Machine Learning for the Detection of Arctic Melt Ponds from Infrared Imagery

Marlena Reil
University of Osnabrück
IUP, University of Bremen
mareil@uni-osnabrueck.de

Gunnar Spreen
IUP, University of Bremen
gunnar.spreen@uni-bremen.de

Marcus Huntemann
IUP, University of Bremen
huntemann@iup.physik.uni-bremen.de

Lena Buth
Alfred Wegener Institute
lena.buth@awi.de

Dennis G. Wilson
ISAE-Supaero, University of Toulouse
dennis.wilson@isae.fr

Abstract

Melt ponds are pools of water on Arctic summer sea ice that play an important role in the Arctic climate system. Retrieving their coverage is essential to better understand and predict the rapidly changing Arctic, but current data are limited. The goal of this study is to enhance melt pond data by developing a method that segments thermal infrared (TIR) airborne imagery into melt pond, sea ice, and ocean classes. Due to temporally and spatially varying surface temperatures, we use a data-driven deep learning approach to solve this task. We adapt and fine-tune AutoSAM, a Segment Anything-based segmentation model. We make the code, data, and models available online[1].

1 Introduction

The Arctic region is undergoing rapid transformation due to anthropogenic climate change. Temperatures in the Arctic are rising three to four times faster than the global average, a process known as Arctic amplification [2, 38]. Since the beginning of satellite retrievals, scientists have observed a continuing loss of sea ice, with the current rate estimated at 12.7% per decade during summer months [26]. Future projections indicate a summer ice-free Arctic by the middle of this century [31]. These changes have significant implications for local communities and ecosystems [12, 42]. In addition, observations of Arctic warming coinciding with extreme weather events in the mid-latitudes are leading to discussions about linkages of the Arctic to the global climate system [3, 9].

Melt ponds are pools of water that form on Arctic sea ice as a result of summer ice and snow melt when temperatures rise above freezing. They can cover up to 60% and 80% of the ice area on flat sea ice at the peak of melt [6], and range in size from square centimeters to square kilometers [34]. Because of their darker color, melt ponds absorb significantly more sunlight than the white sea ice and snow [7, 11, 29, 35]. This causes surrounding areas to warm up, leading to accelerated ice melt — a positive feedback mechanism that is only stopped when the melt ponds start freezing again in fall. Melt ponds also increase the light transmittance of the surface with implications for the under-ice ecosystem [24, 28, 18]. Accurate measurement of melt pond coverage is fundamental for a variety of climate-relevant tasks, ranging from observing the energy balance at the ice surface [7], to improving coupled local and global climate model predictions [8, 36, 44], and sea ice concentration retrievals

[1] https://github.com/marlens123/autoSAM_pond_segmentation

[19]. However, observational data to date are limited. In Appendix A.1, we provide an overview of existing melt pond measurement methods.

Here, we consider a newly collected dataset of helicopter-borne thermal infrared (TIR) imagery to expand data available for melt pond analysis. TIR measures thermal radiation emitted from the surface [49]. An advantage of TIR over optical sensor data is its independence from lighting conditions. This allows retrieval when the sun is low above the horizon. TIR can also detect melt ponds in winter when they are overfrozen and snow covered [48]. To the best of the authors' knowledge, the data used for this study are the first summer melt pond data using thermal infrared.

For further analysis, we segmented the images into different surface classes. Prior work in this application domain has mainly used pixel-based and object-based approaches such as thresholding, pixel-based neural networks, edge detection, or combinations thereof (e. g., [25, 34, 50, 22, 15, 14, 10, 27, 52]). However, temporally and spatially varying surface temperatures[2] hamper the establishment of a fixed set of classification features for TIR images, and we consider a data-driven approach instead.

The goal of this work is to effectively segment TIR images into melt pond, sea ice, and ocean classes. We rely on deep neural networks as the segmentation tool, as they have been shown to perform very well in segmentation tasks [21]. Specifically, we use an automatic and pre-trained version of the Segment Anything (AutoSAM) model as our model base [13]. We then fine-tune this model with the labeled TIR image data for melt pond segmentation. We fully describe the AutoSAM model in Appendix A.2.

2 Methodology

2.1 Data

Our dataset comprises helicopter-borne TIR imagery acquired with an Infratec Vario-CAM HD head 680 camera during the PS131 ATWAICE campaign [17][3]. A total of 16 flights were conducted at an altitude of about 300 m in July and early August 2022, corresponding to the season of peak melt pond coverage [34]. The geographic area of study is the marginal ice zone of the Fram Strait region. Data is available in NetCDF4 file format with 640×480 pixels per image. Each image records the broadband infrared radiation (7.5μm to 14μm) at roughly 1 m resolution. Image gradient and drift correction were applied prior to this study [49]. No georeferencing was performed in advance, which implies geometric distortions at the image boundaries.

