Responses to reviewer comments for the article "Mapping Antarctic Crevasses and their Evolution with Deep Learning Applied to Satellite Radar Imagery"

We would like to thank the reviewers for taking the time to assess our article and for providing such positive and insightful feedback. This document contains point-wise responses to each comment made, and details of changes to the manuscript where suggestions have been implemented or concerns addressed. We have implemented essentially all of the suggestions by the reviewers. Responses to the major comments from the reviewers are written in paragraph form while responses to the minor comments are tabulated underneath.

Both reviews expressed that they would like greater explanation of the method used to train the neural networks, some quantitative validation of the fracture maps and some intercomparison with existing methods. We have implemented each of these suggestions while staying within the scope of the original article. The major changes in the article are in the addition of three new figures (3, 6, 10) regarding explanation of the training procedure, quantitative analysis of the performance of the crevasse mapping procedure and comparison between ours and existing methods respectively. Section 2.1.1 ("Bootstrapping Neural Networks") has been largely re-written, we have enlarged section 2.3 ("Evaluation"), and we have added section 4.7 ("A Comparison of Ice Shelf Crevasse Detection Methods") to the discussion.

We have appended to the bottom of this document a manuscript showing the changes made following the reviewer reports. Red shows where text has been removed and blue where we it has been added.

Responses to major comments from Reviewer #1

Reviewer 1: The authors' effort in developing the method presented for fracture mapping with Sentinel images is highly commendable. If the produced dataset undergoes more careful validation steps as suggested below, it is expected to be a valuable asset for the cryosphere community. The potential application of the author's result to enhance the representation of damage in ice flow models is important. However, I do think that a more comprehensive explanation and validation of the fracture detection methodology utilizing machine learning is necessary to verify the robustness of the method and its results. Below are three major comments:

Response: We thank the reviewer very much for their praise of the article and the many suggestions which have led to its improvement. The table at the bottom of this document details our responses to specific comments about the article and we respond to the three major comments of the review below.

Reviewer 1: Firstly, the evaluation of fracture detection presented in the paper is primarily qualitative. To provide a more thorough assessment of the algorithm's performance in detecting fractures, the authors should include quantitative measures. Although the authors mentioned sensitivity and specificity, they did not provide the actual values of these measures. To address this issue, the authors could obtain a labeled ground truth dataset through other automated methods, such as Izeboud & Lhermitte (2023), or manual annotation. The authors mentions manual labels "is likely to be uninformative given the subjective nature of producing manual annotations" but all of the qualitative descriptions regarding "sensitivity" and "specificity" in the evaluation section essentially were based on the authors' subjective judgment regarding what counts as surface crevasses, rifts, and basal crevasses. Therefore the authors already potentially impose subjective judgements. Therefore, it would be more transparent to provide annotated fractures that represent the authors' judgments, and calculate standard quantitative measures of neural network performance, such as sensitivity, specificity, area under the ROC curve, and F1-score. The authors can surely acknowledge the fracture annotation, just like any labels in the glaciology literature, could contain subjective bias. Assuming that fracture annotation is improved in the future, the future users can follow the author's NN training/evaluation procedure to improve performance.

Response: In general, we agree with the reviewer that quantitative validation is a useful and worthwhile practise. Our initial hesitation to undertake it in this case was based on the idea that typical quality metrics for this kind of data are rarely reliable, and often obscure the more appropriate validation technique of visual inspection. There was no attempt made to prejudice the qualitative discussion of the performance of the method, but the reviewer's point that biases are more transparent to the reader when there is data annotated by the authors is a good one. Also, it is true that there are certain things, such as the benchmarking of different datasets, that are difficult to do without ground truth images. In this spirit, we have "ground truthed" three full Sentinel-1 IW acquisition scenes by manual annotation of the images at 50m resolution. Regarding the use of existing datasets/methods as ground truths, none cover the grounded ice and we consider none to be accurate or reliable enough for the floating ice. These were selected to be *challenging* crevasse-mapping regions, but not abnormal, covering a large range of floating and grounded crevasse features as well as regions of steep topography and persistent surface melt. By doing whole scenes at a time we will hopefully allow greater ease of comparison with other SAR-based methods in future. In the case of type-A crevasses, the annotation is a feasible task. However, for the finer type-B crevasses, especially on grounded ice, it is essentially impossible to annotate individual crevasse features - hence the time taken to develop a processing chain that did not require manually annotated training data. Instead, we have compared the locations of type-B "crevasse fields" in our maps to those we can see in the SAR images. As suggested, we have provided ROC curve and AUROC metric for type-A maps, as well reporting confusion matrices for the type-B fields, and a visual representation of the intersection. We have changed section 2.3 "Evaluation" and added a new figure (6) to the manuscript discussing this in detail. We hope the reviewer considers our changes sufficient in fulfilling the requirements for quantitative validation and we think the suggestion has resulted in a more robust article.

