and with prompts (20 points)

Image	Feature	Figures	F1: no prompt	F1: with prompts
Planet	Icebergs (open water)	4a	0.87	0.91
Sentinel-2	Icebergs (open water)	4b	0.52	0.64
Sentinel-1	Icebergs (open water)	4c	0.53	0.64
Timelapse	Icebergs (open water)	4d	0.83	0.80
Landsat-4	Icebergs	5a	0.49	0.44
Landsat-5	Icebergs	5b	0.76	0.70
Landsat-8	Icebergs	5c	0.77	0.80
Sentinel-2 (zoom-out)	Icebergs (open water)	6a	0.52	0.64
Sentinel-2 (zoom-in)	Icebergs (open water)	6b	0.66	0.71
Planet	Crevasses	7a	0.01	0.44
Sentinel-2	icebergs (sea ice)	7b	0.88	0.88
Sentinel-2	Icebergs (mélange)	7c	0.78	0.71
Sentinel-2	Supraglacial lakes	7d	0.48	0.83
Sentinel-2	Glacier terminus	7e	0.96	0.96

.

that work across platforms are therefore critical in building robust datasets (e.g. Zhao and others, 2022). Here, we assess SAM performance across imagery commonly used in glaciology (Fig. 4). Sentinel-1 is a polar-orbiting C-Band SAR with a spatial resolution of 20×22 m. Sentinel-2 is a polar-orbiting optical satellite with a ground resolution of 10 m. We also test SAM segmentation on a suite of Landsat satellites, namely Landsat-4 (60 m), Landsat-5 (30 m), and Landsat-8 (15 m). We include one optical CubeSat sensor, PlanetScope, which has a spatial resolution of 3 m. Finally, we explore the performance of SAM on *in situ* timelapse photographs. Result metrics are shown in Table 1.

In Figure 4a, automatic segmentation does provide good results on the high-resolution Planet image, but is impacted by false positives. These false positives are eliminated by adding 20 points as prompts (10 points identify the icebergs and 10 points identify the background), resulting in an F1 score of 0.91. We find similar improvements with the coarser Sentinel-2 imagery. Adding prompts improves the F1 score from 0.52 to 0.64 – in particular, the prompts help SAM detect smaller icebergs. Larger icebergs were detected successfully with both approaches. There are several small icebergs (hardly visible in Fig. 4b) that are missed even with the prompts, which suppresses the overall F1 score.

The Sentinel-1 image is an RGB composite made up of HH, HV, and HHxHV polarization bands. As with most SAR images, it is noisy, especially compared to the optical remote sensing images. Despite the noise, the no-prompt based approach successfully segments all prominent icebergs. Adding points improves the model performance, particularly because the background itself is so noisy.

Performance on the timelapse photograph is strong; no-prompt approach has an F1 score of 0.83 and with-prompts approach has an F1 score of 0.80. Depending on the range of pixel intensity throughout the image, and the gradients between features of interest and the background, detection can become complex. When the prompts are provided based on specific features, SAM includes such pixels and potential features as part of segmentation. However, in this manuscript we are only using 10 prompts for the foreground and 10 prompts for the background. So based on the complexity of the image scene, the prompts might not be sufficient. In Figure 4d, an increased number of prompts, across the range of gradients, might improve the performance of the with-prompt model.

The Landsat satellite system is the longest running satellite constellation and is widely used in glaciology research, so we

also assess the performance capabilities of SAM with the multiple spatial resolutions that the Landsat satellite system provides. Here we present the segmentation capabilities of SAM on Landsat-4 (60 meters), Landsat-5 (30 meters), and Landsat-8 (15 meters) images from West Greenland (Fig. 5). We find that the SAM no-prompt approach performs better with lower spatial resolution imagery than the with-prompt approach. This result is evident in the F1 score as well as the precision and recall scores of the two approaches. As we transition to higher spatial resolution images of 15 meters in Landsat-8, we find that the with-prompt approach provides better performance and a higher F1 score. It is likely that

the higher-resolution imagery provides stronger gradients between the background and the foreground.