Along with the TIR images, optical sensor data from a Canon camera was acquired and is being processed in parallel. Future efforts include merging both datasets for annotation enhancement and multimodal analysis. This is currently not possible due to different positions of the TIR and visual cameras on the helicopter and different parameter sets used.

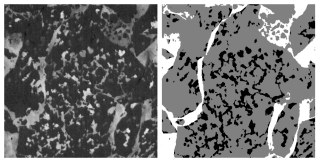
We selected 21 images from flights 7, 9, 10, 11, and 16 (see Table 3.4.2 in [17]) because these were the only flights with good atmospheric conditions[4]. Each flight contains 1000 to 5000 images with overlap. We tried to include a variety of surface conditions, such as different temperatures, sizes, and shapes of ice floes and melt ponds. We did not incorporate images with poor visibility or where surface classes were ambiguous to the human annotator. We hand-labeled each of the 21 selected images pixel-wise into melt pond, sea ice, and ocean using the pencil and filling tool of GIMP 2.10 [47]. In parallel with annotation, we inspected the optical images recorded at the same second to account for visually ambiguous features and reduce annotation uncertainty. For each image, the annotation process took several hours. Figure 1a shows an example of a TIR image with the corresponding annotation.

We converted the temperature values from the TIR camera into grayscale PNG images. To remove major distortions at the image boundaries, we center-cropped the images to

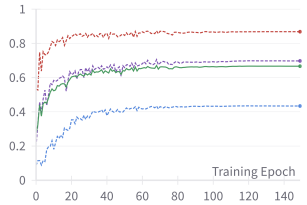
[2]Due to the salinity of ocean water, its freezing point is lowered and ocean can be sometimes warmer, sometimes colder than sea ice.

[3]For more information on the infrared camera, see [49].

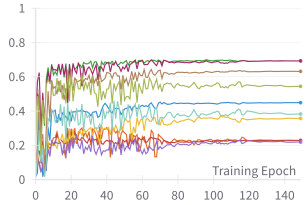
[4]Thermal infrared cannot penetrate through clouds.



(a) Data sample of a helicopter-borne TIR image (left) with annotation mask (right). Black: melt pond, gray: sea ice, white: ocean. Location: Fram Strait region. Date: July 18, 2022 [17].



(b) AutoSAM validation mean IoU (green) and per-class IoU averaged over five runs: melt pond IoU (blue), sea ice IoU (red), ocean IoU (purple).



(c) Visualization of the distribution of AutoSAM's performance on individual validation samples measured in melt pond IoU. Each line and color corresponds to a single validation image.

Figure 1: Data sample and model validation curves.

480×480 pixels. Since SAM is designed for RGB input, we repeated the channel dimension 3 times. The data type was converted from int32 to float32. We applied image-wise z-score normalization, which resulted in better performance than min-max normalization.

We manually split the labeled dataset into 11 training and 10 validation images to ensure balanced image conditions across both sets. The training set was used for model training, and the validation set for both quantitative and qualitative evaluation. We used 10 additional unlabeled images as an independent test set exclusively for qualitative evaluation.

2.2 Model

AutoSAM [13] is an automatic version of the SAM [21] model. SAM's heavyweight pretrained encoder is kept and the mask decoder modified for prompt-free multiclass finetuning. We present a detailed description of the model architecture in Appendix A.2, along with training details in Appendix A.3.

3 Results

We assess model performance using mean and per-class Intersection over Union (IoU) and focus on melt pond IoU to evaluate melt pond detection. Because we observed instability over different trainings, we report metrics averaged over five runs. Hyperparameter findings are reported in Appendix A.4. We tested U-net [40] as comparison model, but as AutoSAM outperforms (Table 1), we leave its descriptions in Appendix A.6.

Table 1: From epoch 120 recorded AutoSAM validation IoUs averaged over five runs, compared to U-net results (Appendix A.6).

	mean	**melt pond**	**sea ice**	**ocean**
AutoSAM	0.667	0.435	0.868	0.698
U-net	0.582	0.320	0.823	0.602

AutoSAM achieves a mean IoU of 0.667 as shown in Figure 1b and Table 1. We observe a high performance difference between the training classes. While the sea ice class is predicted with an IoU of 0.868, melt pond is significantly lower with a difference of 0.433. We attribute this partly to the class imbalance in the training data (melt pond: 0.129, sea ice: 0.651, ocean: 0.220). Figure 1c gives an insight into the performance distribution over different validation samples, which varies greatly. While the validation melt pond IoU is 0.69 for the highest scoring images, it is 0.226 for the lowest scoring image.