Reviewer 1: Secondly, to gain a better understanding of what this method captures in comparison to other existing methods, it is important to conduct some comparisons with existing fracture maps such as Izeboud & Lhermitte 2023 (also Sentinel image with a different method) or Lai et al 2020 (same method with lower resolution images). The code for Izeboud & Lhermitte's method and the fracture map produced by Lai are openly available. It is likely that this method complements existing techniques, as the authors' Unet captures the sharpest fractures, while Lai's map captures smoother/larger features visible in MOA images. The authors' method also appears to detect fine-scaled fractures, which are also captured by Izeboud & Lhermitte's method.

Response: We agree with the reviewer that this kind of intercomparison is a very useful exercise, however, a full and detailed comparison of the different techniques is certainly beyond the scope of this work. For other datasets such as observed mass loss or ice sheet model performance, community intercomparison projects are the primary method for these comparisons, and we think that as more crevasse-mapping methods and datasets become available this sort of project would be of great value. To add to the differences between our dataset and previous efforts in this area identified by the reviewer, our dataset covers the grounded *and* floating parts of the Antarctic Ice Sheet, while previously published efforts have focused on the floating ice. Hence, a comparison on grounded ice would require a greater level of work including collaboration across groups than we cannot commit to for this article. For example, preliminary comparisons I have carried out between our method and that of Izeboud & Lhermitte (2023) for a small section of Pope Glacier indicate that their method is inappropriate in its current form for the extraction of grounded crevasses from SAR data. However, it is not clear whether some combination of SAR image resolution and a particular choice of parameters would yield an acceptable solution (as the method is focussed on floating ice, there is no recommended set of parameters for such images). The method of Izeboud & Lhermitte (2023) is also too time and resource-intensive to conduct many experiments ($\sim$ 3 hours of processing time for a single SAR frame at 50m resolution with a window size of 10x10 pixels using 10 processes on a single node (Intel Xeon processor (E5-2640 v4)) - close to an order of magnitude more compute time than making type-A/B composites using our method). Hence, we believe a comprehensive comparison between our maps and other potential methods is beyond the scope of this work.

However, we do take the reviewers suggestion on board and have therefore performed a small comparison considering only floating ice shelves and of limited coverage, which will allow people to see the main differences in the datasets when it comes to ice shelves. We have performed this comparison for a single Sentinel-1 scene covering the Crosson and Dotson Ice Shelves (one of those used in the new quantitative evaluation section). As per the suggestion of the reviewer, the comparison is between the method presented in this work, the method of Izeboud & Lhermitte 2023, and that of Lai et al., 2020. We respectively refer to these as “S23”, “I23” and “L20” here and in the revised article. There are a number of ways in which to apply I23, and we tried various options for the free parameters. We have tried to be fair in our application of their method to compare its maximum capability with our own, though we cannot guarantee that this maximum has been reached by the standards of the designers of that method. We hope the reviewer feels as we do that collaboration between different research groups on this small section of the article is beyond the scope of this work, but that due diligence has been paid to fairly representing the other methods. The new section is 4.7 “A Comparison of Ice Shelf Crevasse Detection Methods” and the new figure is Fig. 10.

