We have implemented the corrections as promised to the two reviewers and the community comment:

Answer to RC1: https://egusphere.copernicus.org/preprints/egusphere-2024-498#RC1

Dear anonymous reviewer,

Thank you very much for your detailed feedback to our manuscript! We greatly appreciate the time spent on formulating detailed comments and suggestions. Please find below the answers to both specific and technical comments:

Specific comments:

We will add a general paragraph at the beginning of the method section, that introduces and describes the overall analysis of the paper better.

We will define validation and test set and strive to correct sentences with ambiguous meaning.

We will add a few sentences to describe the key elements of the HRNet+OCR algorithm in simple words, before entering the technical details.

We will correct the tense in the method and the result section.

Technical comments:

Line 28: We will add the missing comma.

Line 34 : We will change heightens to increases.

Line 39 : We will use in opposition instead of opposed.

Line 50 : We will change this sentence as proposed.

Line 52: We will rephrase this sentence for clarity.

Line 67 : When discs are used to encode clicks the whole area specified by the radius is given the same weight. When clicks are encoded as Gaussians the weight is a Gaussian distribution with decreasing from the center of the click over the area specified by the radius. We will explain what we mean by Gaussians in the revised version of the manuscript.

Line 69 : We will add the missing comma.

Line 73 : We will add the missing comma.

Line 75 : We will remove the word “make”.

Line 78 : The challenge with Interactive Deep Learning models is that during training the human user interactions (clicks in our case) are modeled. In the worst case the modeled behavior does not represent reality well, and the model is useless when used by a human.

Consequently, testing if the way the user interactions are modeled lead to a model speeding up segmentation for a human using it is essential. We will change this passage to clarify for future readers.

Data:

Line 85 : We will change the structure of this sentence as suggested.

Line 86 : We will add a placeholder between the number and unit.

Line 98 : We will rephrase this sentence for clarity.

Line 99 : We will add the missing comma.

Figure 1: We will adapt this figure to improve readability and information retrieval.

Line 104 : We will add the missing comma.

Line 105 : We will add the missing comma.

Line 109 : We will add the missing commas.

Line 113 : We will add the missing comma.

Line 123 : We will add the missing comma.

Line 125 : We will add the missing comma.

Methods:

Line 132 : Yes, this is it 5 pixels, we will add this information to the manuscript.

Line 133 : We will add the missing comma.

Line 137 : We will add the missing comma.

Line 138 : We are not entirely sure we understood your question: The images are not randomly selected, instead there is a fixed split to a training, validation, and test set (see Table 1). The random and iterative sampling strategies in this sentence refer to the simulation of user input in the form of clicks for training, validation, and testing. The simulated user clicks are partially placed in random locations and partially in the area with the largest error (see next line). There is no human clicking involved in training the model. This is also the reason we did a user study (see also comment to line 78)

Line 142 : We will add the missing comma.

Line 157 : We will add the missing comma.

Line 159 : We will split this sentence up for better clarity: “Achieving a high IoU after few clicks makes the model most useful. Consequently, we compare the IoU at click k (for k = 1,2,....,20) averaged over all the images (mIoU@k).”.

Line 162 : We will add the missing comma.

Line 163 : We will change the abbreviation for threshold to T in the whole document to avoid confusion with the timestep in Figure 4.

Line 164 : We will add the missing comma.

Line 167 : We will add the missing comma.

Line 173 : We will add the missing comma.

Line 180 : Yes, this is hyperparameters, we will specify this in the text.

Line 183 : We will add the missing comma.

Line 190 : We will put “we carried out a small user study” at the beginning of the sentence as suggested.

Line 191 : We will add the missing comma.

Line 192 : We will add the missing comma.

Line 194 : We will add the missing comma.

Line 197 : We will add the missing comma.

Line 198 : We will add the missing comma.

Line 201 : We will add the missing comma.

Results:

Line 204 : We will adapt this sentence to improve clarity.

Line 208 : We will add the missing comma.

Line 209 : We will add the missing comma.

Line 212 : We will add the missing comma.

Figure 6 : We do not agree that the text and figures need to have exactly the same font. To improve readability and appearance we will harmonize the font in our figures.

Line 217 : We will add the missing comma.

Figure 7 : It is ground truth, yes. We will replace the abbreviation with “Ground truth” in Figure 7, 9 and 10.

Line 220 : We will add the missing comma.

Line 224 : We will add the missing comma.

Line 225 : We will add the missing comma.