Our qualitative results are shown in Figure 2 and Figure 3. We further refer to individual images in the figures as "samples" and indicate them by numbers shown on top of the

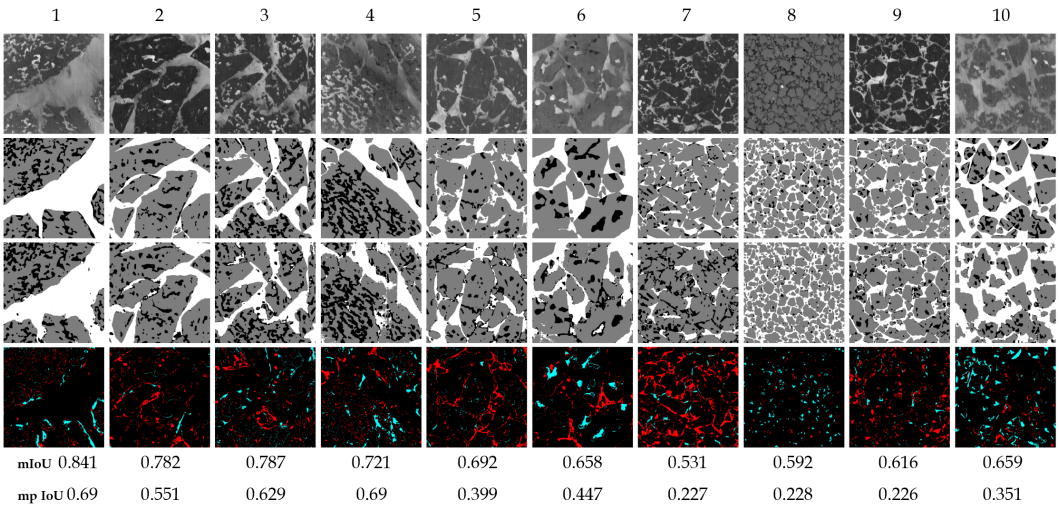


Figure 2: AutoSAM prediction results on the validation images. Rows from top to bottom: input images, ground truths, predictions, error map (shows melt pond false positives in red and melt pond false negatives in blue). Mean IoU (mIoU) and melt pond IoU (mp IoU) of each image below. Black: melt pond, gray: sea ice, white: ocean. Note that while the error map gives a sense of the error patterns, the absolute number of errors is dependent on the occurences of melt ponds, which shifts across the images.

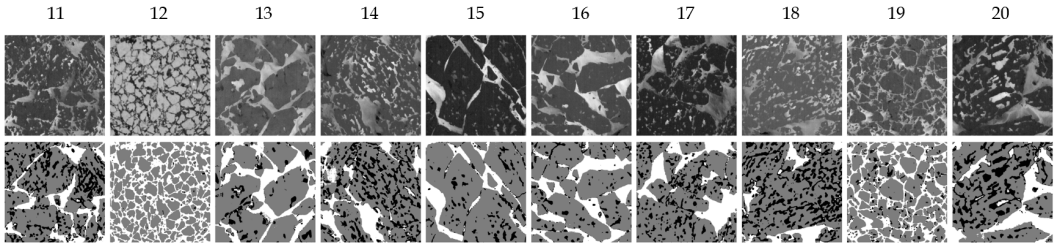


Figure 3: AutoSAM prediction results on unseen images not contained in the training and validation sets.

columns. Generally, AutoSAM is able to capture different surface types independent of their relative temperature difference (e. g., samples 2 and 8). In both the validation and unseen set, we observe good performance on images with large ice floes and visually well separated surface classes, while melt ponds on smaller floes are captured worse. We find that in some predictions, ocean gaps between adjacent floes are misidentified as melt ponds (samples 5,6,7, and 13). Input images with fuzzy boundaries result in interspersed class predictions (samples 4 and 10). We also observe several cases where the interior of a correctly delineated melt pond is misidentified as ocean (samples 3,4,6,11,13,14, and 20). Melt ponds at the edges of ice floes are also partly misidentified as ocean (samples 3,10, and 12). Referring to the error map shown in Figure 2, we see an overestimation of melt pond boundaries (marked red). However, this may partly be caused by annotation uncertainty due to mixed pixels and low resolution in the input images. We provide further analysis in Appendix A.5.