Reviewer 1: Lastly, the method used to generate training data is not well explained in the paper. It would be helpful if the authors could provide visual examples of the training data used in the 4-5 iterations described in lines 96-102. Additionally, the authors should clarify how a neural network used for detecting calving can eventually detect surface crevasses and even surface expression of basal crevasses that appear quite distinct from a calving front. The authors mentioned “manually selected images for which the network performed well at the task of crevasses detection to form an updated training 100 dataset“. However, they did not explain how they selected these images. Does this manual selection include some but not all basal crevasses-like features, so that the final NN represents basal crevasses with low but nonzero prediction probability? Again this training data generation step already involves subjective judgment that the neural network learns from. Therefore, it would be beneficial to provide visual examples of the training data to allow for a more thorough understanding of their methodology.

Response: We thank the reviewer for pointing out this confusion, as has the other reviewer. As such, we have made efforts to rewrite this section of the article (Sec. 2.1) to clarify some of the presentation and have added a new figure (Fig. 3) demonstrating the original calving-front-based training dataset as well as schematic illustration of the bootstrapping procedure with example images.

Responses to minor comments from Reviewer #1:

Reviewer 1		
ID	Reviewer Comment	Response
1	Line 46: Re: "We produce continent-wide maps of fracture". This appears contradictory with lines 136-138 where the authors mentions that large parts of Ronne and Ross ice shelves can't be mapped.	Thank you for pointing out this error. To convey the idea of the maps spanning the continent without implying complete coverage, we have changed instances of "continent-wide" to "pan-continental".
2	Line 62: Please cite classical references, Irwin 2057, for the definition of Mode I, II, III fracture: - Irwin, George R. "Analysis of stresses and strains near the end of a crack traversing a plate." (1957): 361-364.	Thank you for providing this useful reference, we have added it where suggested.
3	Line 65: Add the following references that clearly demonstrate "basal crevasses which can result in visible large-scale depressions in the surface": - Luckman, A., D. Jansen, B. Kulessa, E. C. King, P. Sammonds, and D. I. Benn. "Basal crevasses in Larsen C Ice Shelf and implications for their global abundance." The Cryosphere 6, no. 1 (2012): 113-123. - McGrath, Daniel, Konrad Steffen, Ted Scambos, Harihar Rajaram, Gino Casassa, and Jose Luis Rodriguez Lagos. "Basal crevasses and associated surface crevassing on the Larsen C ice shelf, Antarctica, and their role in ice-shelf instability." Annals of glaciology 53, no. 60 (2012): 10-18.	Again, thank you for the references, we have added them where suggested.
4	Line 66-69; line 84: Define "sharp/narrow." The authors emphasize that sharper features are easier to detect with the method presented here, including surface crevasses, rifts and the "sharpest looking basal crevasses/narrow surface depressions". It is unclear what are "sharpest looking basal crevasses". Can the authors explicitly state what sharpness means, e.g. what are the characteristic sharpness for the fractures to be detected? Can the authors simply use a few altimetry elevation data to demonstrate sharpness over the rifts, surface crevasses and the "sharpest looking basal crevasses" identified using Sentinel 1? Wang et al. 2021's paper includes a few examples of the ICESat-2 data over fractures on Amery that shows the steepness of the crevasse walls. - Wang, Shujie, Patrick Alexander, Qiusheng Wu, Marco Tedesco, and Song Shu. "Characterization of ice shelf fracture features using ICESat-2–A case study over the Amery Ice Shelf." Remote Sensing of Environment 255 (2021): 112266.	In image statistics/computer vision "sharpness" refers to the magnitude of intensity gradients in the image. In this case, in the direction perpendicular to the crevasses. We have added the following statement in brackets "(i.e. where there are large-magnitude intensity gradients perpendicular to the crevasse walls)" reflecting this in line 70.