Line 226 : We will change the sentence to: “For more than one fourth of all avalanches, the AvaWeb never reaches the NoC20@85, while for the AvaPic and AvaMix less than 1% of all avalanches never reach an IoU of 85%”

Line 227 : We will add the missing comma.

Line 230 : We will add the missing comma.

Line 232 : We will add the missing comma.

Line 235 : We will add the missing comma.

Figure 8-9-10 : We do not agree that the text and figures need to have exactly the same font. To improve readability and appearance we will harmonize the font in our figures.

Line 241 : We will add the missing comma.

Line 248 : We will add the missing comma.

Line 250-251 : We will rephrase this sentence to make it more clear, or maybe make two sentences.

Line 254 : We will rephrase this sentence to “For clicks 1 to 5, where we had enough samples from all participants, we tested if the differences between the highest and the lowest mIoU are statistically significant. The differences are not significant for IoU@1 and IoU@2 (t-test: p-value: >0.05) but they are statistically significant for IoU@3 (p-value= 0.045), IoU@4 (p-value= 0.034) and IoU@5 (p-value= 0.035).”

Line 257 : We will replace “eachother” by “each other”.

Discussion:

Line 263 : We will remove “outlines” from the sentence.

Line 265 : We will add the missing comma.

Line 271 : We will add the missing comma.

Line 276 : We will add the missing comma.

Line 277 : We will add the missing comma.

Line 279 : We will add the missing comma.

Line 281 : We will add the missing comma.

Line 282 : We will add the missing comma.

Line 284 : We will add the missing comma.

Line 286 : We will add the word “that” to this sentence.

Line 288 : We will add the missing comma and we will harmonize the way “user study” is written throughout the text.

Line 288 : We will change to past tense.

Line 294 : We will add the missing comma.

Line 295 : We will add the missing comma.

Line 296 : We will add the missing comma.

Line 297 : We will add the missing comma.

Line 298 : We will change to “were” in this sentence

Figure 14 : We will add a North arrow to the map.

Line 305 : We will add the missing comma.

Line 206 : We will move the comma.

Line 309 : We will add the missing comma.

Line 310 : We will add the missing comma.

Line 316 : We will add the missing comma.

Line 317 : We will change this sentence to “Compared to the traditional way of mapping avalanches, IAS saves over 90\% time. We believe that the time saved may be even greater since the avalanches with a time recording were rather small (mean size 1.75) and all located in an area well known to the person mapping.” to increase clarity.

Line 317 : We will merge this sentence with the above to avoid a one-sentence-paragraph.

Conclusion:

Line 321 : We changed the tense to were.

Line 322 : We will add the missing comma.

Line 324 : We will add the missing comma.

Line 331 : We will add the missing comma.

Line 331 : We will change to “Assuming the camera position and area captured is stable, the georeferencing can be reused for all subsequent images. In the past this has been done for webcam-based snow cover monitoring (Portenier et al., 2020).”

Line 332 : We will add the missing comma.

Line 338 : We will add the missing comma.

Line 343 : We will remove “as is “ in “The model as is may also be used to”.

Line 343 : We will replace “These” by “this”.

Line 345 : We will add “avalanche annotations” to “thereby getting more accurate and reliable avalanche annotations in the future.”.

Line 345: We will add the missing comma.

Answer to RC2: https://doi.org/10.5194/egusphere-2024-498-RC2

Dear anonymous reviewer,

thank you very much for your detailed comments and suggestions to improve the quality of our manuscript!

Please find below the answers to your comments:
Fig. 1: We will make this figure better readable by changing the way the field of view per camera is displayed.

Sect. 2.2. The dataset includes selected avalanches twice, captured under different illumination conditions. These avalanches are of course not split between the datasets for training and testing. We believe this helps the model to become robust and independent of the illumination conditions. We see no influence on the ability of the AvaWeb to generalize better or worse to unknown view angles.

Sect 3.1. We will restructure and expand the model description for better readability and easier understanding for the readers with little deep-learning background.

Sect 3.3. COCO+LVIS is a combination of COCO (an image dataset for object detection) and LVIS (a large-scale instance segmentation dataset). Both are publicly available datasets with a total of 104k images and 1.6M instance-level masks. They are widely used for training, testing, and comparing models. In our case baseline refers to the HRNet+OCR trained on this dataset. The AvaWeb, AvaPic, AvaMix refer to the HRNet+OCR trained with Avalanche images “on top of” COCO and LVIS. In other words, we use pretrained weights from COCO+LVIS to then fine-tune to avalanches.
HRNet+OCR is a model for semantic segmentation, initially not used for interactive object segmentation. Conv1S is the solution to feed user corrections into the HRNet+OCR without losing the advantages of the pretraining on COCO+LVIS.