4 Discussion and Conclusion

Melt ponds are essential components of the Arctic climate system but their parameterization is hampered by a lack of observational data. In this study, we addressed this gap by hand-labeling a set of airborne TIR images and using them to fine-tune the AutoSAM segmentation model.

We obtain a mIoU of 0.667 and a melt pond IoU of 0.435, and acknowledge that results are preliminary. Future work can refine the model by constraining physically impossible cases, such as melt ponds surrounded by ocean and ocean surrounded by melt ponds. In addition, Appendix A.6 suggests that combining AutoSAM with a U-net model may be beneficial.

We note that the samples used in this study may not capture the entire data, as images with poor visibility were excluded from annotation and prediction.

However, AutoSAM has shown promise for melt pond segmentation from single-channel TIR data, as it can predict varying surface conditions. This shows the advance of a machine learning approach over physically based methods that are unable to identify surface classes under changing conditions. By refining the method, we contribute to the incorporation of TIR data into melt pond analysis, enabling a light- and season-independent study of the Arctic climate system and helping to understand linkages to global climate change.

References

[1] Lukas Biewald. *Experiment Tracking with Weights and Biases*. Software available from wandb.com. 2020. URL: https://www.wandb.com/.

[2] Intergovernmental Panel on Climate Change (IPCC). "Summary for Policymakers". In: *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014, 1–30. DOI: 10.1017/CBO9781107415324.004.

[3] Judah Cohen et al. "Recent Arctic amplification and extreme mid-latitude weather". In: *Nature geoscience* 7.9 (2014), pp. 627–637.

[4] Jia Deng et al. "ImageNet: a Large-Scale Hierarchical Image Database". In: June 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[5] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *CoRR* abs/2010.11929 (2020). arXiv: 2010.11929. URL: https://arxiv.org/abs/2010.11929.

[6] H Eicken et al. "Hydraulic controls of summer Arctic pack ice albedo". In: *Journal of Geophysical Research: Oceans* 109.C8 (2004).

[7] Florence Fetterer and Norbert Untersteiner. "Observations of melt ponds on Arctic sea ice". In: *Journal of Geophysical Research: Oceans* 103.C11 (1998), pp. 24821–24835.

[8] Daniela Flocco et al. "Impact of melt ponds on Arctic sea ice simulations from 1990 to 2007". In: *Journal of Geophysical Research: Oceans* 117.C9 (2012).

[9] Jennifer A Francis and Stephen J Vavrus. "Evidence for a wavier jet stream in response to rapid Arctic warming". In: *Environmental Research Letters* 10.1 (2015), p. 014005.

[10] Niels Fuchs. *PASTA-ice Github Repository https://github.com/nielsfuchs/pasta_ice*. Version v2023.01. Jan. 2023. DOI: 10.5281/zenodo.7548469. URL: https://doi.org/10.5281/zenodo.7548469.

[11] Thomas C Grenfell and Donald K Perovich. "Seasonal and spatial evolution of albedo in a snow-ice-land-ocean environment". In: *Journal of Geophysical Research: Oceans* 109.C1 (2004).

[12] Grete K Hovelsrud et al. "Arctic societies, cultures, and peoples in a changing cryosphere". In: *Ambio* 40 (2011), pp. 100–110.

[13] Xinrong Hu, Xiaowei Xu, and Yiyu Shi. *How to Efficiently Adapt Large Segmentation Model(SAM) to Medical Images*. 2023. arXiv: 2306.13731 [cs.CV].

[14] W. Huang et al. "Melt pond distribution and geometry in high Arctic sea ice derived from aerial investigations". In: *Annals of Glaciology* 57.73 (2016), 105–118. DOI: 10.1017/aog.2016.30.

[15] Jun Inoue, Judith A Curry, and James A Maslanik. "Application of Aerosondes to melt-pond observations over Arctic sea ice". In: *Journal of Atmospheric and Oceanic technology* 25.2 (2008), pp. 327–334.

[16] L. Istomina et al. "Melt pond fraction and spectral sea ice albedo retrieval from MERIS data – Part 1: Validation against in situ, aerial, and ship cruise data". In: *The Cryosphere* 9.4 (2015), pp. 1551–1566. DOI: 10.5194/tc-9-1551-2015. URL: https://tc.copernicus.org/articles/9/1551/2015/.