5	Figure 1a: How do the authors know 3 and 4 correspond to basal crevasses and surface crevasses based on Satellite image, given that they look quite similar on the imagery? The authors can check with ICESat-2, radar profile, higher resolution satellite image, or simply use basal crevasse locations that have been identified in the literature (e.g. Luckman et al 2012; McGrath et al 2012). Given that this paper is focused on fracture mapping and the word "basal crevasses" was mentioned several times, I believe it is important to justify the existence of basal crevasses in a few places where the author indicates the UNet identifies basal crevasses.	This is a very good point and we thank the reviewer for providing references which show basal crevasse locations. The use of Crosson Ice Shelf as an example for the discussion of different crevasse features visible in the SAR images is convenient given the great variety of features that can be seen. The use of "basal" refers in general to crevasses that, heuristically, are visually similar to the basal crevasses that have been identified on, e.g. the Larsen-C Ice Shelf or those following the melt-channel on the Dotson Ice Shelf. However, as the reviewer states, this is not very precise. (I am now of the opinion that those crevasses in box 4 are probably basal crevasses as well, with a rift in the middle, so have removed the box from the figure.) Unfortunately, the only viable options for a dataset that could be helpful in discriminating between the classes are ground-penetrating radar or subshelf imagery/morphology data. As "basal" is used as a class of features throughout the article to refer to many different regions, I have elected to qualify its use rather than to try and justify it with additional datasets.
6	Line 101: Re: "1000 images". Remind the reader how large each image is.	We have changed the sentence "... training dataset of $\sim 10^3$ images ..." to "... training dataset of $\sim 10^3$ 256×256-pixel, 64-bit images ...".
7	Line 110: Re: "a threshold can be applied to produce binary maps, with values varying from 0.3 to 0.5 depending on the features of interest." Can the authors explain how these thresholds are determined?	These had been determined using small examples of annotated type-A features (much smaller than those supplied in the revised version of the manuscript), along with true and false positive rates (TPR & FPR respectively) for different thresholds. The optimal thresholds were chosen to maximise TPR-FPR, for either basal-like and rift-like features. However, rather than explain this in the article, we have chosen to replace the line "... with values varying from 0.3 to 0.5 depending on the features of interest" to "... with optimal values that depend on the features of interest and the desired balance between performance metrics."
8	Figure 3: Is "D" neural network output (mentioned in appendix) or fracture density?	We thank the reviewer for pointing out this confusion. For type-A it is the network output and for type-B it is the filtered/enhanced network output. I have added the following sentence to the end of section 2.1.4: "This results in a normalised score $D \in (0, 1)$ with 1 indicating 'fracture' and 0 indicating 'no-fracture' for each pixel in the map." and the following sentence to the figure caption: "$D \in (0, 1)$ is a normalised score with 1 indicating 'fracture' and 0 indicating 'no-fracture' for each pixel."

9	Line 176-178: Re: "A formal validation of the accuracy of the crevasse mapping results presented here, for example, by comparing our maps with manually annotated satellite images, is challenging. This is especially true for our method which produces a continuous, rather than binary, output" This is not entirely true. Classification with a probability output is a stanford ML task and there are several standard validation metrics used to evaluate the performance of models that produce non-binary output, such as the "Area under the ROC Curve".	We thank the reviewer for pointing out this useful evaluation tool. An early ambition for the dataset was to produce a map with values that should not be compared to the classes "crevasse" and "no crevasse". Instead, the values should represent different types of crevasses, with higher values more likely indicating rifts or sharp surface crevasses, and lower values perhaps representing features that are less likely to be full-thickness or important for ice dynamics. However, this complicates things a fair bit, and is difficult to do in a reliable way. Hence, we have removed this kind of suggestion from the article. For example, we have removed from the evaluation section the following paragraph: "This has a benefit, however, as the preference for a continuous output over binary crevasse maps allows for different features to be represented differently in the same maps without training a network to identify and classify different crevasse types, a task that is difficult even to human observers. I.e. features that look only weakly fracture-like in the SAR images will appear fainter in the crevasse maps. For example, heavily crevassed ice shelves such as the Stancomb-Wills Ice Tongue (Fig. 3a) display a nuanced picture of large rifts and smooth, shallow-looking features." As per the reviewer's helpful suggestion, we use ROC curve/AUROC curve as one of the main measures of performance in the newly written section 2.3.
10	Line 190: Re: "the methods we have developed extract the vast majority of features in the backscatter images to produce our crevasse maps while highlighting very few features erroneously." Without comparison between this method with ground truth, we don't know if the method predicts very few features erroneously.	We hope the reviewer agrees that the quantitative evaluation section has resolved this comment.