We will expand this section in the revised version of the manuscript to make it easier to understand, also for the readers with little deep-learning background. We do not think showing details of the HRNet+OCR in Figure 4 would increase understanding. We will re-emphasize the difference in the size of training datasets in Section 3.3.

Technical comments.

L4. We will change to "becoming more frequent" instead of "getting more frequent".

L32. We will use "often" rather than "oftentimes".

L33. We will change "an" to "a".

L75. We will change to “we propose to use webcam infrastructure” according to the suggestion from reviewer 1.

L85. We will correct "webcams network" to "webcam network".

L122. We will change "their UIBK" to "the UIBK".

L123. We will change "cropped" to "cropping".

L134. We will remove one instance of Fig.4.

L182. We will change "evaluate on the" to "evaluate the".

L190. We will change "users who's" to "users whose".

L225. We will change one fourth to one quarter.

L226. We will change "while this for the" to "while for".

L246. We will change "to a more" to "to greater".

L289. We will replace “beat” by "exceed" to use language more appropriate for a scientific publication.

Answer to CC: https://doi.org/10.5194/egusphere-2024-498-CC1

Dear Ron Simenhois,

thank you for taking the time to read our manuscript and for giving detailed feedback about passages that need improvement for better readability and passages that are ambiguous and need to be corrected for clarity.

Please find below the answers your comments:
Different models: We will go over the manuscript and make it clear that our models

(AvaWeb,..) differ in the data used to train them (and the number of epochs like noted in 3.3), but not in the model architecture.

Model comparison: In our discussion we mention that we believe the coarseness of the annotations in the AvaPic prevents the model from learning all it could from such a large and diverse dataset. We expect the best model performance from training with a large dataset with fine annotations covering various perspectives, avalanche types, avalanche sizes as well as snow and illumination conditions. We will expand this section in the discussion chapter and explicitly describe the implications and recommendations for practice that we have found, picking them up again in the conclusion.

We will correct the typos, punctuation, and grammar inconsistencies in the manuscript as already promised to Reviewer1.

Line 32/324: We will replace the “between 10 and 60 minutes” with a general statement or leave it altogether.

Line 39: True, this would be a good place. We will introduce segmentation here, before the reader needs to know what it is to understand what s/he is reading.

Line 62: We will remove “mask” here.

Line 70: You are right, it is possible to georeference any image where enough persistent objects in the image with known coordinates are identified. We meant to emphasize that for cameras in a stable position this process can be done once and reused for all subsequent images. In contrast, each image with a unique perspective needs to be individually georeferenced, resulting in a comparably higher effort per image. We will adapt our manuscript specifying what we mean.

Line 132: We will add “pixels” as a unit here.

Line 159: We will remove the redundant word “the”.

Line 164: Fox et al. (2023) **trained** the model with an Intersection Over Union (IoU) threshold of 0.2 and a confidence threshold of 0.25. But when **testing** they used an IoU threshold of 0.05 and a “confidence threshold which maximizes the model F1 score” (see caption to Table 2 in Fox et al., 2023).

Line 175: We will rewrite this description to make our point clearer.

Line 227: You are right, it does not come as a surprise that having seen avalanches in training is beneficial to segmenting one later. We believe describing this in more detail in Sect 4 does not fit, but we will mention this in the discussion section.

Line 230ff: Fox et al. (2023) state they achieve an F1 score of 64.0 ± 0.6 which we have correctly copied to Table 4. For the number in this line, we meant to compare F1 scores neglecting standard deviation. The difference is however 0.12 and not 0.13 (0.64 vs. 0.76). We will correct this mistake and add the most important F1 scores to the text to avoid confusion and allow the text to stand-alone.

Table 5: We will add the number of images part of the *UserPic* to the caption of this table.

Line 286: It is identical IoU of 5% for the bounding boxes: Fox et al. (2023)'s confidence threshold for this F1 score is unknown (see comment to line 164). We thresholded our raw predictions, which could also be called model confidence, at 0.5 (see line 153). This value was determined by analyzing mean IoU scores per click on the validation set. Consequently, the F1 scores we compare are both based on the confidence threshold the respective authors found to work best.

Line 296/318: We will add the appropriate unit, in this case the European avalanche size scale.

Line 321: We replace prediction with segmentation.