Revision Notes, egusphere-2025-3146

Dear Editor, Reviewers, and Community Members,

Thank you very much for your time and exceptionally valuable feedback on our manuscript, "**Rapid Flood Mapping from Aerial Imagery Using Fine-Tuned SAM and ResNet-Backboned U-Net**."

We have carefully addressed all suggestions and thoroughly revised the manuscript based on your thoughtful and insightful comments. Your feedback has significantly helped us enhance the clarity, strength, and overall quality of our work.

Our detailed, point-by-point responses to Reviewer 1, Reviewer 2, and Community comments are highlighted in green, blue, and red, respectively, within the revised document.

We are confident that these revisions have substantially strengthened the manuscript and believe they are now ready for further consideration.

Thank you once again for your contribution to this process. We look forward to your assessment of the revised submission.

The author's reply to the comments of Reviewer 1 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 suggestions. 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 manuscript and added a short "future work" plan that describes practical approaches (e.g.,	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.**

	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 output can be directly used for field operations and addressing location-specific help requests (Please see lines 441-444).	
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 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	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.**

	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 (Please refer to 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	We thank the reviewer for pointing this out. The Dice Loss and Cross-Entropy Loss were combined by taking their	To minimize the divergence between the predicted and the observed values, we used DiceCELoss, a loss function that

loss functions were combined? For example, were they summed, averaged, or weighted differently?	average. This clarification has been added to the revised manuscript (Please see lines 210-213 in blue, since it is common comment with Reviewer2).	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 preferable for long-term accessibility.	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.	

The author's reply to the comments of Reviewer 2 is highlighted in blue.

Comments	Responses	Manuscript Change
In their paper, Hadi Shokati et al. Propose methods to improve rapid flood mapping from Aerial Imagery using Fine-Tuned SAM and ResNet-Backboned U-Net. This paper is a valuable contribution to remote sensing and rapid disaster assessment. Although none of the comments and suggestions are critical, I would like to ask authors to incorporate and address these issues and suggestions before the paper's publication. Although the methods in this paper are related to floods, they do not directly discuss flood itself. Therefore, it would be	We appreciate the reviewer's insightful comment. We replaced 'Flood' with 'Flood Mapping' in Keywords to increase the visibility of the paper. Please see line 27.	**Keywords:** Flood Mapping, ResNet, SAM, UAV, U-Net

beneficial for the readers and also enhance the paper's visibility to replace "flood" in keywords with "flood mapping" or "segmentation of flood", which are more relevant to the presented study.		
The introduction and methods sections are well written, addressing the main issues and research question. However, there was a minor absence of the reference in line 210 regarding the choice of DiceCELoss. It is claimed in the paper that: *"DiceCELoss is often used to improve segmentation performance by leveraging both the pixel-wise accuracy (via Cross-Entropy) and the structural similarity (via Dice coefficient)."* Please include at least one reference to support this claim and the choice of this loss function.	Thank you for your valuable comment. The suggested reference has been added to the manuscript to support the statement regarding the choice of DiceCELoss. Please see lines 210 - 215.	To minimize the discrepancy between observed and predicted flood extents, we used the Dice-Cross-Entropy Loss, which averages the Dice loss and cross-entropy loss. This composite loss function is widely used in model training, as it balances the strengths of both components **(Hadlich et al., 2023; Shokati et al., 2025)**. It facilitates rapid convergence and often improves final performance, particularly enhancing the Dice coefficient **(Hadlich et al., 2023),** which is critical for accurately capturing the spatial overlap between predicted and actual flood areas.
Although the terms and names of the methods are well described throughout the paper, their usage in the text and figures is inconsistent. For example, Segment Anything Mode is consistently abbreviated as SAM, but the versions (point prompts or points prompts) are referred to in various inconsistent forms. Please use consistent terminology for methods in the entire manuscript, especially for the main methods. For instance, here are a few examples: • **SAM (Points prompts)** on page 14 and **SAM (Point prompts)** on page 15 in the figures. • **Points** in Figure 4.	We appreciate the reviewer's valuable comment regarding this inconsistency. In response, we have carefully revised the entire paper to ensure consistency in terminology. Specifically, we unified all variations of the method names: For Bounding box, we consistently used the full form "**bounding box prompts**" throughout the text. For Point prompts, we used the full form "**point prompts**" consistently. For Segment Anything Model, we consistently used its abbreviation "SAM" in the text, however, in the figures, we kept the full form (e.g., "Segment	All the revised and standardized terms have been highlighted in blue throughout the manuscript.

