We are grateful to the reviewer's insightful and constructive comments. Please see the following point-to-point response.

（1）The discharge is usually gauged at a regular and stable channel. The channel the camera set is irregular and there are lots of rocks scattered in the channel. It makes the flow turbulent and unstable. I don't think it is a suitable location for monitoring discharge. The method of retrieving image features for discharge monitoring could distinguish the magnitude of discharge qualitatively, but it seems impossible to conduct quantitatively without labeled discharge from a gauge. It means the method is hard to be applied independently, especially at an irregular channel without stable stage-discharge relationship.

Re: We agree that it is challenging to monitor mountain discharge at upstream irregular channels due to the steep terrain and inaccessibility for field personnel. As the reviewer pointed out, most of the streamflow gauges are still set at downstream river channels nowadays. Therefore, we attempted to propose a method of monitoring the discharge of mountain streams by retrieving image features with deep learning models. Particularly, we tested two hypotheses: (1) the features of mountain streams (e.g., coverage of water surface, flow direction, flow velocity) embedded in RGB images can be recognized by suitable deep learning approaches to achieve effective discharge monitoring, and (2) proper image pre-processing and categorization can improve accuracy of image-based discharge monitoring of mountain streams. The proposed method provides a possible alternative apparatus for continuous discharge monitoring at rocky upstream mountain streams, where it is challenging to identify the cross-section shape or establish a stable stage-discharge relationship. Site-specific field data is still needed to identify the criteria for image categorization and model validation. The reviewer has pointed out an important direction for future research. Further efforts are needed to develop a method that can be transferred directly to other sites without model calibration. However, we believe that we have achieved our goal in this study despite the method's limitations.

(2) With the proposed method, it may be possible to generate more accurate estimation of discharge compared to large-scale particle image velocimetry (LSPIV) and particle tracking velocimetry (PTV) at a regular channel. I suggest a comparison among the these methods to make it more convincing.

Re: LSPIV and PTV have been widely used to acquire the flow velocity distribution on water surface in recent years. We indeed have tried these two methods when we started this study. Unfortunately, we found LSPIV and PTV unsuitable for monitoring rocky mountain streams due to the steep terrain, shallow water depth, and complex cross-section. As shown in the figure below (**Fig. 1**), the velocity distribution of the stream derived from PIV (Tauro et al., 2018; Thielicke and Sonntag, 2021) is irregular and chaotic, making it hard to calculate the average flow velocity at a certain cross-section, and thus discharge.