[17] Thorsten Kanzow. *The Expedition PS131 of the Research Vessel POLARSTERN to the Fram Strait in 2022*. Ed. by Horst Bornemann and Susan Amir Sawadkuhi. Bremerhaven, 2023. DOI: 10.57738/BzPM_0770_2023.

[18] Christian Katlein et al. "Seasonal evolution of light transmission distributions through Arctic sea ice". In: *Journal of Geophysical Research: Oceans* 124.8 (2019), pp. 5418–5435.

[19] Stefan Kern et al. "The impact of melt ponds on summertime microwave brightness temperatures and sea-ice concentrations". In: *The Cryosphere* 10.5 (2016), pp. 2217–2239.

[20] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[21] Alexander Kirillov et al. *Segment Anything*. 2023. arXiv: 2304.02643 [cs.CV].

[22] Thomas Krumpen et al. "HELIOS, a nadir-looking sea ice monitoring camera". In: *Cold Regions Science and Technology* 65.3 (2011), pp. 308–313.

[23] Jack Landy et al. "Surface and melt pond evolution on landfast first-year sea ice in the Canadian Arctic Archipelago". In: *Journal of Geophysical Research: Oceans* 119.5 (2014), pp. 3054–3075.

[24] Bonnie Light, Thomas C Grenfell, and Donald K Perovich. "Transmission and absorption of solar radiation by Arctic sea ice during the melt season". In: *Journal of Geophysical Research: Oceans* 113.C3 (2008).

[25] Peng Lu et al. "Sea ice surface features in Arctic summer 2008: Aerial observations". In: *Remote Sensing of Environment* 114.4 (2010), pp. 693–699.

[26] Walter N Meier and Julienne Stroeve. "An updated assessment of the changing Arctic sea ice cover". In: *Oceanography* 35.3/4 (2022), pp. 10–19.

[27] Xin Miao et al. "Object-based detection of Arctic sea ice and melt ponds using high spatial resolution aerial photographs". In: *Cold Regions Science and Technology* 119 (2015), pp. 211–222. ISSN: 0165-232X. DOI: https://doi.org/10.1016/j.coldregions.2015.06.014. URL: https://www.sciencedirect.com/science/article/pii/S0165232X15001433.

[28] Marcel Nicolaus et al. "Changes in Arctic sea ice result in increasing light transmittance and absorption". In: *Geophysical Research Letters* 39.24 (2012).

[29] Marcel Nicolaus et al. "Seasonality of spectral albedo and transmittance as observed in the Arctic Transpolar Drift in 2007". In: *Journal of Geophysical Research: Oceans* 115.C11 (2010).

[30] Hannah Niehaus et al. "Sea Ice Melt Pond Fraction Derived From Sentinel-2 Data: Along the MOSAiC Drift and Arctic-Wide". In: *Geophysical Research Letters* 50.5 (2023), e2022GL102102.

[31] Dirk Notz and Julienne Stroeve. "The trajectory towards a seasonally ice-free Arctic Ocean". In: *Current Climate Change Reports* 4 (2018), pp. 407–416.

[32] Lucas Prado Osco et al. "The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot". In: *International Journal of Applied Earth Observation and Geoinformation* 124 (2023), p. 103540. ISSN: 1569-8432. DOI: https://doi.org/10.1016/j.jag.2023.103540. URL: https://www.sciencedirect.com/science/article/pii/S1569843223003643.

[33] Adam Paszke et al. "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32 (2019).

[34] DK Perovich, WB Tucker III, and KA Ligett. "Aerial observations of the evolution of ice surface conditions during summer". In: *Journal of Geophysical Research: Oceans* 107.C10 (2002), SHE–24.

[35] DK Perovich et al. "Seasonal evolution of the albedo of multiyear Arctic sea ice". In: *Journal of Geophysical Research: Oceans* 107.C10 (2002), SHE–20.

[36] Chris Polashenski, Donald Perovich, and Zoe Courville. "The mechanisms of sea ice melt pond formation and evolution". In: *Journal of Geophysical Research: Oceans* 117.C1 (2012).

[37] Alec Radford et al. "Learning transferable visual models from natural language supervision". In: *International conference on machine learning*. PMLR. 2021, pp. 8748–8763.

[38] Mika Rantanen et al. "The Arctic has warmed nearly four times faster than the globe since 1979". In: *Communications Earth Environment* 3 (Aug. 2022), p. 168. DOI: 10.1038/s43247-022-00498-3.