3.2 SAM performance across different zoom levels

Segmentation results from SAM also depends on the size of the object relative to the size of the image. In other words, very small objects, surrounded by a lot of background, are hard to segment. Creating smaller sub-images (from the larger image), adjusts the relative size of the objects, thereby allowing SAM to segment these small objects that were previously discarded in
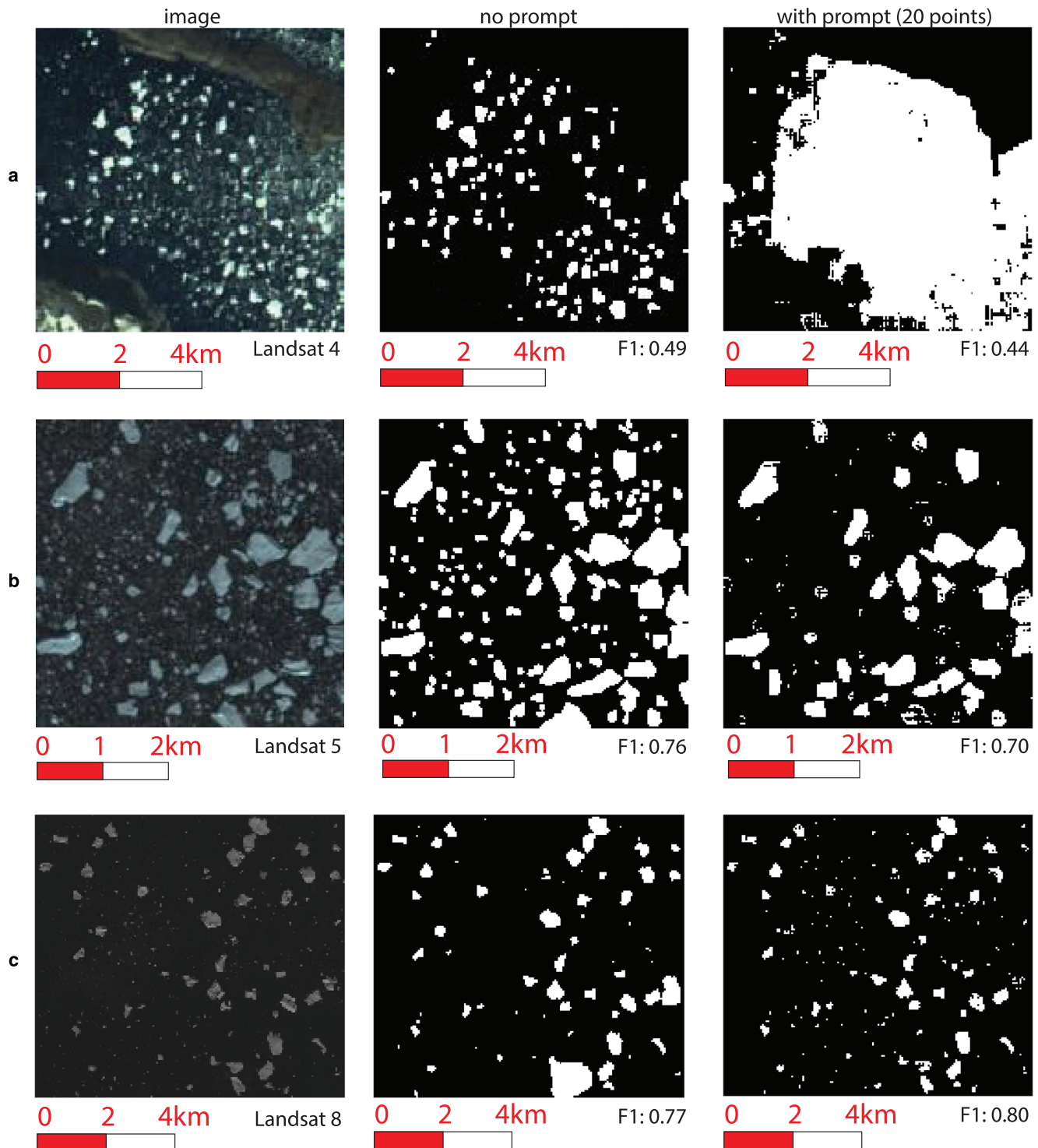

