TSSA: TWO-STEP SEMI-SUPERVISED ANNOTATION FOR RADARGRAMS ON THE GREENLAND ICE SHEET

Atefeh Jebeli[1], *Bayu Adhi Tama*[1], *Vandana P. Janeja*[1,*], *Nicholas Holschuh*[2], *Claire Jensen*[2], *Mathieu Morlighem*[3], *Joseph A. MacGregor*[4], *Mark A. Fahnestock*[5]

[1]University of Maryland, Baltimore County, MD, USA [2]Amherst College, MA, USA
[3]Dartmouth College, NH, USA [4]NASA Goddard Space Flight Center, MD, USA
[5]Geophysical Institute, University of Alaska Fairbanks, AK, USA
[*]*Corresponding author*

ABSTRACT

Ice-penetrating radar surveys have been conducted across the Greenland Ice Sheet since the 1960s, producing radargrams that measure ice thickness and detect the ice sheet's radiostratigraphy. However, these radargrams are relatively under-explored and not yet fully annotated, mapped, or interpreted glaciologically. We aim to move towards automatic radargram annotation using deep learning-based methods. To provide a training set for these methods, we develop a two-step semi-supervised annotation (TSSA) approach that uses an existing unsupervised layer annotation (ARESELP) method and a deep learning-based segmentation approach (U-Net) to detect surface, and bottom reflectors (representing the bedrock) layers in radargrams. Here we focus on two evaluations of our approach: 1. Surface and bottom annotations; and 2. Data augmentation and transfer learning techniques for improving the performance of deep learning methods. Our study is a foundation for improving the efficacy of AI-based methods for auto-annotation of radargrams, where the training set is generated seamlessly through unsupervised learning.

***Index Terms*—** ice sheet, ice penetrating radar, supervised learning, unsupervised learning, deep neural network

1. INTRODUCTION

Ice sheets are important indicators of past and present climate change. The Greenland Ice Sheet (GrIS) has been a particular investigation priority, with regular surveys using ice-penetrating radar since the 1960s [1] producing radargrams that measure ice thickness and capture its radiostratigraphy. These radargrams are used to extrapolate age-depth relationships away from deep ice cores, calculate paleo-accumulation rates and variability, and make inferences concerning historical and contemporary ice dynamics [2, 3]. However, the full potential of these radargrams has yet to be realized because they are not yet fully annotated.

Currently, there are some artificial intelligence (AI)-based approaches for annotation with promising results for synthetic data or expert-curated semi-supervised images [4, 5], requiring extensive time from experts for annotation and interpretation [6, 7]. There are relatively few studies exploring the use of AI techniques to detect ice surface and ice bottom interface in radargrams [8]. One study employed wavelet transforms to reduce speckle noise in radargrams and then conducted training on a deep convolutional neural network using a dataset comprising 920 images, aiming to detect the surface and bottom interfaces [9]. Dong et al. [6] adopted a synthetic approach to generate radargrams, which were subsequently used for training an AI model.

Here we propose a two-step semi-supervised annotation (TSSA) approach using unsupervised learning and deep learning-based segmentation methods to move towards automatic annotation of radargrams. Our approach leverages an existing unsupervised layer annotation (ARESELP [10]) to provide a training set for the deep learning-based segmentation approach (e.g., U-Net [11]). Our study aims to demonstrate the efficacy of AI-based methods for the auto-annotation of radargrams. We focus on two aspects of its evaluation: 1. Surface and bottom annotations; and 2. Data augmentation and transfer learning techniques for improving the performance of deep learning methods.

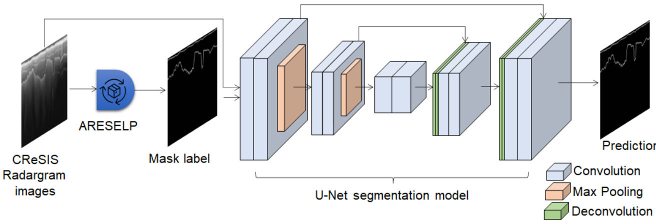


Fig. 1. Overview of the proposed two-step semi-supervised technique for the auto-annotation of radargrams.