11	Line 222-227: The authors mention some smooth features, likely basal crevasses, are also detected with this method. As the method's training data include only sharp edges, I'm puzzled how the NN eventually develops the ability to pick up smooth features. Was that related to the "bootstrapping" procedure to make training data?	Yes, as you correctly identify, this is due to the bootstrapping procedure and the fact that the network was not trained until convergence on the calving front dataset. Essentially, we treat calving fronts as a subset of the set of features corresponding to large, linear, textural discontinuities in the backscatter images. This subset is defined in the end by high-level, semantic information like the identification of sea on one side of the calving front, and ice on the other. Early on during training, the U-net learns to act like an edge detector. We have added to section 2.1.1 a number of lines of text helping to clarify the intuition behind this. Applying this partially trained network to a number of unseen images shows the network performs as a sophisticated edge detector, ignoring small edges due to speckle and picking up macroscopic ones, many of which are smoother in nature than the calving fronts. During bootstrapping, one can favour picking up smoother edges so the network learns that these features are admissible.
12	Line 258: Fracture density. For rifts, does the fracture map count the areas within the rift as fracture, or only the two "sharp-sides" edges? If the latter, the current definition of fracture density can underestimate the effect of fracture on reducing ice viscosity, as it doesn't include the void space in between the rift walls.	This is an interesting point. Whether the network fills in the gap depends on the width of the rift and the presence of icebergs/mélange in the void space between the rift walls. However, this makes little difference in determining the effect on ice dynamics as stresses can't be transmitted through ice that sits between, but is disconnected from, the crevasse walls.
13	Figure 5: Should "Buttressing ratio" be "buttressing number"? Use kappa as defined in the appendix. Shouldn't the buttresing number be < 1? How can it be as large as 3??	We have changed "buttressing ratio" to "buttressing number" as suggested by the reviewer and thank them for pointing out this inconsistency. The formula for *kappa* as written here is the same as defined as the appendix (except for an erroneous factor of 2 that has been fixed). The reason $\kappa > 1$ is that e_2 is often negative. The form $\kappa = 1 - \frac{e_2}{N}$ rather than $\kappa = \frac{e_2}{N}$ is so an unconfined ice shelf spreading out in two directions, which should display little buttressing, would have a buttressing number that tends to 0 rather than 1.
14	Line 480: Clarify what resolution count as "the most-high resolution cases"	I have clarified that I consider less than 10m to be high resolution, and have also added a "reliably" to qualify the statement.

15	Line 490: Re: “There are, however, certain disadvantages to the use of SAR. Some ice shelf crevasses appear smoother on the surface than in optical imagery. For example, the crevasses on Dotson Ice Shelf, that appear faintly in SAR backscatter images at 50 m resolution, can be seen more clearly in the MODIS MOA image over the same region.” Are these fractures visible in MOA mapped in Lai et al 2020?	Yes, this is a good point. These features are mapped very well in the Lai et al., 2020 maps. See the newly written comparison section that shows these Dotson Ice Shelf features well in the MODIS MOA used by Lai et al., 2020.
16	Line 500: Note that it would be extremely difficult to have a reliable crevasse-depth estimate as the crack tip can be generally narrower than the data resolution.	Thank you for pointing this out. I have changed: “... the simplest to implement would be to locate coincident ICESat-2 observations with identified crevasses to assess depth, or to look for discontinuities ...” to: “... the simplest to implement would be to locate coincident ICESat-2 observations with identified crevasses to assess some measure of depth, with the understanding that a reliable estimate of true depth is unattainable in many cases due to the sub-resolution width of the crack tip, or to look for discontinuities ...”.

Responses to major comments from Reviewer #2

Reviewer 2: The manuscript presents a method for automatic detection of certain types of crevasses in Sentinel-1 imagery of the Antarctic Ice Sheet. Further, the derived dataset is introduced and first analyses are derived from the data. Of note is the monthly resolution of the data product and the near pan-Antarctic coverage. Overall, the manuscript is very well written and shows clear promise. I believe that the resulting crevasse maps will be of high use for Antarctic research and allow for a better understanding of ice sheet dynamics. The main concern is with the description of the used methodology and the validation thereof. The scientific value of the derived crevasse maps, and in turn the analyses, hinges in large parts on the reliability of the used neural network approach, which should be more thoroughly evaluated.

Response: We thank the reviewer very much for their complements regarding the article, data and derived analyses. We thank them also for their insightful feedback which has led to a more complete and robust manuscript. A table below document details responses to specific comments, and we address the three major comments below.

