Authors point-to-point responses Referee 'Comment on tc-2023-80', Anonymous Referee #2, 9 Nov 2023 Please find the author's responses in blue below the reviewer's comments.

Many thanks for the review, which will help to improve the quality of the manuscript.

General comments:

This paper presents a convolutional neural network (CNN) approach to measure and quantify surface elevation change in Greenland and the Antarctic ice sheets via satellite radar altimetry data. Through extensive analysis, the authors show that their proposed method displays improved performance and reduced uncertainty over traditional retrackers.

The primary strengths of this paper are in the thoroughness of analysis of the performance of AWI-ICENet1 and in comparisons to conventional retracking algorithms. Cross point error analysis is a good way of comparing the performance of each method for identifying the ice surface, as it does not rely on a ground truth (as is typical in supervised machine learning).

Another strength of the paper is the construction of a synthetic dataset that, after training a CNN on it, performs at least as well as (if not better than) conventional methods. It is an impressive contribution in itself to be able to construct a synthetic dataset that is sufficiently close in distribution to the training and testing data such that a deep learning model can be adequately trained on the synthetic data alone.

Specific comments:

Despite the strengths and contributions, my main concern for this paper is that it does not situate itself within the context and literature of deep learning approaches applied on data from satellite or airborne sounding of ice sheets. To my knowledge, the majority of this work has involved using deep learning to track ice and bedrock layers beneath the ice surface, but these approaches still seem quite relevant, at least to briefly discuss. These are some such prior works:

1. 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 Trans. Geosci. Remote Sens., vol. 60, pp. 1–12, 2021.
2. 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 Trans. Geosci. Remote Sens., vol. 60, pp. 1-14, 2022.
3. M. H. Garcia, E. Donini, and F. Bovolo, "Automatic segmentation of ice shelves with deep learning," in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2021, pp. 4833–4836.
4. H. Kamangir, M. Rahnemoonfar, D. Dobbs, J. Paden, and G. Fox, "Deep hybrid wavelet network for ice boundary detection in radra imagery," in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2018, pp. 3449–3452.
5. R. Ghosh and F. Bovolo, "TransSounder: A hybrid TransUNet-TransFuse architectural framework for semantic segmentation of radar sounder data," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, 2022.
6. E. Donini, F. Bovolo, and L. Bruzzone, "A deep learning architecture for semantic segmentation of radar sounder data," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2021.
7. Y. Cai, S. Hu, S. Lang, Y. Guo, and J. Liu, "End-to-end classification network for ice sheet subsurface targets in radar imagery," Appl. Sci., vol. 10, no. 7, p. 2501, Apr. 2020.

The authors discuss some prior machine learning methods applied to the cryospheric sciences, but this discussion only includes one deep learning approach (Fayad et al. (2021)). I would recommend that the authors include a brief discussion of what distinguishes Fayad et al. (2021)'s setting/model from the current paper. I would also recommend the authors incorporate an additional discussion of the above (and related) references on page 3, or where relevant.

Thank you for this comprehensive list of additional references. We will insert them in the introduction as ML methods used in radar stratigraphy as most of them are dealing with identifying bed rock and/or internal layers or are used for classification of different ice regimes in radar images. We will also add a brief discussion on the differences of our approach and the one of Fayad et.al. and try to make clear how our approach is different to the given references.

Most of these prior approaches applying CNNs to identify ice and bedrock layers beneath the ice surface use 2D CNNs to capture spatial correlations in the along-track direction. However, to my understanding AWI-ICENet1 only performs 1D convolutions in the radar return at a specific waveform in time. Why was this design choice made? It seems likely that capturing spatial correlations could aid the prediction of a deep learning model, especially in regions where data is noisy and measurements are highly variable. Please add a discussion/comparison of AWI-ICENet1 to prior 2D CNNs methods in the paper.

Thank you for raising this question.

Our choice for a 1D CNN has several reasons. First, it is simple and fast and can directly be applied to level 1B waveform data without any preprocessing. Second, our focus is the individual waveform and not like in radar stratigraphy continuous layering or bedrock. Third, each altimeter waveform data file provided by ESA consist of highly variable numbers of waveforms and this could bear problems for 2D CNNs as they usually expect images of similar dimensions or at least subimages of the smaller dimension. Furthermore, an altimetry 2D radargramm looks much different to airborne soundings as the receiving range window is adjusting to topographic changes. This means that the position of the waveform within the window can suddenly jump. A 2D CNN would try to use spatial or alongtrack correlations and would possibly misinterpret such jumps.

We agree that spatial correlations could help to handle noisy measurements and we could imagine that especially for coastal or ocean altimetry a 2D approach might be suitable as well, as sudden waveform jumps are not expected or could be reduced by a preprocessing which shifts the waveform to a constant range gate. However, over the ice sheet this not appropriate as we face large elevation differences of a couple of hundred meters along track. Lastly "our simple" 1D approach shows very good results and we don't see a need to make it more complicated. In addition our 1D single parameter waveform based approach can be extended to a multi parameter approach to gather more information than just the retracking point from the waveform itself.

The authors motivate the use of a synthetic dataset by discussing how ground truth data cannot be obtained by using airborne or ground-borne sounders due to the different footprint sizes. While the answer may be clear to someone in the cryospheric community, some members of the machine learning community may ask why the ground truth cannot simply be set to be the output from a retracking algorithm that the CNN can simply learn to approximate (albeit potentially improving runtime). I would recommend that the authors briefly address this question in the introduction as well.

Thanks for this advice. We will briefly address this and try to make clear that a radar satellite measurement is an integrated signal over a large area including an unknown and variable signal contribution from the subsurface. The topography in this area is dominating the waveform shape and the volume contribution changes the shape and especially the leading-edge width as well. Thus, the ConvNet cannot simply provide a surface elevation when it is trained by point measurements as obtained by ground truth data like GNSS, airborne laser or satellite laser measurements.

Can the authors also provide a brief description/comparisons of runtimes between the algorithms?

Yes, we will add a table showing the runtimes of the different algorithms

Technical corrections:

There are several typos in the paper, and some of the language is unclear; please proofread the paper closely again. For instance, there are

We will proofread the paper.

two typos in line 144

Thanks, will be changed

line 30 “esa” should be “ESA.”

Thanks, will be changed.

On line 270, there seems to be an extra x.

Thanks, will be removed

In lines 504-505, it is unclear what is meant by “the nature of things.”

We will change the sentence to:

The reason that the correlations are lower for AWI-ICENet1 is that the seasonal h-anomalies are already strongly suppressed, resulting in a much lower correlation with backscatter or LEW and thus reducing the correction.

The wording in line 93 should also be tweaked for grammar and combined with the previous sentence: “Reason is the very different footprint size of the two systems.”

We will change the sentence to:

The reason for this are the very different footprints of the two systems. While the ICESat-2 laser points to areas of less than 0.02 km2 , satellite radar altimeters illuminate large areas of up to 10 km2 , so that the two are not spatially assigned and cannot be directly compared with each other.