**Reviewer #4**

The manuscript presents a low-cost method with deep learning method to monitor discharge of mountain streams. Excellent performance was achieved with several preprocessing methods and coupled deep learning methods. However, there are still some issues that need further revision.

Thank you for reviewing our manuscript. Your constructive suggestions are crucial for enhancing its quality. We sincerely appreciate your time and effort, and we have carefully considered all of your suggestions. Below are our responses to your questions and our revisions to the manuscript.

1. The author coupled RF and SVM with CNN. However, training methods of these models are different, which means they cannot be trained together. Please provide details about how do you train coupled models.

Re: Thank you for your reminder. We acknowledge that while we did provide individual training details for the two coupled models (CNN+SVM and CNN+RF) in their respective sections (L314-316 for CNN+SVM and L325-326 for CNN+RF), our descriptions were overly ambiguous, which may have caused confusion among readers. Therefore, we have revised and clarified these sections to address their differences as follows:

(1) CNN+SVM (the training procedure is applied to SVM): The extracted image features, coded with a "one-vs-all" scheme, were used to train binary SVM classifiers. Specifically, one SVM classifier with a linear kernel function was trained for each discharge class to distinguish that class from the rest. The hinge loss function was employed to optimize the entire model by maximizing the margin between discharge classes.

(2) CNN+RF (the training procedure is applied to RF): We here used an RF comprising 350 decision trees and five decision leaves for discharge calculation. The coupling method of CNN+RF mirrors that of CNN+SVM, using the same pooling outputs of CNN as inputs for RF discharge classifier. RF is trained to assign optimal weights to each decision tree and leaf without a specific loss function.

2. Accuracy and F1 score are generally used for classification evaluation, while $R^2$ and RMSE are used for regression evaluation. In this study, the author conducted a classification task for discharge. It is unsuitable to use $R^2$ and RMSE for evaluating the results because the predicted discharge is discrete but the ground truth is continuous.

Re: Thank you for your feedback. Accuracy and F1 score are indeed commonly used in classification tasks. Although the discharge is treated as discrete in this study, we also consider the value difference between the simulated discharge and the flowmeter's discharge, especially if model simulations exhibit significant bias from the ground truth. This aspect cannot be adequately captured by accuracy and F1 score alone.

For instance, it was observed that the accuracy of CNN+$RF_{BZ}$ is 4.8% higher than that of $CNN_{CE}$, while the RMSE is also 0.05 m higher. This result suggests that while CNN+$RF_{BZ}$ has a higher likelihood of recognizing the true discharge, it also has a greater chance of identifying incorrect discharges with significant bias. Therefore, evaluating the model's performance solely based on accuracy and F1 score is incomplete. Incorporating regression metrics like $R^2$ and RMSE is appropriate to reflect the model's robustness in handling extreme outliers, which is equally crucial for discharge monitoring.

3. Lines 354-358, how did the author get the $-\bar{R}+7.5\bar{G}-6.5\bar{B}$ ? Is there any theory of this correlation? Please provide details about how to get the equation in the "Methods" section

Re: Thank you for your feedback regarding Section 3.1. Our aim in this section was to explore the relationship between image R/G/B characteristics and discharge values. Initially, we established the equation $a\bar{R}+b\bar{G}+c\bar{B}$, where coefficients *a*, *b*, and *c* were to be determined. Through a systematic exploration of various combinations of *a*, *b*, and *c*, we identified that the characteristics derived from $-\bar{R}+7.5\bar{G}-6.5\bar{B}$ exhibit a high correlation with discharge values. This finding supports our assertion that discharge can be inferred directly from RGB matrices without the need for preliminary extraction of cross-sections and flow velocity, which forms the theoretical foundation of our study.

We acknowledge that this section serves as an introduction to subsequent sections. Our forthcoming work will leverage more sophisticated deep learning models to enhance the retrieval of discharge-related features under dynamic environmental conditions, ensuring robustness and stability.

In response to your suggestions for improved readability, we have incorporated additional details in Section 2.4, explaining how to get the equation.

4. The section "Correlation analysis" is not related to the theme of this study and not included in the flowchart. Why did the author build such a linear relationship in the study focused on deep

learning application?

Re: As previously explained, this section is essential as it serves as a preliminary step to demonstrate the feasibility of deriving flow discharge directly from RGB matrices without first extracting cross-sections and flow velocity. This foundational work establishes that discharge-related features are indeed embedded in images, thereby justifying and facilitating the subsequent use of deep learning models to extract flow discharge.

To aid in reader comprehension and avoid confusion, we have added the aim of this section to the manuscript.