Reviewer 2: 1. Neural Network Training: The methodology used is not entirely clear. The bootstrapping approach is non-standard, so it might be helpful to explain it in a bit more detail. Currently, it is not entirely clear which criteria are evaluated to select the bootstrapping samples. Further, please clarify which calving front dataset was used for the initial training round.

Response: We thank the reviewer for pointing out the lack of clarity in this section of the article. We have attempted to present the methods more completely, going into greater detail about the initial training and bootstrapping procedure, and explaining the intuition behind why the training method was designed this way. There is also an additional figure (Fig. 3, section 2.1.) that illustrates schematically part of the training process and shows some example images at different stages.

Reviewer 2: 2. Evaluation of the Model: Quantifying the accuracy of the proposed method is critical for estimating the credibility of the predicted crevasse maps and the subsequent analyses. While I agree that it is not possible to thoroughly compare it to human annotators on a pan-Antarctic+multi-year scale, it would still be interesting to see such a comparison for single scenes. Further, comparing it to simpler (non-DL) methods or existing DL approaches for crevasse detection (e.g. [1], [2]) might be helpful for the readers to better understand the advantages and drawbacks of the newly proposed method.

Response: This point was also raised by Reviewer #1. Rather than repeating much of the response here, please see our response to the first major comment from Reviewer #1. In short, we agree with the reviewer that this kind of validation is a useful practise so we have implemented the suggestion of quantifying the performance of our crevasse maps across three complete Sentinel-1 acquisitions. We have extended section 2.3 ("Evaluation") and added a new figure (Fig. 6) detailing the results of our quantitative evaluation.

Regarding comparison with existing methods, we believe a complete and wide-ranging comparison to be beyond the scope of this article. For grounded crevasses, this is not possible as existing processes focus on floating ice shelves. The method of Izeboud & Lhermitte could potentially be applie in future to grounded ice though our attempts using Sentinel-1 SAR data suggest the use of other sensors would be necessary. However, a small comparison, covering floating ice is possible, and we have carried this out using a single Sentinel-1 frame covering the Crosson and Dotson Ice Shelves. We compare our method with that of Izeboud & Lhermitte 2023, and Lai et al., 2020. This is very far from an exhaustive comparison, but we think it conveys the largest differences between these three methods which we regard as being the most likely to be used for widespread analysis. We find that our method looks to perform best for rifts, shear margins and fine surface crevassing (type-B), while that of Lai et al, 2020 performs well for the basal-like features we do not retrieve in our data. We were unaware of the work of Zhao et al., 2022 referenced by the reviewer, but the data is not publicly available so was not included in the comparison. However, the method is very similar to ours so future intercomparison should be performed. We have added a reference to this work in the introduction in acknowledgement that they were the first to combine SAR data and deep learning for crevasse detection on ice shelves.

Reviewer 2: 3. Inspection of the crevasse maps uploaded as review assets suggests that the method is also sensitive to local changes in texture which are not related to crevasses, like the calving front or ice mélange. This should be discussed.

Response: This is a good point, and we are grateful that the reviewer has taken time to assess the uploaded datasets. The network architecture we created (a smaller and shallower version of the U-Net) was chosen in part because many of the crevasse features we are interested in are textural in nature. Combined with the training process, it is natural that many other things (calving fronts, the edges of icebergs, texture on sea ice and in mélange, and the edges of polynias etc) are features of the network outputs - particularly visible in the type-A maps.

We have added a note to the end of the results section (Sec. 2.2) that discusses this point. We will also mask the sea in the crevasse map dataset when the crevasse maps are uploaded. We have also extended the description of the network architecture and training procedure in section 2.1.1 which will help explain these extra features in the crevasse maps.