[39] Simiao Ren et al. "Segment anything, from space?" In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024, pp. 8355–8365.

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.

[41] Anja Rösel and Lars Kaleschke. "Comparison of different retrieval techniques for melt ponds on Arctic sea ice from Landsat and MODIS satellite data". In: *Annals of Glaciology* 52.57 (2011), 185–191. DOI: 10.3189/172756411795931606.

[42] Jasmine E Saros et al. "Arctic climate shifts drive rapid ecosystem responses across the West Greenland landscape". In: *Environmental Research Letters* 14.7 (2019), p. 074027.

[43] RK Scharien et al. "First-year sea ice melt pond fraction estimation from dual-polarisation C-band SAR–Part 2: Scaling in situ to Radarsat-2". In: *The Cryosphere* 8.6 (2014), pp. 2163–2176.

[44] David Schröder et al. "September Arctic sea-ice minimum predicted by spring melt-pond fraction". In: *Nature Climate Change* 4.5 (2014), pp. 353–357.

[45] Yasuhiro Tanaka and Randall Kenneth Scharien. "Potential of melt pond fraction retrieval from high spatial resolution AMSR-E/2 channels". In: *IEEE Geoscience and Remote Sensing Letters* 19 (2020), pp. 1–5.

[46] Yasuhiro Tanaka et al. "Estimation of melt pond fraction over high-concentration Arctic sea ice using AMSR-E passive microwave data". In: *Journal of Geophysical Research: Oceans* 121.9 (2016), pp. 7056–7072. DOI: https://doi.org/10.1002/2016JC011876. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JC011876. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JC011876.

[47] The GIMP Development Team. *GIMP*. Version 2.10.12. June 12, 2019. URL: https://www.gimp.org.

[48] Linda Thielke et al. "Preconditioning of summer melt ponds from winter sea ice surface temperature". In: *Geophysical Research Letters* 50.4 (2023), e2022GL101493.

[49] Linda Thielke et al. "Sea ice surface temperatures from helicopter-borne thermal infrared imaging during the MOSAiC expedition". In: *Scientific Data* 9.1 (2022), p. 364.

[50] Mark A. Tschudi, J. A. Curry, and J. A. Maslanik. "Airborne observations of summertime surface features and their effect on surface albedo during FIRE/SHEBA". In: *Journal of Geophysical Research: Atmospheres* 106.D14 (2001), pp. 15335–15344. DOI: https://doi.org/10.1029/2000JD900275. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000JD900275. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900275.

[51] Mark A. Tschudi, James A. Maslanik, and Donald K. Perovich. "Derivation of melt pond coverage on Arctic sea ice using MODIS observations". In: *Remote Sensing of Environment* 112.5 (2008). Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue, pp. 2605–2614. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2007.12.009. URL: https://www.sciencedirect.com/science/article/pii/S0034425708000047.

[52] N. C. Wright and C. M. Polashenski. "Open-source algorithm for detecting sea ice surface features in high-resolution optical imagery". In: *The Cryosphere* 12.4 (2018), pp. 1307–1329. DOI: 10.5194/tc-12-1307-2018. URL: https://tc.copernicus.org/articles/12/1307/2018/.

[53] Nicholas C Wright and Chris M Polashenski. "How machine learning and high-resolution imagery can improve melt pond retrieval from MODIS over current spectral unmixing techniques". In: *Journal of Geophysical Research: Oceans* 125.2 (2020), e2019JC015569.

[54] J.J. Yackel and D. Barber. "Melt ponds on sea ice in the Canadian Arctic Archipelago. Part 2. On the use of RADARSAT-1 synthetic aperture radar for geophysical inversion". In: *Journal of Geophysical Research* 105 (Sept. 2000), pp. 22061–22070. DOI: 10.1029/2000JC900076.

A Appendix

A.1 Methodology to Observe Melt Ponds

Retrieving melt pond coverage is hampered by the remoteness of the Arctic Ocean. Most knowledge to date is based on in situ measurements, where melt pond data are collected directly at the location of interest (e. g., [23, 36, 6]). These methods are needed to understand the underlying processes, but are rare and limited to small areas, thus not suitable for representing the spatially and temporally varying melt pond coverage for the entire Arctic.

Satellites are a promising method in the long term because they can cover large parts of the Arctic on a regular basis. To date, most satellite retrievals of melt ponds have used low- and medium-resolution satellite imagery (e. g., [51, 41]). This does not resolve individual ponds, making accurate acquisition a challenge. Satellite images with higher resolution are available (e. g., [30]), but still limited in spatial coverage [53].