Figure 5. SAM's detection of icebergs in (a) Landsat-4, (b) Landsat-5, (c) Landsat-8. The second column shows segmentation results with no added prompts; the third colum shows results with-prompts. The corresponding confusion matrices of the images can be viewed in Figure S5.
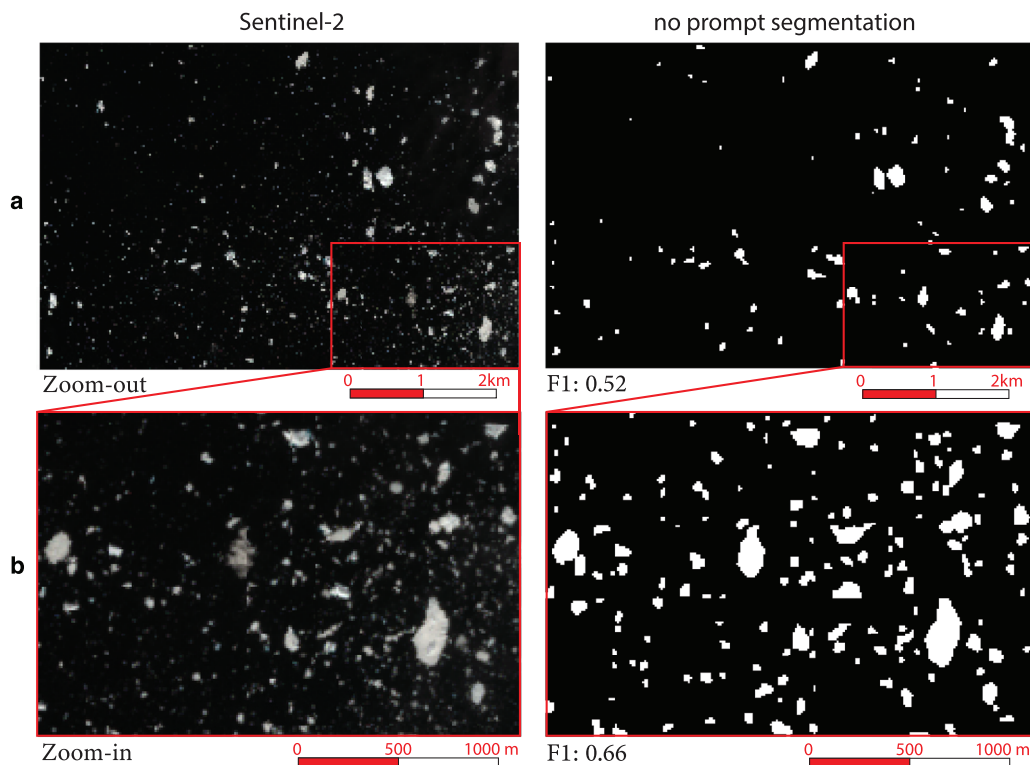
Figure 6. SAM segmentation results scale with zoom level. (a) Segmentation results on a larger Sentinel-2 scene. (b) An inset from the large Sentinel-2 scene showing that at this zoom level, smaller icebergs are detected.

the larger image (Fig. 6a). This approach provides the user control over the feature size of interest. The F1 score improves when zooming in because the model detects more of the smaller icebergs (which were included in the manual labels used to calculate the F1 score, see Supplementary Fig. S1).

Creating subsets of original images, and then mosaicking them together after SAM implementation, does create an additional pre- and post-processing step that is needed when working on larger regions of interest such as a fjord or basin.