Responses to minor comments from Reviewer #2

Reviewer 2		
ID	Reviewer Comment	Response
1	Lines 104f.: If my understanding is correct, two networks (N_A, N_B) are employed to map two types of crevasses (type-A and type-B), and then the results are stacked and a softmax is employed. Is this also done during training or only for inference? What is the benefit of employing two separate networks over a single, multi-class UNet?	We apologise to the reviewer for this confusion, this should have been described more clearly in the article. The two networks are employed to map the two types of crevasses, with a softmax applied to each independently (simply to scale the outputs to between 0 and 1 for the "crevasse" class). The type-A and -B maps are eventually combined (by taking a max over the separate maps) but only after the continent-wide mosaics are made with each type independently. We arrived at this procedure really after a process of trial and error, with the separation of the two classes in the end vastly simplifying the training procedure. We are in the process of combining the two networks into a "dual-headed" approach, where the encoder parts of the networks are kept separate, but the latent representations are combined and share a multi-class decoder - reducing the cost of training and throughput. The decision to separately process the two types of crevasses was also made in part to allow for the future addition of other classes of crevasses without impacting the current performance. For example, type-A mapping does not perform particularly well on the basal-crevasse-like features. More training, starting with N_A could be performed to focus on these features. We hope the reviewer finds the rewritten methods section easier to decipher.
2	Line 106: The mention of "scalar outputs" is in contradiction with the softmax, which is a vector-valued function.	Again, we apologise for the confusion. We take only the first component of the softmax output which gives us the pixel-wise "probability" of an identified fracture. We have changed the line: "A softmax function was used to normalise the output of the networks to the range $(0, 1)$ for each pixel." to "For each network, a softmax function was applied to the output and the channel corresponding to "crevasse" was selected so that the outputs were normalised to the range $(0, 1)$ for each pixel, where 1 represents a high confidence of a crevasse, and 0 a low confidence."

3	Lines 131f.: The quality of the type-B detection depends on the availability of SAR acquisitions from multiple look angles.131f.). With the failure of Sentinel-1B in 2021, the number of available acquisitions has roughly halved. Does this affect the type-B crevasse maps?	This is an astute point! Yes, it has affected the type-B crevasse maps, but not uniformly. The grounded crevasse fields are still recovered well in the monthly mosaics as there are still number of acquisition angles. This is especially true in the Amundsen Sea Sector, where the coverage is greatest. Elsewhere, there is a marginal increase in noise away from crevassed regions. For specific analyses, targeted on a certain area or time-span, some thinking would be required to judge the optimal temporal window over which to mosaic the type-B features. We hope that by highlighting this point it may help justify more SAR data acquisitions at a wide variety of look angles on different tracks, to remove any sampling artifact from future versions of the product.
4	Line 177f.: "This is especially true for our method which produces a continuous, rather than binary, output". This argument can be made for any neural network-based method, so it is not quite convincing as a reason for not providing quantitative evaluations. Further, metrics like AUROC exist for such cases.	We thank the reviewer for pointing this out and refer them to our response to minor comment number 9 of Reviewer #1. As per the reviewer's helpful suggestion, we use ROC curves/AUROC curves as one of the main measures of performance in the newly written section 2.3.
5	Line 181: Were the presented example SAR images hand-picked or randomly chosen?	The presented SAR images were indeed hand-picked. As reviewer #1 noted, this allows for our subjective bias, though we chose these images in an attempt to show the good and bad elements of the data. We chose to display the Amundsen Sea Sector in full as it is the most well-studied part of the AIS so we can assume the reader has some familiarity with its geometry. Additionally, it contains the full range of crevasses seen across the continent. We chose eastern Dronning-Maud Land to show the poor performance of the processing over regions of slush/persistent surface melt (near the grounding line of Baudoin Ice Shelf - c1 & d1) as well as the effectiveness of the type-B crevasse mapping over regions where the features are hard to see in individual SAR frames (upstream of Shirase Ice Shelf c2 & d2).
6	Line 528: "However, a greater number of crevasses can be seen at 10 m resolution". How readily can the proposed methodology be adapted to higher resolutions? Maybe a sentence could be added here outlining the ease/difficulty of adapting to higher resolutions.	This is a good point. It is expected that the transition to higher resolution data should be relatively straightforward, with the largest difficulty coming from tuning the lengthscales in the Parallel Structure Filtering algorithm. Increasing the resolution of the backscatter images we process to the SLC acquisition range resolution of 20m shows that the type-A detection continues to work well. The type-B detection, that appears more sensitive to noise, suffers from the additional speckle we get when the SLC images are not multi-looked in the range direction.

7	Line 381: Typo: “Firstly, it likely”	Thank you for spotting this typo, it has been corrected in the manuscript.