As a trade-off between scale and resolution, airborne campaigns are used. Airborne measurements have the advantage of flexibility, as they can be targeted to areas of interest and avoid cloud cover. They can be used to validate satellite retrievals (e.g., [30, 16]).

At the sensor level, current data range from mostly optical (e. g., [25, 34, 52]) to passive microwave (e. g., [46, 45]) and synthetic aperture radar (e. g., [54, 43]).

A.2 AutoSAM

Our goal is to train a model that can segment TIR data under changing temperature conditions and with limited labeled data available. Our approach is based on the pre-trained Segment Anything (SAM) model [21], which is the first of its kind foundation model for image segmentation. SAM is pre-trained on a large segmentation dataset from the natural image domain (SA-1B). For remote sensing tasks, fine-tuning is recommended [39, 32].

Fine-tuning SAM requires the definition of prompts in the form of points, boxes, or text that identify the object to be segmented. To avoid the costly prompt acquisition for the small and numerous melt ponds, we consider AutoSAM [13], a prompt-free version of SAM.

We introduce essential SAM components and AutoSAM modifications. SAM comprises three components: an image encoder, a prompt encoder, and a mask decoder. The image encoder adopts a ViT-based (ViT-H/16) [5] architecture. It includes 14×14 windowed attention and four global attention blocks. Input images of any size are rescaled to 1024×1024 and downsized to 64×64 output embeddings. The number of channels is reduced to 256 using convolutions. The prompt encoder takes sparse prompts (points, boxes, or text) or dense prompts (masks). Sparse prompts are represented as positional encodings and combined with learned embeddings, or transformed using text encoder [37]. Mask embedding is done by convolutions and the results are combined element-wise with the image embedding. The mask decoder is designed to map image embedding and prompt tokens with an added learned output token to a final mask. It includes a repeated modified transformer decoder block with token self-attention and image-token cross-attention. Thereafter, the image embedding is upsampled and masks are generated based on the point-wise product between the upscaled image embedding and updated tokens.

AutoSAM retains SAM's powerful pre-trained encoder and modifies the (lightweight) mask decoder to allow fine-tuning on non-promptable multiclass images. It does this by removing the prompt tokens. For multiclass segmentation, output tokens and image embedding are replicated by the number of classes. Figure 3 in [13] provides a visualization of the AutoSAM mask decoder.

A.3 Training Details

Our code is based on the AutoSAM implementation [13] using PyTorch [33]. We trained on a NVIDIA GPU Tesla T4 with 16 GB RAM. During training, we froze the SAM encoder so that only the decoder parameters were updated. We applied flipping and rotation augmentation on the fly with a probability of 0.5 each and nearest interpolation for rotation to preserve the mask values. We used a linear combination of categorical crossentropy and dice loss, balanced by class weights. We used the Adam optimizer with a learning rate of 0.0005 and $(\beta 1, \beta 2) = (0.5, 0.999)$ [20]. We used a batch size of 2 and the ReduceLROnPlateau learning rate schedule with a patience of 20. We trained for 150 epochs.

For monitoring our metrics, we used the Weights & Biases dashboard [1]. We chose melt pond IoU as selection criterion for saving model weights and restricted selection to after 100 epochs, as we observed convergence at this point.

A.4 Hyperparameter Findings

We tried more augmentation such as cropping, sharpen/blur, and Gaussian noise to simulate changes in flight altitude and atmospheric effects. We found no improvement in performance. Contrary to the findings of [21, 13], using a larger encoder scale (ViT-l instead of ViT-b) resulted in worse performance on melt ponds. We experimented with a smaller learning rate of 0.0001, but saw no improvement. We tested dropout in the last layer with probabilities of 0.2 and 0.5, but both resulted in worse performance.

A.5 Further Model Analysis

For further analysis, we provide the distribution of melt pond false positives and melt pond false negatives (Figure 4). The plot suggests a general overestimation of melt ponds, as the amount of false positives is higher. False positive melt ponds are more often sea ice ground truth. We assert this to the false boundary predictions found in Section 3. False negative melt ponds are instead more often predicted as ocean, which may be due to the misidentification of ocean within melt ponds and at ice egdes.

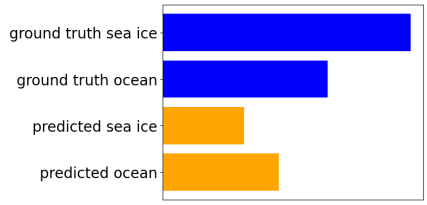


Figure 4: Melt pond false positives (blue) and melt pond false negatives (orange). Computed from the AutoSAM prediction results of the entire validation set.