3.3 SAM performance across different glaciology features

We assess the broader utility of SAM for cryosphere research by testing it on five different cryosphere features: crevasses, icebergs in sea ice, icebergs in mélange, supraglacial lakes, and a glacier terminus (Fig. 7). We use Sentinel-2 imagery for all the features except crevasses which are generally too narrow to segment in 10 m imagery. For crevasse segmentation, we use Planet imagery.

Our SAM results show that the with-prompt approach provides highly accurate results particularly for supraglacial lakes and terminus positions. Crevasses were essentially undetectable without any prompts, and the F1 score improved to 0.44 with prompts. There are several narrow and short crevasses that, even with 20 prompts, SAM did not detect. It is likely that additional prompts, or a zoomed-in image, would improve this performance. The F1 score for icebergs in sea-ice was consistently strong both with and without prompts (0.88). Icebergs in the mélange were better-detected without prompts (0.78) than with prompts (0.71); in this scenario, the similarity between background and features caused the prompted model to over-estimate the features. In the supraglacial lakes example, we find that due to the presence of two large false-positives in the predicted image, the F1 score of SAM is low at 0.48. The precision and recall scores show that the precision of the model is low for foreground detection (supraglacial lakes) due to the false positives, thereby

impacting the overall F1-score. Adding prompts improves the model substantially.

For iceberg segmentation, no-prompt SAM segmentation provides a fast and fairly consistent result across all sensors, spatial resolutions, and environmental conditions such as open water, sea-ice, and mélange. For the prompt-based approach, prompts were placed in radiometrically different locations to make sure that the model gets a range of sampling. For all features, prompts generally produce better SAM segmentation results.

4. Discussion and conclusion

SAM as a foundational model has been trained on unlabeled training data through self-supervised learning. This training allows the model to be generalized. Additionally, the training dataset for SAM is comprised of 10 million images and over 1.1 billion masks, thereby creating very diverse training data. Convolutional Neural Networks (CNNs) are conventionally trained on large numbers of training data that allows the model to successfully segment objects within an image. However, the overhead of computational efficiency, labeling large and diverse training data, and having enough convolutional layers, makes implementation of the CNN models challenging.

