Dear Reviewer 1:

We appreciate your careful reading and critical reflections on our manuscript. We value the concerns raised, but respectfully hold a different perspective on several key points. Below, we address each comment in detail, presenting our reasoning and the corresponding revisions to be made to the manuscript.

Comment 1:

The most critical concern lies in the overall suitability and motivation for an approach that learns the absolute stage directly from images. The reliance on on-site gauge data for training at every new location significantly limits its utility, particularly for ungauged catchments, which are the primary target for innovative remote sensing techniques. As gauged catchments already possess well-established, high-accuracy stage measurement methods, the practical added value of this camera-based approach for these sites is questionable. Also, there are already studies discussing the potential and limits of directly learning the stage from images, which are not mentioned in this study (e.g. Vanden Boomen et al., 2021). Furthermore, the risk is high that the approach is highly sensitive to any movements (internal or external geometry) of the camera setup. Such movements would likely necessitate a complete re-learning of the model, which is a significant practical limitation and is not adequately addressed in the current work. Finally, the authors' premise that obtaining accurate stage data is a critical challenge for all DL-based camera gauges is debatable. For approaches relying on photogrammetry, the stage data serves only as a reference, not as the primary input for the AI-model, thereby mitigating this "critical challenge." A stronger, more refined motivation for this specific DL-only approach is needed.

Response to Comment 1

We thank the reviewer for this insightful comment regarding our technical route. In response, we would like to offer complementary perspectives on the intrinsic value of image-based river monitoring.

First, we acknowledge that our approach currently requires physical gauge data at a new site for initial model training. However, the data dependency is limited to the early stage of deployment. For a specific site, during the initial co-existence period of cameras and gauges at the early phase, the latter can serve as reference data to help AI models learn and stabilize. Once the physical gauges reach the end of their lifespan, a well-trained camera-AI system can replace their function at a low cost, enabling continuous observation. With the continuous increase in the number of river camera and physical gauge observation sites, as well as the accumulation of image-water level pairs, DL models that directly establish image-to-water-stage mappings without relying on additional camera parameters or environmental information such as terrain provide a foundation for developing one-fit-all water stage estimation models. In contrast to other methods such as photogrammetric approaches, these models are

independent of site-specific auxiliary data, thereby demonstrating strong generalizability and the potential to be directly extended to ungauged rivers without reliance on physical gauges or in-situ topographic surveys.

Second, we believe that our method possesses a certain degree of robustness to variations in the visual observation environment. As the reviewer pointed out, factors such as camera displacement and geometric variation are important sources of potential uncertainty that can affect the robustness of image-based water level estimation. Taking photogrammetric estimation as an example, this method requires the extraction of water masks and their subsequent overlay with topographic data. When the camera viewpoint changes but the pre-established projection coordinate system is still used, substantial biases in the derived water level can occur. In contrast, the DL framework we adopted, which directly maps images to water stages, can inherently mitigate such effects and maintain robustness. DL models can automatically extract high-level features that capture relative spatial relationships between different objects, such as water bodies and riverbanks, rather than depending on the absolute position of any single target. Consequently, minor camera displacements exert limited influence on the results, while larger shifts can be effectively corrected using modern camera systems equipped with automated calibration and pan–tilt–zoom (PTZ) mechanisms capable of dynamically compensating for geometric variations in real time.

Overall, we consider our chosen technical route to be feasible in terms of generalizability and robustness. Within this framework, the quality of water-stage values obtained from physical gauges—used as training labels for the AI model—is critical. However, influenced by contact-based measurement errors, these labels often contain either *random but minor* deviations or *systematic and substantial* biases. Random noise caused by turbulence or backflow can typically be attenuated during deep-learning parameter optimization. In contrast, systematic errors—such as those arising from sediment accumulation or rapid riverbed changes—necessitate targeted correction strategies. Accordingly, our study's core, innovative contribution is the introduction of a multi-task learning framework that uses the water-pixel proportion from images as an auxiliary label to mitigate systematic errors from physical gauges.

In the revised manuscript, we will refine and integrate these perspectives into the *Introduction* and *Discussion* sections to better articulate the rationale and significance of the proposed deep-learning-enabled, image-based water level estimation approach. Regarding the reviewer's comment on the citation of previous studies, we agree that a more in-depth discussion of their limitations would be beneficial. Therefore, we will incorporate and critically discuss these studies in the revised version to more clearly highlight both the potential and the limitations of prior research in this area. We also recognize that the previous phrasing, "*obtaining accurate stage data is a critical challenge for all DL-based camera gauges*", was inaccurate. This challenge pertains specifically to the technical route used in our study, rather than to all DL-based

camera gauge approaches in general. The statement will be revised in the manuscript to reflect this clarification.