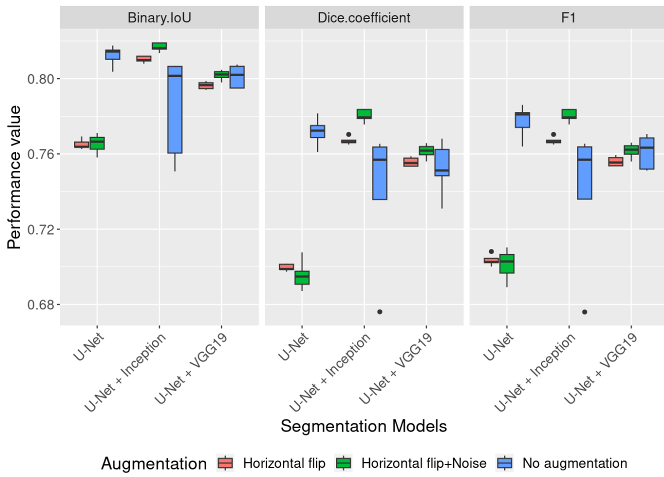


Fig. 2. Performance values of different segmentation models across fold with respect to various augmentation techniques

2. METHODS

The overall proposed two-step semi-supervised annotation technique is shown in Figure 1. The true label for glaciological features (e.g., ice surface and bottom) is generated by ARESELP [10]. ARESELP is an automatic englacial annotation technique that relies on the continuous wavelet transform (CWT)-based peak detection and Hough transform to trace the annotation. In addition, ARESELP avoids the manual selection of seed points by evaluating the peak prominence with wavelet coefficients. We employ U-Net, a widely used deep learning approach for semantic image segmentation [11], which is a method for labeling each pixel in an image with its corresponding class. The U-Net architecture has two main parts: an encoder and a decoder. The encoder network processes the input image using a series of convolutional and maximum pooling layers for feature extraction. These features are then passed to the decoder network, which uses up-sampling and convolutional layers to reconstruct the segmentation mask at the original resolution of the input image. We combine the original U-Net with two different pre-trained models: VGG19 [12] (U-Net + VGG19) and Inception (Inception + VGG19) [13]. Both architectures are built by replacing the encoder part with pre-trained models while keeping the decoder part. We use the pre-trained VGG19 model as a feature extractor, capturing high-level features from the input image, while the decoder path of the U-Net is responsible for upsampling these features to generate the final segmentation map. The U-Net+Inception architecture is built by combining inception blocks with the convolutional layers of the original U-Net architecture. Parallel layers were used only in the feature extraction stage to reduce the number of parameters and computation time due to a large network size [13].

Table 1. Mean performance of various segmentation models under $5CV$. Bold indicates the best performance value.

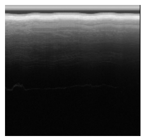
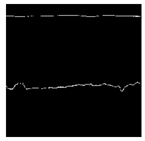
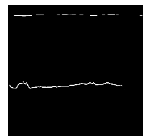
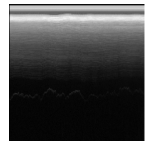
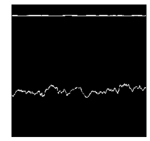
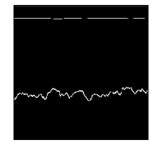
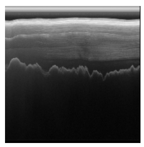
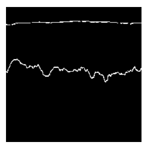
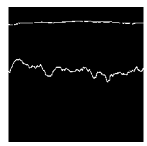
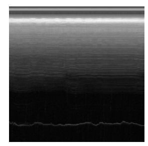
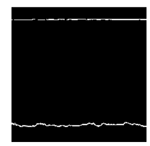
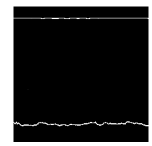
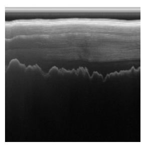
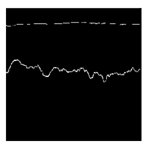
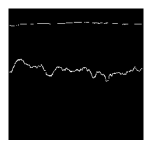
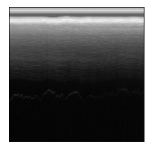
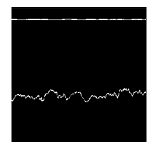
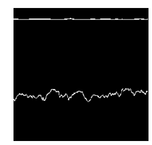
Model	Size of Data	Augmentation	Binary IoU	Dice coefficient	F1
U-Net	425	No	0.812	0.772	0.777
U-Net + VGG19			0.801	0.752	0.761
U-Net + Inception			0.785	0.740	0.740
U-Net	850	Horizontal flip	0.765	0.700	0.704
U-Net + VGG19			0.796	0.756	0.756
U-Net + Inception			0.810	0.767	0.767
U-Net	1275	Horizontal flip + Noise	0.765	0.696	0.701
U-Net + VGG19			0.802	0.761	0.762
U-Net + Inception			**0.817**	**0.780**	**0.780**