Our implementation of the SAM model shows that it is a robust segmentation model with adaptability across different satellite sensors in no-prompt and with-prompt workflows. In noisy images, such as Sentinel-1 SAR, we find that the no-prompt approach identifies all major icebergs in images robustly; using prompts helped the model also detect smaller icebergs along with the prominent icebergs. This is a huge advantage for the SAR community in climate and Earth science, as the adaptability of the SAM model to produce semantic and instance segmentation datasets promotes data fusion workflows, thereby resulting in an overall improvement of temporal resolution. Another important example showing the adaptability of the SAM model
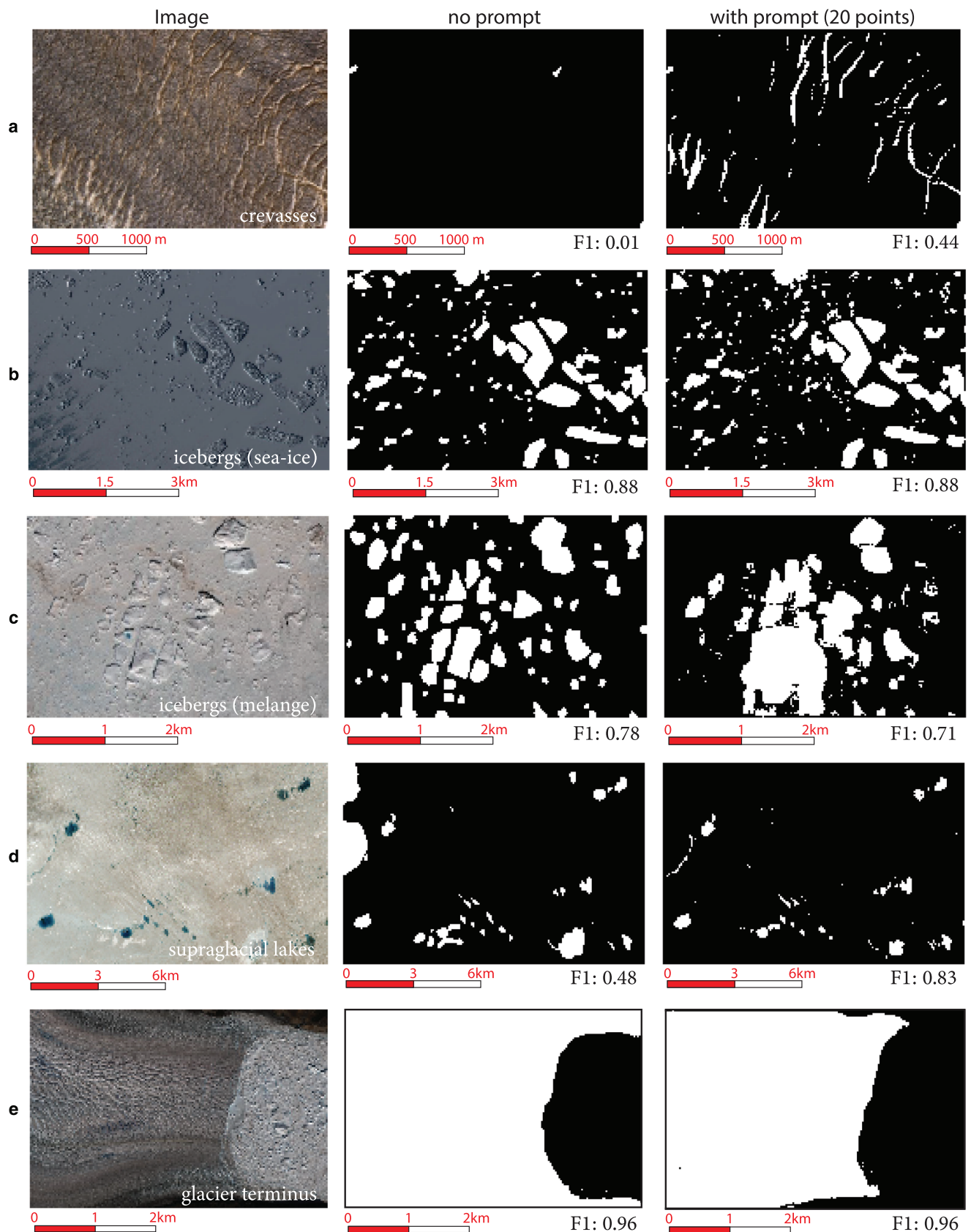
Figure 7. SAM performance on different cryosphere features: (a) crevasses, (b) icebergs in sea ice, (c) icebergs in pro-glacial mélange, (d) supraglacial lakes, and (e) a glacier terminus. All imagery is Sentinel-2, except for Panel (a) that is from © Planet Labs Inc. 2023. All Rights Reserved. The corresponding confusion matrices of the images can be viewed in Figure S6.

is in identifying icebergs in timelapse photographs as shown in Figure 4d. Timelapse is an extremely popular form of imagery in polar research and is used in classification, kinematics, and feature tracking (Messerli and Grinsted, 2015; Giordan and others,

2020). Both the no-prompt and with-prompt segmentation results were strong with an F1 score of 0.83 and 0.80 respectively.

In Rezvanbehbahani and others (2020) iceberg segmentation is done using the CNN model UNet, applied to PlanetScope

imagery. In that study, the F1 score of the iceberg semantic segmentation is 0.89 after extensive training on more than 10 000 manually-digitized iceberg labels. In comparison, semantic segmentation of icebergs on PlanetScope imagery using SAM is 0.87 with no-prompt approach and 0.91 in with-prompt approach. Our results are similar to the highly trained CNN UNet model, without any training data added. With minimum to no-training, successful semantic segmentation of complex glaciological features can be done for images that the model has not been trained.

