The reviewer comments are in black, and our answers are in red.

The proposed paper presents a data set containing a rock glacier inventory (RoGI) in the Tibetan Plateau. The vastness of the study area combined with the number of rock glaciers that have been inventoried are the main, bot not the only, arguments that make this data set to be of significant importance for the scientific community.

We thank the reviewer for recognizing the scientific significance of our inventory.

The methodology used to compile the inventory complies with good practices, recommendations and the latest guidelines in rock glacier inventorying as proposed by the Rock Glacier Inventories and Kinematics community (RGIK).

Although, it doesn't use a specific kinematic method (e.g. remote sensing, GNSS), the geomorphological approach in estimating the activity status of RGs is an accepted method by the scientific community.

Estimating the activity status of rock glaciers is our future effort. We will use InSAR data to attribute the kinematic information. See lines 512-513: "The InSAR data will be used to attribute kinematic information (RGIK, 2022)."

The performance of the deep learning algorithm varies significantly between subregions, performing poorly in at least five of them, but the manual revise of the deep learning mapping solves this shortcoming.

Yes, manual improvement is crucial for producing this inventory.

Thus, the resulting data set is rigorous enough to be considered for publication. As for the manuscript itself, I recommend some minor revisions and clarifications.

Specific comments:

Line 238 – 240

Please explain in more detail how the "Retrieve" operation in the "Manual improvement and independent validation" was performed. Specifically in what areas were the rock glaciers added, as I assume you didn't check the entire study area.

The reviewer is right. We didn't check the entire study area but focused on areas in proximity to the polygons identified by the deep learning model. Specifically, when we observed a missing rock glacier within the window scope during the manual inspection of candidate rock glacier polygons, we added it to our inventory.

Relevant information has been added to the manuscript in lines 240-241: "When missing rock glaciers were identified during the manual inspection of nearby candidate polygons, they were added to the inventory.". and lines 446-448: "Additionally, the "Retrieve" operation focused on areas where missing rock glaciers were observed near the polygons identified by the deep learning model. Consequently, some rock glaciers may have been missed without conducting an exhaustive examination of the entire study region.".

Line 251 – 258

It is not clear if the two reviewers have made changes to the RoGI or if their only purpose was to evaluate the accuracy of the inventory. In other words, a rock glacier that was drawn by the seven mappers and the which was found to be incorrectly identified by the reviewers is still part of the inventory? Please clarify.

The two reviewers have not made changes to the RoGI. Their only purpose was to evaluate the accuracy of the inventory. We have included an attribute "ADDI_INF" in the inventory dataset to provide this information. See lines 267-268: "The ADDI_INF provides information on whether the rock glacier has been recognized as a false identification by the reviewers.".

Paragraph 6.1.2 Limitations of the deep learning model (lines 408 – 427)

It is important to acknowledge the limits of the model as they are the based on which the model can be improve in the future. And the most important model limitation, in my opinion, is the number of bands that can be used as input data. The use of morphometric data (e.g. slope, terrain roughness) and lithological data might significantly increase the model accuracy. Please consider acknowledging this limitation or have an argument on why this is not a significant limitation.

Thanks for pointing it out. Yes, it is an important limitation of our model and hasn't been discussed in the current manuscript. We have added this in our revised manuscript. See lines 417-420: "A key limitation of the current deep learning model is the restricted number of input bands. Our model only utilizes RGB bands, while inherently excluding crucial topographic information such as slope and elevation. As rock glacier occurrence is closely related to topography and underlying geology, the

absence of morphometric inputs like terrain roughness and slope, as well as lithological data, may hinder the model performance.”.

Table 3 (page 15) The first two data columns are expressed in the number of rock glaciers, while the following three are expressed as the total areas of rock glacier. If there are rock glaciers that greatly vary in size, then some metrics (e.g. F1 score) might be influenced by it. Also, tables 4 and 5 are expressed in number of rock glaciers. Please explain why you choose to have this inconsistency or consider revising the figure in order that all the data columns are expressed in the same unit. Consistency is important for accurate assessment and for easy reading.

Thanks for the comment. The three tables serve different purposes, thus may have some inconsistencies. Table 3 aims to evaluate the performance of our deep learning model. Here, we chose to evaluate the area instead of counting numbers, because the area-based evaluation is more relevant to the pixel-based evaluation of a deep learning model.

However, for Tables 4 and 5, the evaluation and comparison are conducted on the inventory rather than deep learning model. For evaluating an inventory, as we pointed out in lines 252-253: “Given the difficulty in accurately evaluating the delineated boundaries, our validation focused primarily on verifying the primary markers.”. Thus, it is more practical to compare the numbers instead of the areas for these two tables.