5. Lines 460-463, is preprocessing time included in the comparison?

Re: The preprocessing time is not included in the comparison. The mentioned time refers to the duration taken by the three models to calculate discharge with preprocessed images as input. It is worth noting that preprocessing is significantly faster than discharge computation (approximately 10 times faster), which is why it was not considered in the overall timing.

To enhance readability and clarity, we have revised the description in the manuscript.

6. One of the application limitations of the study is that it could not be applied without labeled discharge from a gauge. It would be better to discuss more about the limitation and further improvement in the "Discussion" section.

Re: Thank you for your very constructive suggestion. We have added a further discussion on the limitations of our method and outlined future directions in Section 4, as follows:

"Moreover, site-specific field data is crucial for identifying the criteria for image categorization and model training, which restricts the broader applicability of our approach in ungauged basins, where such field data may not be readily available. Further research on integrating multiple data sources and surveying approaches is warranted for developing a more generalizable method."

**Reviewer #3**

(1) “acoustic doppler current profiler”, the word doppler should be capitalized.

Re: Thank you. We have corrected it in the revised manuscript.

(2) In Introduction, the authors have focused on explaining image methods: PIV, PTV, STIV. Have deep learning techniques ever been used in hydrological monitoring? Previous studies on this topic should be discussed in this section.

Re: Thank you for your suggestion. In the Introduction, we have indeed focused on widely used image-based methods such as PIV, PTV, and STIV, which often lack applicability for mountain streams. Our study is the first attempt to retrieve flow discharge directly from RGB images without prior knowledge of river geometry and cross-sections using deep learning models, thereby addressing the challenges faced by these popular methods. Hence, we emphasized the comparison with these image-based methods.

We acknowledge that previous studies on deep learning techniques in hydrological monitoring are rare. However, we have discussed their promising application potential in Lines 83-90. Additionally, we identified a relevant study titled “RivQNet: Deep Learning Based River Discharge Estimation Using Close-Range Water Surface Imagery” (Ansari et al., 2023), which introduces the application of deep learning methods in river monitoring. This study demonstrates the representativeness and advantages of deep learning models, although the discharge estimation still relies on cross-section information and derives surface velocity using CNN.

We have incorporated a brief discussion of this literature in the Introduction to highlight the advantages of deep learning models.

(3) L61, lacking should be replaced by lack.

Re: Thank you. We have corrected it in the revised manuscript.

(4) 3.3.2 Comparison of discharge models. What is “discharge model”? It needs to be clarified because you have used “discharge classification model” through the paper.

Re: Thank you. We have standardized its terminology throughout the text to “discharge classification model(s)”.

**Reference:**

Ansari, S., Rennie, C., Jamieson, E., Seidou, O., and Clark, S.: RivQNet: Deep Learning Based River Discharge Estimation Using Close‐Range Water Surface Imagery, Water Resources Research, 59, 10.1029/2021WR031841, 2023.

**Reviewer #2**

1. Lines 98-100. The first hypothesis “the features of mountain streams (e.g., coverage of water surface, flow direction, flow velocity) embedded in RGB images can be recognized...” is better to be responded or discussed in the discussion or conclusion section.

Re: Thanks for your constructive suggestions. We have addressed the two hypotheses in the "Conclusion" section to better summarize our work and facilitate reader understanding.

2. Section 2.2. I suggest a hydrograph for the period July 20th to September 27th 2022 with high quality image data marked.

Re: Thank you for your constructive suggestions. To improve the presentation, we have added a hydrograph demonstrating the in-situ discharge with excellent quality to Figure 1, as follows:

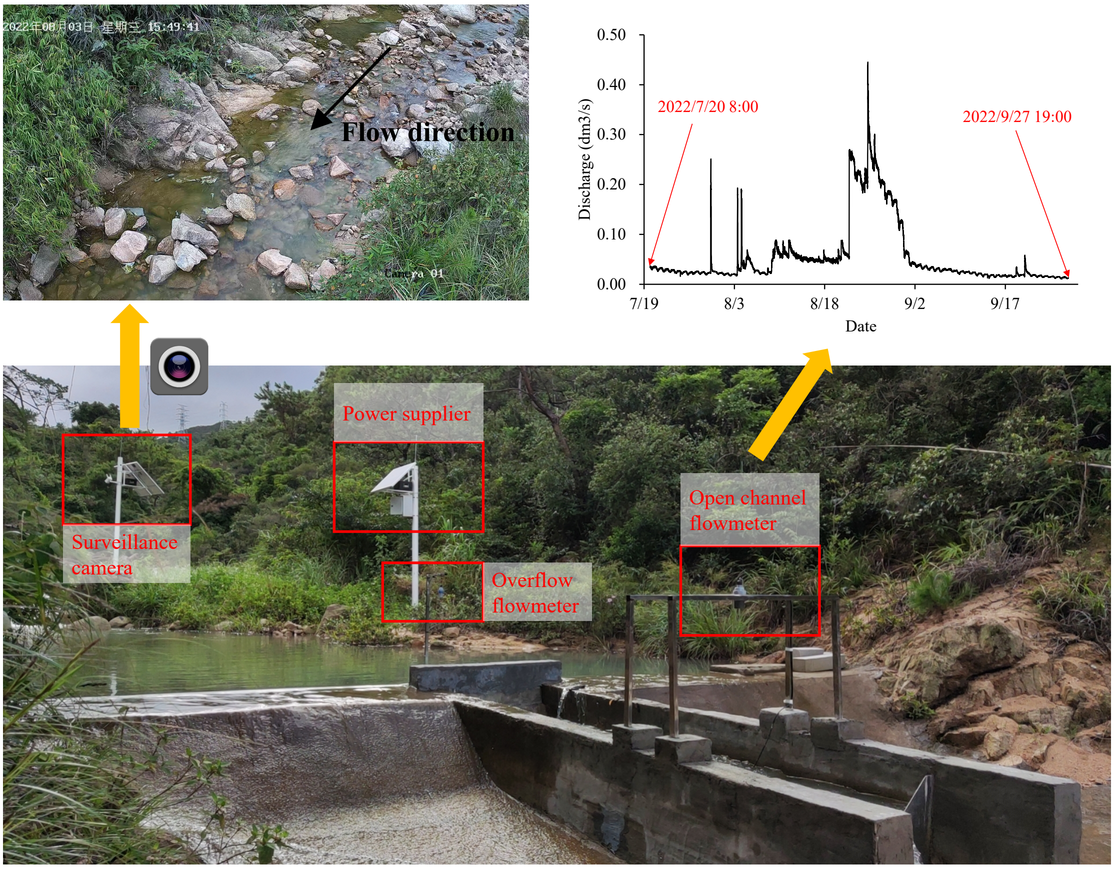


Figure 1. Camera setup. The camera is set on the left bank of the stream, about 3 m above the water surface, and 8 m upstream of a gauging weir. The top right panel demonstrates the changes in the flowmeter’s discharge during the measurement period.

3. Line 333. “Accuracy” should be “Classification accuracy” and in other sections.

Re: Thank you very much. We have standardized the terminology to “Classification accuracy” when referring to the performance of our discharge classification models. This includes updating the labels in Figures 6 and 8.

4. The discussion section is about the advantage, limitations, and the role of key procedures of the new method. It is better to reorganize this section to make it clearer.

Re: Thank you for your constructive suggestions. We have reorganized the Discussion section to make it clearer by presenting the advantages, limitations, and potential directions for future improvement in that order. Additionally, we have added some content highlighting the primary limitation and suggesting possible improvements.

5. Lines 539-541. The general performance of the method evaluated by R2 and RMSE is needed.

Re: Thanks for your reminder. We have included the comparison of $R^2$ and RMSE as follows,

"In this case, the classification accuracy, F1 score, and R2 of CNN+SVM and CNN+RF were 9.1%~14.4%, 0.084~0.115, and 0.006~0.010 higher, respectively, while RMSE was 0.31~0.51 m lower compared to CNN."

Additionally, we have supplemented the demonstration of these metrics in Sections 3.3.2 and 3.3.3 to facilitate better comparison and provide a clearer highlight of the best models and color-enhancing methods.

**Reviewer #1**

The authors have not yet answered the question of how the proposed method can be applied to practical flow measurements. Moreover, 37 traffic samples are too small, so the sample data in this paper are not representative.

During floods, the light is dark and storms often occur. Therefore, the traffic monitoring method established only from the color classification may be feasible in a certain condition, but it is difficult to apply in more scenarios. The paper lacks sufficient traffic monitoring data to verify the rationality of the results under different environmental conditions, so it is recommended to be rejected.

Re: Thank you for your comment. Discharge monitoring at rocky upstream mountain streams has been a difficult task for a long time due to the complex topography. We agree that the samples presented in this study did not cover all environmental conditions, which affects the applicability and transferability of the models. We have thoroughly discussed about these limitations in the manuscript. However, we believe that our study represents a new direction for applying deep learning techniques in acquiring high-frequency discharge data through image analysis. Although the method was only tested at one single site, it provides a different idea that could serve as an alternative apparatus, or integrated into traditional monitoring approaches to improve data quality. The paper also tackled a few important issues in streamflow image processing, including the treatment of images affected by the disturbances of water reflection and vegetation shadow, and the tradeoff between speed and accuracy when using different color enhancing methods. These attempts could provide useful reference for streamflow observation at other sites facing similar challenges.