Comment 2:

The paper utilizes pixel information from segmented images to provide relative stage information but lacks sufficient discussion on the segmentation process itself. This is a significant omission, especially since several established studies (e.g., Eltner et al., 2021; Zamboni et al., 2025, Moghimi et al., 2024) already perform this kind of water segmentation for stage measurement, and the potential for segmentation errors and their influence on the multi-task learning is not discussed at all. Furthermore, the review fails to include relevant, state-of-the-art photogrammetric approaches that use water segmentation (e.g., Blanch et al., 2025). Given that the study site appears highly suitable for these methods, a direct comparison and justification for choosing the DL-only approach is necessary. Also, the achieved accuracy, appearing to be in the decimeter (dm) range, is not competitive with the centimeter (cm) accuracy demonstrated by other camera gauge studies, particularly those using robust photogrammetric methods (e.g., Eltner et al., 2021, Erfani et al., 2023, Blanch et al., 2025). Therefore, also the title of the manuscript is misleading because I think, the achieved accuracies cannot be described accurate. Finally, the approach involves combining two loss functions, which necessitates the fine-tuning of the lambda value. This introduces a hyperparameter that must be manually tuned, complicating the model's reliability and generality.

Response to Comment 2:

First, we understand the reviewer's focus on segmentation-based approaches. However, it is important to clarify that the image segmentation module is not the innovation of this study. In our previous work (Wang et al., 2025), we have already proposed and validated an advanced water-body segmentation algorithm, which demonstrated state-of-the-art performance on representative datasets.

The algorithm integrates a domain-specific DL-based water segmentation model with a foundation model, Segment Anything Model (SAM). Compared with the DL-enabled semantic segmentation models used by Erfani et al. (2023) and Blanch et al. (2025), our method exhibits stronger generalizability and lower dependence on local data. In contrast to the SAM-based segmentation approaches proposed by Moghimi et al. (2024) and Zamboni et al. (2025), our method further incorporates an additional DL module as an automated prompter, which enhances usability and facilitates automated deployment. Without any local fine-tuning, the proposed segmentation approach achieved an IoU exceeding 0.9 across four river camera stations in Tewkesbury, UK, demonstrating sufficient performance for generating auxiliary labels within the multi-task learning process of this study.

Overall, the principal contribution of the present work lies in the design of a multi-task learning framework that integrates relative water-level features derived from the segmentation task with the direct water-level regression task. This joint formulation effectively mitigates the impact of biased contact-based gauge data during model training, thereby improving the robustness of water level estimation. A detailed discussion of the segmentation algorithm itself would divert attention from this core methodological innovation — the proposed multi-task learning structure. Therefore, the detailed description of the segmentation method will not appear in the *Introduction*. Instead, following the reviewer's suggestion, we will add appropriate citations and a concise description of the segmentation algorithm in the *Methods* section to justify the use of our internally developed segmentation approach. Moreover, we do not deny the value of photometric approaches that combine water segmentation with topographic projection. In fact, in this study, the ground-truth water stage data were obtained using such a photogrammetric method, as the physical gauge measurements were affected by contact-based biases.

In addition, we would like to clarify a misunderstanding regarding the reported accuracy.The accuracy of our results is expressed in centimeters. This unit convention follows the same precision definition used in previous studies conducted at the same site, ensuring comparability and methodological consistency. Nevertheless, we concur with the reviewer that the use of "*accurate*" in the title may convey a subjective impression. The title will be revised to more clearly highlight the methodological focus and problem addressed, with the effectiveness of the approach objectively demonstrated through the presented results.

