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 blue. 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 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 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.	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
The introduction and methods sections are well written,	Thank you for your valuable comment. The suggested	To minimize the discrepancy between observed and predicted

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.	reference has been added to the manuscript to support the statement regarding the choice of DiceCELoss. Please see lines 210 - 215.	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. • **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	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 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.	All the revised and standardized terms have been highlighted in blue throughout the manuscript.

"bounding box" and other times as **"Bbox."** • ...		
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 396 - 398.	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 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	**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 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

	performance with similar dataset sizes. Please see lines 409 - 425.	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.