Revision Notes, egusphere-2025-3146

Dear Editor and Reviewers,

We would like to express our sincere gratitude for your time and thoughtful comments on our manuscript, " **Rapid Flood Mapping from Aerial Imagery Using Fine-Tuned SAM and ResNet-Backboned U-Net**." Your insightful feedback has been extremely valuable in helping us improve the clarity, strength, and overall quality of our work.

We have carefully considered all suggestions and addressed them point-by-point in the revised manuscript. For your reference, we have highlighted our responses to your comments in green. We believe these revisions have significantly strengthened the manuscript and we are confident that it is now ready for further consideration.

Thank you again for your valuable contribution to this process. We look forward to your feedback on the revised manuscript.

The author's reply to the comments is highlighted in green.

Comments	**Responses**	**Manuscript Change**
The research is well designed and written. It contributes to the development of a strong and user-friendly AI tool that can provide quick and effective support in flood-affected areas where urgent assistance is needed, without requiring harmonized or standardized procedures for image collection from different sources. As a limitation of the research, I believe it would be valuable to suggest including the geolocation of the final flood map to facilitate relief efforts.	Thank you for your careful reading and constructive suggestion. We fully agree that adding geolocation to the final flood maps would substantially increase the usefulness of the system for emergency responders and for insurance loss assessment. We would like to clarify that the dataset used in this study consisted of 290 aerial images and their corresponding manually created masks provided by a third party; these images did not include GPS/INS metadata or any georeferenceable files (e.g., GeoTIFF, orthophotos). Because the original dataset lacks precise location information, it was not possible to produce geolocated outputs in this work. We have now explicitly stated this limitation in the revised	Despite these promising results, there is still room for further research. **A limitation of the present study is that the Flood Area dataset we used (290 aerial images and associated masks) did not include GPS/georeferenced metadata, so it was not possible to produce delocalized results. Addressing this dataset limitation in future works would enable more accurate and actionable relief.**

	manuscript and added a short "future work" plan that describes practical approaches (e.g., collecting GNSS/RTK-enabled UAV imagery, using ground control points and photogrammetric orthorectification, or aligning masks to georeferenced basemaps) to enable georeferenced flood maps in follow-up studies. We appreciate the suggestion and will prioritize geolocation in our future data collection and system development so that the model outputs can be directly used for field operations and addressing location-specific help requests (Please see lines 423-426).	
Furthermore, the reasons behind the superiority of SAM-Points should be discussed. Compared to other methods, this approach appears to be more effective in distinguishing bare soil from flooded areas.	Thank you for raising this important point. We agree that further clarification is necessary. In our study, the superior performance of SAM with point prompts over bounding box prompts can be explained by several dataset-specific characteristics. First, in flood imagery, water often extends across the entire scene with highly irregular and amorphous boundaries. Bounding boxes in such cases tend to cover almost the whole image and thus provide little discriminative information to the model, sometimes even introducing ambiguity between flooded and non-flooded regions. By contrast, multiple dispersed point prompts explicitly highlight localized regions within the	Theoretically, this lower granularity of information from Bbox prompts leads to poorer performance in such cases. **In addition, the inherently diffuse and irregular nature of flood boundaries makes point prompts stronger cues for guiding the model, while bounding boxes typically include both flooded and non-flooded regions, providing the model with less discriminatory guidance.**

	flood extent and along its boundary, which allows SAM to capture fine-grained differences more effectively (as previously mentioned in the manuscript). Second, flood boundaries are less sharply defined compared to other object segmentation tasks, and point prompts serve as stronger anchors for delineating these diffuse regions. Together with our automatic prompt generation strategy (which ensured dispersed placement of points within flooded areas), these factors explain why SAM-Points outperformed SAM-Bbox in this context. We have revised the manuscript to emphasize these aspects more clearly (Please see lines 307-310).	
Upon re-reading the manuscript, I noticed that in lines 200–203 you mention the use of various data augmentation techniques. Could you please clarify the probability settings assigned to each augmentation method?	Thank you for your comment. We have revised the manuscript to clarify the probability settings of the data augmentation methods. The revised text (Lines 201–204) now specifies that random horizontal and vertical flips, rotations (±30°), Gaussian blur, and random grayscale conversion were each applied with a probability of 0.5.	These included geometric transformations such as random horizontal and vertical flips and rotations of **up to 30°** as well as color-based transformations such as random grayscale transformations and Gaussian blurs **with a kernel size of 3**, **all applied with a probability of 0.5.**
In lines 201–203, it is not clear whether the augmentation was applied exclusively to the training dataset. Providing this clarification would enhance the transparency of the methodology.	Thank you for your insightful comment. We confirm that data augmentation was applied exclusively to the training dataset to increase its diversity. This clarification has been added to the revised manuscript (Please see lines 200-201).	Data augmentation was applied **exclusively to the training set** to increase the diversity of the training data.

Still in lines 201–203, it would be highly valuable to explicitly include details regarding the number of images before and after data augmentation, as well as their distribution across the training, validation, and test sets. Such information is critical to ensure reproducibility.	We sincerely thank the reviewer for this valuable comment. We applied data augmentation exclusively to increase data diversity rather than the number of samples. Consequently, the total number of images in each split (training: 204, validation: 43, testing: 43) remained unchanged. This clarification has been incorporated into the revised manuscript (Please see lines 200-201).	Data augmentation was applied exclusively to the training set to increase **the diversity** of the training data.
In lines 209–219, you mention the use of both Dice Loss and Cross-Entropy Loss. Could you please specify how these two loss functions were combined? For example, were they summed, averaged, or weighted differently?	We thank the reviewer for pointing this out. The Dice Loss and Cross-Entropy Loss were combined by taking their average. This clarification has been added to the revised manuscript (Please see lines 210-213).	To minimize the divergence between the predicted and the observed values, we used DiceCELoss, a loss function that integrates Dice Loss with Cross-Entropy Loss (CE Loss). **Specifically, the two components were combined by taking their average,** leveraging both the pixel-wise accuracy (via Cross-Entropy) and the structural similarity (via Dice coefficient) to improve segmentation performance.
I appreciate that the code is publicly available on GitHub. However, I could not locate the corresponding datasets in the repository. Based on the README file, it seems that the authors expect users to obtain the data from an external source. While this is acceptable provided that the source remains reliably available, hosting a copy of the datasets within your GitHub repository would be	We thank the reviewer for this valuable suggestion. We have now added a direct link to the datasets in our GitHub repository to improve accessibility and ensure long-term availability. The README file has been updated accordingly.	

preferable for long-term accessibility.		