An interesting aspect in the assessment of SAM was with the implementation of no-prompt and with-prompt approach. We assessed the performance of both the approaches in different conditions, spatial resolutions, and glaciological features. Our results show that, in general, the no-prompt approach works well for images across a wide range of spatial resolutions. A uniform distribution of objects within the image helps provide consistent segmentation results from SAM. However, when this distribution changes or if the objects do not have a strong gradient, then results from the SAM no-prompt approach are more susceptible to false positives. The with-prompt approach provides more support for non-uniform distribution of objects within the images. The with-prompt approach does provide high amount of control in delineation of objects in close proximity or of objects that are difficult to detect, such as crevasses. Low-resolution images, such as Landsat-4 (60 m) and with objects in subtle gradients such as icebergs in mélange, the with-prompt approach was unable to delineate objects. This shortcoming can also be due to the limited number of prompts for the foreground and the backgrounds. For such images increasing the number of prompts and improving the location of prompts will likely improve segmentation results.

We find that SAM is highly adaptable across the different image types that we compared in this study. Results of SAM for PlanetScope, Sentinel-1, Sentinel-2, Landsat-4, Landsat-5, Landsat-8, and timelapse photographs for multiple glaciological features with minimum to no user input, are encouraging. This study shows that developing a segmented dataset across multiple remote sensing platforms is feasible, even in the absence of labeled datasets. Additionally, new remote sensing datasets can be included without sacrificing pre-existing workflows.

4.1 Conclusion

SAM's ability to identify features in simple and complex images means that high temporal resolution datasets can be created by combining segmentation results from optical and SAR remote sensing imagery. For larger regions where the coverage area is several square kilometers, we find that creating subsets of the image by grids aid the model in focusing on smaller features such as lakes and icebergs. This additional step will allow study of glaciological features in more detail as well as scaling SAM to process regions as large as an ice-sheet.

Overall, SAM does provide a comprehensive approach of implementing deep learning in glaciology with faster setup, high accuracy, and minimum to no user input to generate robust segmentation results for different use cases. The no-prompt approach provides consistent results for different features, image types, and spatial resolutions. However, images where foreground object gradients are subtle or features are not as pronounced (e.g. crevasses), the no-prompt approach is unable to segment successfully. The with-prompt approach provides greater control for the object segmentation in such images. Low resolution images however do limit segmentation of smaller features such as icebergs. A potential work around to overcome such limitations of with-prompt approach can be to increase the number of prompts.

This will improve the data diversity and potentially aid the model in identifying features. The dearth of good training holds true for a lot of state of the art deep learning studies in glaciology, thereby limiting the generalizability of such models. For studies where a state of the art model is preferred, SAM can act as an efficient tool for generating large amounts of training data thereby enabling creation of a more generalized model.

References

Baumhoer CA, Dietz AJ, Kneisel C and Kuenzer C (2019) Automated extraction of Antarctic glacier and ice shelf fronts from Sentinel-1 imagery using deep learning. *Remote Sensing* **11**(21), 2529. doi:10.3390/rs11212529

Bentes C, Frost A, Velotto D and Tings B (2016) Ship-iceberg discrimination with convolutional neural networks in high resolution SAR images. In *Proceedings of EUSAR 2016: 11th European conference on synthetic aperture radar*, VDE, pp. 1–4.

Brown T and 9 others (2020) Language models are few-shot learners. *Advances in neural information processing systems* **33**, 1877–1901.

Chudley TR and 7 others (2021) Controls on water storage and drainage in crevasses on the Greenland Ice Sheet. *Journal of Geophysical Research: Earth Surface* **126**(9), e2021JF006287. doi:10.1029/2021JF006287.

Giordan D, Dematteis N, Allasia P and Motta E (2020) Classification and kinematics of the planpincieux glacier break-offs using photographic time-lapse analysis. *Journal of Glaciology* **66**(256), 188–202. doi:10.1017/jog.2019.99

Gorelick N and 5 others (2017) Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment* **202**, 18–27. doi:10.1016/j.rse.2017.06.031

Hinton GE, Osindero S and Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural computation* **18**(7), 1527–1554. doi:10.1162/neco.2006.18.7.1527

Jeni LA, Cohn JF and De La Torre F (2013) Facing imbalanced data–recommendations for the use of performance metrics. In *2013 Humaine association conference on affective computing and intelligent interaction*, IEEE, pp. 245–251.