3. DATA AND EXPERIMENTAL SETTINGS

The Center for Remote Sensing of Ice Sheets (CReSIS) [1] acquires, processes (including SAR focusing), and provides the radargrams. We examined 425 radargrams, each with a dimension of 1408×1024 pixels and one "color" channel (grayscale radar amplitude). Additionally, a "ground truth" set of ice radar imagery is generated using an unsupervised model (ARESELP [10]), resulting in images with the same dimensions as the original radargrams. Given the limited size of the available dataset, two augmentation techniques are employed to increase the number of images and assess the potential improvement in model performance. Due to the nature of the data, a horizontal flip technique and a combination of horizontal flip with added noise are also applied, resulting in a total of 850 and 1275 images, respectively.

To train our segmentation models, we use an appropriate loss function, specifically the $dice$ coefficient loss, optimize it using the $Adam$ optimizer, and resize images to 512×512. The dataset is divided into training, validation, and test sets to monitor the model's performance and prevent overfitting. For model training, we employed a 5-fold cross-validation ($5CV$) approach and implemented early stopping techniques to ensure the model's generalizability and avoid the model overfitting. Each model is trained for 200 epochs with a batch size of 8. To evaluate model performance, we use several metrics, including binary intersection over union (binary IoU), dice coefficient, and $F1$ score. IoU is the ratio of the area overlap between the predicted segmentation and the true label to the area of union between the predicted segmentation and the true label. For binary-class segmentation, the mean image IoU is determined by averaging the IoU of each class. Dice coefficient and $F1$ are equal to $2\times$, the area of overlap between the predicted segmentation and the true label, divided by the total number of pixels in both the predicted segmentation and the true label images.

Table 2. Example predicted image for 3 segmentation models with horizontal flip and combined horizontal flip + noise.

Model	Horizontal flip			Horizontal flip + noise		
	Radargram	True label	Predicted image	Radargram	True label	Predicted image
U-Net						
U-Net + VGG19						
U-Net + Inception						

4. RESULTS

We implemented three different segmentation models and data augmentation approaches, resulting in nine combinations of segmentation-data augmentation approaches. Table 2 compares the mean performance for each segmentation model over $5CV$. The results indicate that training the model with U-Net+Inception along with data augmentation (horizontal flip and noise added) yields superior results, achieving a performance measure of 0.817 (IoU) as compared to others, such as U-Net + VGG19 at 0.802 and base U-Net at 0.765. The U-Net + Inception architecture has the highest performance scores across all performance metrics. After implementing data augmentation, the effectiveness of the data is enhanced. In particular, the horizontal flip + noise technique increases prediction accuracy, outperforming horizontal flip and no data augmentation. By incorporating the Inception block in U-Net, the model can capture local and global context information, allowing it to better segment the ice surface and bottom at different scales. Figure 2 denotes the performance variability of all segmentation techniques relative to each fold in $5CV$. We find that U-Net + Inception has greater variability (blue in Figure 2) compared to the other two models (U-Net and U-Net + VGG19).

5. DISCUSSIONS

Incorporating augmented data through horizontal flipping during the training process of the U-Net + Inception model demonstrated significantly improved outcomes when compared to training the model exclusively with the original dataset. We found that U-Net + Inception has increased performance variability across the metrics. Increasing the sample size through data augmentation can reduce this variability and provide a more representative sample for every fold [14, 15]. This is also reflected in our findings, where the variance of all performance metrics for U-Net + Inception with data augmentation is significantly lower (red and green in Figure 2) than that of U-Net + Inception without data augmentation.

To better understand and interpret the output of each segmentation model, we show an example of predicted images in Table 2 to qualitatively assess the performance of segmentation models across different data augmentation techniques. Here we compare the predicted images with ground truth annotations from the unsupervised model. It is evident that U-Net + Inception model with horizontal flip + noise data augmentation technique more accurately reproduces the ice surface and bottom annotations, and demonstrates good object completeness. On the other hand, the U-Net model with horizontal flip data augmentation technique does not accurately capture the entire object.