Regarding the hyperparameter that balances the two tasks, we have conducted dedicated experiments and ablation analyses to examine their effects. The results and parameter recommendations are reported in the manuscript and can serve as a practical reference for future applications (Line 195 - Line 204):

"*When training the ShuffleNet model under the multi-task learning paradigm, the number of training iterations was set to 20, the learning rate was set to 0.001, and the mini-batch size was set to 8. Additionally, to investigate the impact of the parameter* λ *used to balance the regression and ranking tasks on model training, nine scenarios were set (* λ *=0, 1, 5, 10, 15, 20, 25, 30, 35), and the iterative changes in MSE loss were visualized for each scenario to determine the optimal* λ *value. As illustrated in Figure 5, an increase in* λ *was associated with a corresponding rise in the initial MSE loss value, and the loss 200 did not consistently converge to zero throughout the iterative process. This suggests that multi-task learning effectively mitigates overfitting to the inherently erroneous gauge stage data. Beyond a* λ *value of 10, further increases in* λ *did not result in significant changes in the MSE loss variation. In addition, continuously increasing* λ *may cause the model to prioritize ranking performance over accurate regression, thereby limiting its ability to constrain water stage estimates within reasonable numerical ranges. Consequently, a* λ *value of 10 was selected for subsequent analysis.*"

Comment 3:

The suggested automatic detection of gauge errors appears effective only for very strong and obvious errors. It is unclear why an established statistical approach would not be equally or more effective for this task. The authors apply an automatic post-processing/filtering step to refine the training data, assuming the error resides in the stage data and not the camera imagery. This assumption needs stronger justification. The lack of provided code is a serious concern, particularly for a technical note. This does not comply with the FAIR principles, which are essential for research reproducibility.

Response to Comment 3:

The contact-based biases in physical gauges can generally be divided into random and systematic components. DL models inherently possess a strong capability to smooth out stochastic noise during the optimization process. In our study, an automatic detection module was specifically developed to identify and mitigate significant systematic errors.

Regarding the assumption issue, the strength of the multi-task learning framework lies in its ability to balance and integrate errors across different tasks, thereby enhancing the learning performance of each task simultaneously. It is not necessary to impose a prior assumption about the superiority of one data source over another, and indeed, we have not tried to make such assumption in this study. Furthermore, minor geometric or imagery-related shifts caused by slight camera movement or displacement do not adversely affect the learning performance of DL models, as we have explained in our response to comment 1.

With regard to traditional statistical approaches, these methods typically rely on causal relationships, in which additional meteorological variables — such as precipitation and temperature — are incorporated to correct or explain variations in stage data. In the absence of such external drivers, however, it becomes challenging to distinguish genuine hydrological dynamics from observation errors. In contrast, our approach is grounded in correlative relationships derived directly from camera imagery, providing a visually supported validation pathway that operates independently of meteorological inputs. While future extensions may integrate these meteorological factors, the current framework is designed to address challenges that traditional statistical methods cannot effectively resolve without them. Therefore, the two approaches are conceptually distinct and not directly comparable. The *Discussion* section of the revised manuscript will further elaborate on the theoretical distinctions between causal and correlative modeling paradigms

Finally, we appreciate the reviewer's valuable suggestion regarding code accessibility. We will make the full source code publicly available and include the corresponding

repository link in the revised manuscript to facilitate reproducibility and future research use.

Reference

Blanch, X., Grundmann, J., Hedel, R., & Eltner, A. (2025). AI Image-based method for a robust automatic real-time water level monitoring: A long-term application case. https://doi.org/10.5194/egusphere-2025-724

Erfani, S. M. H., C. Smith, Z. Wu, E. A. Shamsabadi, F. Khatami, A. R. J. Downey, J. Imran, and E. Goharian. 2023. "Eye of Horus: A Vision-Based Framework for Real-Time Water Level Measurement." Hydrology and Earth System Sciences 27 (22): 4135–4149. https://doi.org/10.5194/ hess-27-4135-2023 .

Moghimi, A., M. Welzel, T. Celik, and T. Schlurmann. 2024. "A ComparativePerformance Analysis of Popular Deep Learning Models and Segment Anything Model (SAM) for River Water Segmentation in Close-Range Remote Sensing Imagery." Institute of Electrical and Electronics Engineers Access 12:52067–52085. https://doi.org/10.1109/ACCESS.2024.3385425

Wang, Ze, Heng Lyu, Yanqing Guo, Shun'an Zhou, and Chi Zhang. 2025. How to Use General AI for Task-Specific Applications: A Case Study of Monitoring Water Level Trends with River Cameras. Environmental Modelling & Software 192:106550. https://doi.org/10.1016/j.envsoft.2025.106550.

Zamboni, P. A. P., Blanch, X., Marcato Junior, J., Gonçalves, W. N., & Eltner, A. (2025). Do we need to label large datasets for river water segmentation? Benchmark and stage estimation with minimum to non-labeled image time series. International Journal of Remote Sensing, 46(7), 2719–2747. https://doi.org/10.1080/01431161.2025.2457131