• **point prompts** in line 309 • **point prompt** in line 100 • **"Bounding boxes"** is abbreviated as **"Bbox"** in line 135, but this is inconsistently used throughout the text and figures, sometimes as **"bounding box"** and other times as **"Bbox."** • ...	Anything Model (SAM)") to ensure that readers can interpret the figures independently without referring back to the text. All the revised and standardized terms have been highlighted in blue throughout the manuscript.	
Figure 6 lacks a brief description of the subplots labeled a, b, ..., h in the caption.	Thank you for your insightful comment. Subplots (a–h) correspond to different samples from the dataset we used. However, to make the caption clearer, we added a brief description. Please see lines 401 - 403.	Figure 6: Example segmented images using the Segment Anything Model with point and bounding box prompts (SAM-Points and SAM-Bbox models, respectively) and the U-Net model with ResNet-50 and ResNet-101 backbones. **Subplots (a–h) correspond to different samples from the dataset of Karim et al., (2022).**
I would like to ask the authors to elaborate on why 290 images with different geographic regions and diverse flood events are sufficient for this study. We recognize that transfer learning enables us to train our models with a limited sample size by leveraging pre-trained data; I would appreciate a discussion on how this sample size captures the variability needed for a robust model. Including this clarification would strengthen the manuscript by addressing potential concerns about the dataset.	We thank the reviewer for this comment. In response, we have added a new section to the manuscript. In this section, we explain that in transfer learning, the number of labeled samples required depends on task complexity, model architecture, and the similarity between the pre-trained model and the target task. Our dataset of 290 images, covering flood events in Germany, India, Malaysia, and Bangladesh, provides broad geographic and environmental variability. The inclusion of UAV and helicopter imagery with different angles and altitudes, combined with data augmentation techniques, further increases the effective diversity. Empirical results show that the fine-tuned SAM model achieved	**3.5 Dataset Size and Diversity Considerations** Determining the optimal dataset size in transfer learning does not depend on a fixed number but rather on several factors, including task complexity, model architecture, and the similarity between the pre-trained source domain and the target task. In transfer learning, large-scale pre-trained models such as SAM (Kirillov et al., 2023) and ResNet (He et al., 2016) already capture rich, generalized feature representations from millions of natural images. As a result, a relatively small number of labeled samples is often sufficient for fine-tuning to achieve high performance in specialized applications. Our dataset consists of 290 images

	an IoU of 0.90 and an accuracy of 0.96 on unseen images, confirming that the dataset captures sufficient variability for reliable flood segmentation. Comparable studies (e.g., Ghaznavi et al., 2024; Shokati et al., 2025) also demonstrate strong performance with similar dataset sizes. Please see lines 414 - 430.	covering flood events in countries such as Germany, India, Malaysia, and Bangladesh. This geographic diversity ensures variability in environmental conditions, land cover types, flood characteristics, and illumination. The inclusion of both UAV and helicopter imagery from different camera angles and altitudes further increases this variability, providing a robust basis for model generalization. Additionally, data augmentation techniques (such as rotations, flips, grayscale transformations, and Gaussian blur) increased the effective training diversity and reduced the risk of overfitting. Empirically, our results (Table 1) demonstrate that the fine-tuned SAM model achieved an IoU of 0.90 and an accuracy of 0.96 on unseen data, confirming that the dataset sufficiently captured the variability required for reliable flood segmentation. Comparable studies on environmental and remote sensing tasks (e.g., Ghaznavi et al., 2024; Shokati et al., 2025) have reported strong performance using datasets of similar size, reinforcing the suitability of our sample in the context of transfer learning–based flood segmentation.

The author's reply to the Community Comments is highlighted in blue.

Comment	Response
I'd like to suggest that you consider referencing our results, as they align with your findings and could strengthen your discussion section. For instance, we observed that SAM performs very well in general. However, when segmenting images from the same	Dear Armin, Thank you for your valuable feedback. We truly appreciate the time and effort you have taken to engage in our work.

area, UNet actually produced even better results. This nuance might enrich your discussion, especially when highlighting the practical performance differences between models. Another point is related to the computational aspects: SAM typically operates on 1024×1024 patches, and when fine-tuning with a frozen ViT backbone, it still requires significant computational resources. The choice of ViT backbone also matters—ViT-H is quite heavy and not ideal for fine-tuning, whereas the smaller variants (like Tiny ViT and Medium ViT) tend to perform better with fewer resources. Lastly, regarding datasets: One could argue that with SAM, we might not need large annotated datasets anymore. While SAM reduces the need for manual annotation, I would still say that datasets are necessary. The real question is: how many do we actually need? To help address this, you might consider referencing this recent paper by Professor Anette Eltner from Dresden University: https://doi.org/10.1080/01431161.2025.2457131 It discusses the sensitivity of model performance to dataset size and could help you frame this as a potential advantage of SAM over UNet. I'd love to hear your thoughts and see how you might incorporate some of these ideas into your discussion. Warm regards, Armin	We agree that referencing your results strengthens our discussion significantly. We have now incorporated your findings into the **Results and Discussion** section to provide a more comprehensive analysis in three specific aspects: 1. **Model Comparison:** We aligned our findings that fine-tuned SAM generally outperforms U-Net with your results, while distinguishing our use of point/bounding box prompts versus your automated approach. Please see lines 340 - 346. 2. **Computational Aspects:** We cited your detailed analysis regarding parameter counts and training times to highlight the trade-offs between SAM's precision and U-Net's computational efficiency. Please see lines 327 - 333. 3. **Sky Effects:** We referenced your observations regarding the challenges posed by sky reflections and dynamic water textures to corroborate our findings on how sky elements affect segmentation accuracy. Please see lines 351 - 353. Regarding the ViT backbone, we evaluated multiple variants and observed that their effectiveness for flood-affected area segmentation was comparable. To optimize computational resources, we selected ViT-Base, as it provides a favorable balance between accuracy and efficiency. Please see lines 143 - 148. We also appreciate your recommendation of Professor Anette Eltner's recent article. It provides an excellent perspective on dataset size requirements. The number of labeled images needed depends on the complexity of the task. For instance, in our previous study on erosion and deposition segmentation (https://doi.org/10.1016/j.catena.2025.108954) we worked with about 400 labeled images and observed clear performance gains with increasing dataset size, an effect that was

	particularly relevant given the higher complexity of that task compared to flood mapping. This increased complexity was because eroded and non-eroded soil often have very similar visual characteristics, whereas flooded areas are usually more distinct from their surroundings. We already added a section to discuss dataset size. Please see lines 414-430 (Highlighted in blue, since it was a question for Reviewer 2 as well).