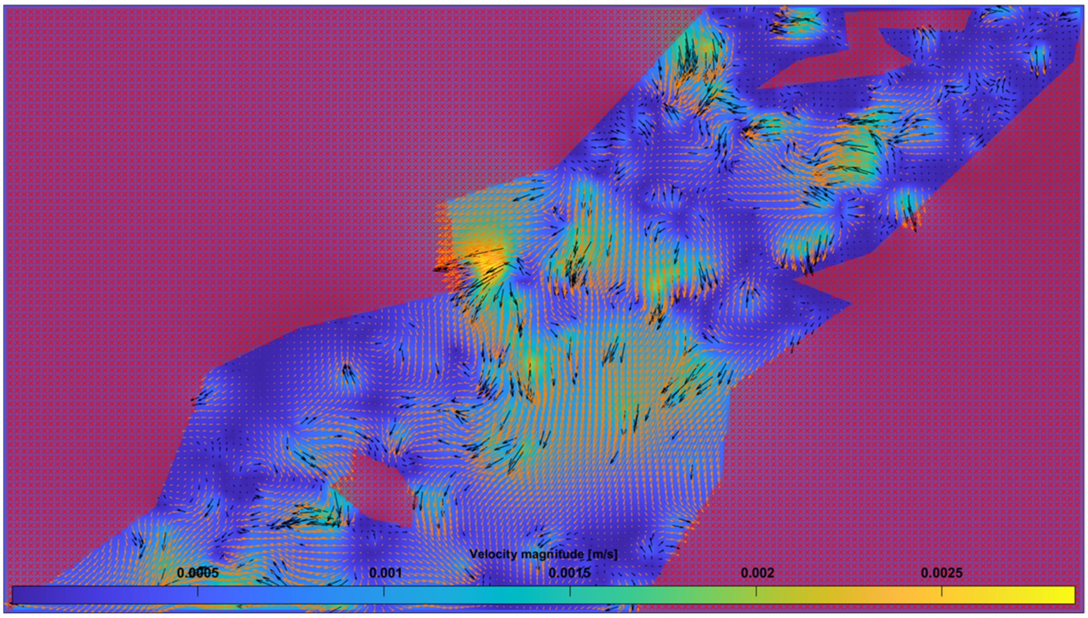

Figure 1. The velocity distribution at the study site at 14:00 pm 2022/9/10 calculated by PIV. The regions filled with red crosses represent streambank and the large rocks lying in the middle of stream. The arrows represent the flow directions, and the color represents the flow velocity.

(3) It is better to give a brief description of the advances of retrieving image features with deep learning in the section of introduction.

Re: Following the reviewer's suggestion, we have added a paragraph in Introduction to describe the advances of retrieving image features with deep learning, as:

"Unlike PIV and PTV, deep learning models possess the capability to extract discharge-related features from images of rivers or streams automatically. These models are able to adjust the weights assigned to each feature, eliminating the need for manual attention and reducing the risk of overemphasizing or misinterpreting features that are unresponsive to flow discharge (Canziani et al., 2016). Besides, deep learning models can extract low-level image features, such as edges, textures, and colors (Jiang et al., 2021). These merits could be essential in retrieving information from images of mountain streams, particularly in regions with intricate cross-sectional profiles.”

(4) Lines 262-268. It is better to put these sentences in the section of introduction or methods.

Re: Following the reviewer’s suggestion, we have added a new subsection (**Section 2.4**) to introduce the purpose and method of analyzing correlation between color information and discharge.

(5) Figure 4. The differences among the sub-figures, like b-1 and b-2 should be pointed out.

Re: We have added a detailed description in the caption of **Fig. 4**:

“Figures b-1, c-1, d-1, and e-1 display the saturation and brightness distributions in Area 1-4 of a “Good quality” sample. Figures b-2, c-2, d-2, and e-2 display the results derived from samples of “Below shadow” (b-2; c-2), “Middle shadow” (d-2), and “Water reflection” (e-2), respectively. Figures f-1, f-2, and f-3 display the saturation and brightness distributions of an entire image, derived from “Good quality”, “Dark”, and “Raindrops” samples, respectively.”

(6) Lines 282-299. There are several thresholds in the section, such as values of 0.3 and 0.4 in line 295, 0.98 and 1300 in lines 297. How these values were determined.

Re: The thresholds for automated image categorization were determined manually by comparing image samples under different environmental conditions. Similar patterns were found in the saturation and brightness distributions of different categories of images. For example, the "Middle shadow" images showed the brightness values with the largest number of pixels in Area 3 were less than 0.4, while that of the other categories of image were higher than 0.4, so we chose 0.4 as the threshold to classify "Middle shadow" images in test 3. We have clarified this in the revised manuscript.

References:

Canziani, A., Paszke, A., and Culurciello, E.: An Analysis of Deep Neural Network Models for

Practical Applications, ArXiv, abs/1605.07678, 7-14, 10.48550/arXiv.1605.07678, 2016.

Jiang, P. T., Zhang, C. B., Hou, Q., Cheng, M. M., and Wei, Y.: LayerCAM: Exploring Hierarchical Class Activation Maps for Localization, IEEE Transactions on Image Processing, 30, 5875-5888, 10.1109/TIP.2021.3089943, 2021.

Tauro, F., Petroselli, A., and Grimaldi, S.: Optical sensing for stream flow observations: A review, Journal of Agricultural Engineering, 49, 199-206, 10.4081/jae.2018.836, 2018.

Thielicke, W. and Sonntag, R.: Particle Image Velocimetry for MATLAB: Accuracy and enhanced algorithms in PIVlab, Journal of Open Research Software, 9, 12-26, 10.5334/jors.334, 2021.