A.6 U-net Experiments

We performed additional experiments on fine-tuning U-net [40] with a ResNet34 backbone pre-trained on ImagNet [4]. We used the same data splits as in the AutoSAM experiments. Training was performed for 300 epochs. The quantitative results are shown in Figure 5. The qualitative results are shown in Figure 6 and Figure 7. We refer our following comparison to Figure 2 and Figure 3 in Section 3.

The mIoU and per-class IoU are lower for U-net than for AutoSAM (Table 1 in Section 3). Compared to the AutoSAM model, U-net performs worse on images where the relative temperature between surface classes changes (samples 8 and 12). U-net also misidentifies correctly delineated melt ponds as ocean (samples 2, 3, 6, 14, and 20), even to a higher extent than AutoSAM. We observe that while AutoSAM seems to overestimate melt ponds (Figure 2 and Appendix A.5), U-net tends to underestimate them (marked blue in the error map). This suggests that a combination of AutoSAM and U-net may be a reasonable direction for future work.

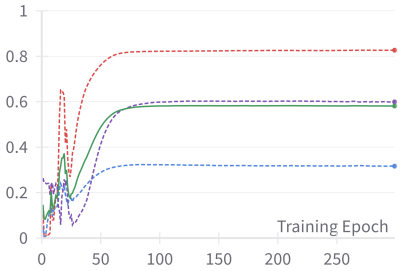


Figure 5: U-net validation mean IoU (green) and per-class IoU: melt pond IoU (blue), sea ice IoU (red), ocean IoU (purple).

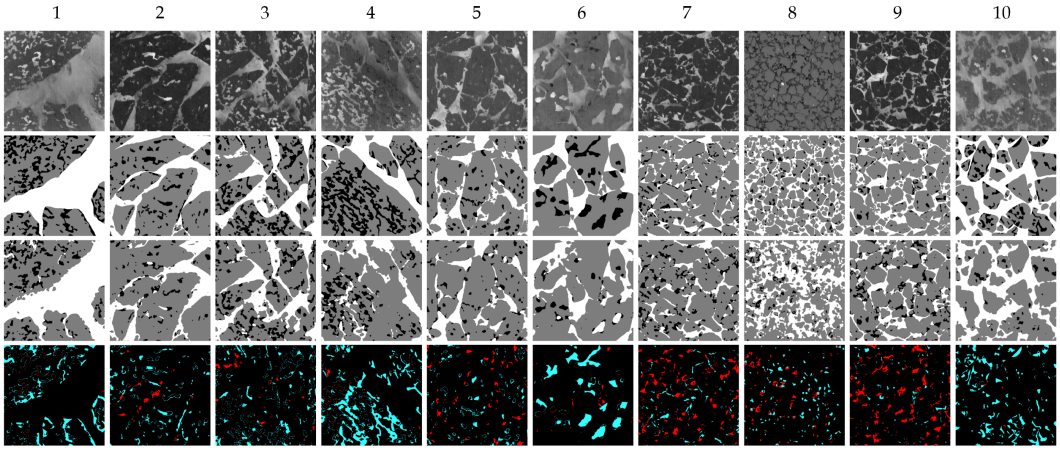


Figure 6: U-net prediction results on the validation images. Rows from top to bottom: input images, ground truths, predictions, error map (shows melt pond false positives in red and melt pond false negatives in blue). Black: melt pond, gray: sea ice, white: ocean.

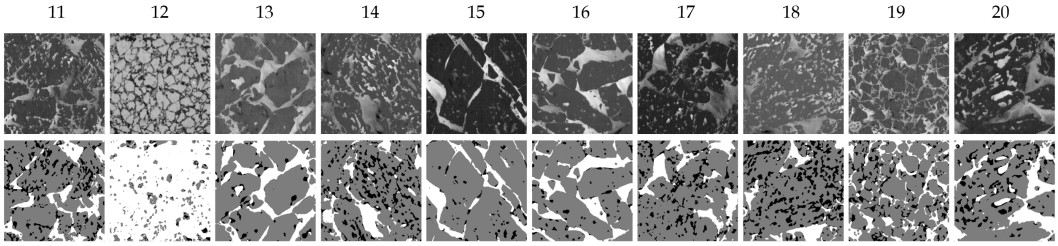


Figure 7: U-net prediction results on unseen images not contained in the training and validation sets.