Our study provides a foundational approach for the auto-annotation of radargrams, with potential applications for radargrams that have not yet been manually annotated. By automating the annotation process, we can significantly accelerate the development of a complete database of radiostratigraphy for the Greenland ice sheet, which can then be used to evaluate spatial variation in the multi-millennial-scale sensitivity of this ice sheet to major climate changes, including the last deglaciation.

6. CONCLUSIONS

We present a two-step semi-supervised annotation technique for the automatic radargram annotation. Our technique starts with unsupervised layer annotation (ARESELP), which informs a deep learning-based segmentation approach (U-Net) to detect features. We focused on detecting the surface and bottom initially. Several pre-trained models and data-augmentation techniques were compared to better understand and interpret the model output. We find that the U-Net + Inception model with horizontal flip + noise data augmentation technique was the most effective model. It exhibited excellent object completeness and accuracy. We expect to expand this approach to also annotate englacial layers, to examine much larger Greenland datasets, and to incorporate other transfer learning techniques. We also note that the U-Net with noise and augmentation declines in performance. In our future work, we want to explore the various mechanisms of introducing the noise in the base model to evaluate the U-Net performance. In general, we observed performance improvements in the augmentation vs. no data augmentation models; however, empirically this effect may also be enhanced due to the increased size of the data. Thus, in our future work we also want to evaluate our models with very large datasets to study the impact of change in data size. In addition, we want to explore more state-of-the-art deep learning models, such as Transformer-based computer vision models, to obtain better segmentation results.

ACKNOWLEDGMENTS

This work is funded by the National Science Foundation (iHARP, Award #2118285). The authors would like to thank Dr. Sanjay Purushotham for insightful discussions.

REFERENCES

[1] CReSIS, “Radar depth sounder data, lawrence, kansas, usa,” Digital Media, 2021. [Online]. Available: http://data.cresis.ku.edu/

[2] J. A. MacGregor, M. A. Fahnestock, G. A. Catania, J. D. Paden, S. Prasad Gogineni, S. K. Young, S. C. Rybarski, A. N. Mabrey, B. M. Wagman, and M. Morlighem, “Radiostratigraphy and age structure of the greenland ice sheet,” *Journal of Geophysical Research: Earth Surface*, vol. 120, no. 2, pp. 212–241, 2015.

[3] S. Franke, P. D. Bons, J. Westhoff, I. Weikusat, T. Binder, K. Streng, D. Steinhage, V. Helm, O. Eisen, J. D. Paden *et al.*, “Holocene ice-stream shutdown and drainage basin reconfiguration in northeast greenland,” *Nature Geoscience*, pp. 1–7, 2022.

[4] M. Rahnemoonfar, M. Yari, J. Paden, L. Koenig, and O. Ibikunle, “Deep multi-scale learning for automatic tracking of internal layers of ice in radar data,” *Journal of Glaciology*, vol. 67, no. 261, pp. 39–48, 2021.

[5] M. Yari, M. Rahnemoonfar, and J. Paden, “Multi-scale and temporal transfer learning for automatic tracking of internal ice layers,” in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 6934–6937.

[6] S. Dong, X. Tang, J. Guo, L. Fu, X. Chen, and B. Sun, “EisNet: Extracting bedrock and internal layers from radiostratigraphy of ice sheets with machine learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.

[7] M. Liu-Schiaffini, G. Ng, C. Grima, and D. Young, “Ice thickness from deep learning and conditional random fields: Application to ice-penetrating radar data with radiometric validation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[8] M. Rahnemoonfar, M. Yari, and J. Paden, “Radar sensor simulation with generative adversarial network,” in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 7001–7004.

[9] H. Kamangir, M. Rahnemoonfar, D. Dobbs, J. Paden, and G. Fox, “Deep hybrid wavelet network for ice boundary detection in radar imagery,” in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 3449–3452.

[10] S. Xiong, J.-P. Muller, and R. C. Carretero, “A new method for automatically tracing englacial layers from MCoRDS data in NW Greenland,” *Remote Sensing*, vol. 10, no. 1, p. 43, 2017.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[12] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[13] I. Delibasoglu and M. Cetin, “Improved u-nets with inception blocks for building detection,” *Journal of Applied Remote Sensing*, vol. 14, no. 4, pp. 044 512–044 512, 2020.

[14] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, “Revisiting unreasonable effectiveness of data in deep learning era,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 843–852.

[15] A. Halevy, P. Norvig, and F. Pereira, “The unreasonable effectiveness of data,” *IEEE intelligent systems*, vol. 24, no. 2, pp. 8–12, 2009.