Kirillov A and 9 others (2023) Segment Anything, preprint arXiv:2304.02643. doi:10.48550/arXiv.2304.02643.

Krieger L and Floricioiu D (2017) Automatic glacier calving front delineation on TerraSAR-X and Sentinel-1 SAR imagery. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, pp. 2817–2820.

Krizhevsky A, Sutskever I and Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. In Pereira F, Burges C, Bottou L and Weinberger K (eds), *Advances in Neural Information Processing Systems*, Vol. 25, Curran Associates Inc.

Lai CY and 7 others (2020) Vulnerability of Antarctica's ice shelves to meltwater-driven fracture. *Nature* **584**(7822), 574–578. doi:10.1038/s41586-020-2627-8

Liu H and Jezek KC (2004) A complete high-resolution coastline of Antarctica extracted from orthorectified Radarsat SAR imagery. *Photogrammetric Engineering & Remote Sensing* **70**(5), 605–616. doi:10.14358/PERS.70.5.605

Messerli A and Grinsted A (2015) Image georectification and feature tracking toolbox: ImGRAFT. *Geoscientific Instrumentation, Methods and Data Systems* **4**(1), 23–34. doi:10.5194/gi-4-23-2015

Mohajerani Y, Wood M, Velicogna I and Rignot E (2019) Detection of glacier calving margins with convolutional neural networks: A case study. *Remote Sensing* **11**(1), 74. doi:10.3390/rs11010074

Moyer A, Sutherland D, Nienow P and Sole A (2019) Seasonal variations in iceberg freshwater flux in Sermilik Fjord, southeast Greenland from Sentinel-2 imagery. *Geophysical Research Letters* **46**(15), 8903–8912. doi:10.1029/2019GL082309

Ramesh A, Dhariwal P, Nichol A, Chu C and Chen M (2022) Hierarchical text-conditional image generation with clip latents. preprint arXiv:2204.06125.

Rezvanbehbahani S, Stearns LA, Keramati R, Shankar S and van der Veen CJ (2020) Significant contribution of small icebergs to the freshwater budget in Greenland fjords. *Communications Earth & Environment* **1**(1), 31. doi:10.1038/s43247-020-00032-3

Schneider J (2022) Foundation models in brief: A historical, socio-technical focus. preprint arXiv:2212.08967.

Seale A, Christoffersen P, Mugford RI and O'Leary M (2011) Ocean forcing of the Greenland ice sheet: calving fronts and patterns of retreat identified by automatic satellite monitoring of eastern outlet glaciers. *Journal of Geophysical Research: Earth Surface* **116**(F3). doi:10.1029/2010JF001847

Sulak DJ, Sutherland DA, Enderlin EM, Stearns LA and Hamilton GS (2017) Iceberg properties and distributions in three Greenlandic fjords using satellite imagery. *Annals of Glaciology* **58**(74), 92–106. doi:10.1017/aog.2017.5

V7Labs (2023) V7 Labs Darwin. https://www.v7labs.com (Online: accessed on 2023-04-23).

Yuan J and 5 others (2020) Automatic extraction of supraglacial lakes in southwest Greenland during the 2014–2018 melt seasons based on convolutional neural network. *Water* **12**(3), 891. doi:10.3390/w12030891

Zhang E, Liu L and Huang L (2019) Automatically delineating the calving front of Jakobshavn Isbræ from multitemporal TerraSAR-X images: a deep learning approach. *The Cryosphere* **13**(6), 1729–1741. doi:10.5194/tc-13-1729-2019

Zhao J, Liang S, Li X, Duan Y and Liang L (2022) Detection of surface crevasses over Antarctic ice shelves using SAR imagery and deep learning method. *Remote Sensing* **14**(3), 487. doi:10.3390/rs14030487