The Response to Comments from Review 2

Comment 1

Table 1, please provide the number of sites where landslides have occurred, along with the number of landslide polygons for each dataset.

Response 1

Thanks for your careful suggestion.

We have revised Table 1 to include the specific events corresponding to each dataset and the number of landslides associated with each event. However, due to the absence of detailed information in the original sources for the GVLM and CAS datasets, some data remain unavailable, resulting in incomplete information. **(P7L161)**

Table.1 *Existing landslide dataset statistics*

Dataset	Bands	events	Tiles	Landslides number	Labeling pixels
Bijie landslide	3	1	2773	770	7.23×10^6
Landslide4sense	14	4	3799	>30000	1.76×10^6
HR-GLDD	4	13	1756	7193	2.96×10^6
GVLM	3	17	17	-	3.24×10^7
CAS Landslide	3	9	20865	-	1.95×10^8

Comment 2

Table 2, please specify the total number of polygons obtained and confirms that the necessary rights for the use of the mentioned images.

Response 2

Thanks for your valuable comment.

1. We have made some revisions to the content of Section 3.1, "Data Collection." In response to your suggestions, we have added Table 2 to provide additional information. This table includes details such as the time, geographic coordinates (latitude and longitude), the number of landslides, and the total area affected by each landslide event. **(P9L179)**

Table.2 Summary table of landslide event information in GDCLD

Events	*Mw*	*time*	*Geographic coordinates*	*Landslide number*	*Total landslide area (km^2)*
Jiuzhaigou	*6.5*	*2017*	*(102.82°E, 33.20°N)*	*2498*	*14.5*
Mainling	*6.4*	*2017*	*(95.02°E, 29.75°N)*	*1448*	*33.6*
Hokkaido	*6.6*	*2018*	*(142.01°E, 42.69°N)*	*7962*	*23.8*
Palu	*7.5*	*2018*	*(119.84°E, 0.18°S)*	*15700*	*43.0*
Mesetas	*6.0*	*2019*	*(76.19°W, 3.45°N)*	*804*	*8.5*
Nippes	*7.2*	*2021*	*(73.45°W, 18.35°N)*	*4893*	*45.6*
Sumatra	*6.1*	*2022*	*(100.10°E, 0.22°N)*	*602*	*10.6*
Lushan	*5.9*	*2022*	*(102.94°E, 30.37°N)*	*1063*	*7.2*
Luding	*6.8*	*2022*	*(102.08°E, 29.59°N)*	*15163*	*28.53*

2. Details regarding data authorization are provided in Section 8, "Data Availability." The Planet data were obtained through the Planet Education and Research Program. Both the Map World and GF-6 datasets were accessed under an image license acquired by our team. The UAV data are under the usage rights of the laboratory affiliated with our team.

"Both the Map World and GF-6 datasets were accessed under an image license acquired by our team. The UAV data are under the usage rights of the laboratory affiliated with our team. If you need to use them, please contact the corresponding author. The original PlanetScope data were obtained through the Planet Education and Research Program. You can get original imageries at https://www.planet.com/ (Planet Team, 2019)***."*** **(P41L713~717)**

Comment 3

In Fig. 4, it's crucial to clarify the distinction between 'Label' and 'Ground Truth,' as they may initially appear similar.

Response 3

Thank you for giving this comment.

In Figure.4, the "label" represents binary pixel value derived from manually interpreted landslide polygons, while the "ground truth" is depicted by overlaying the semi-transparent landslide label on the corresponding position of the image. This approach visually *demonstrates the accuracy of our landslide annotations. In the caption of Figure.4, we added a sentence to explain these words.*

*"The "label" refers to the binary landslide mask, whereas the "Ground Truth" illustrates the concordance between the annotated and actual landslide in images." (**P19L351~353**)*

Comment 4

A clear workflow outlining the entire dataset production process, along with details on personnel involvement, costs, and time invested, would offer valuable insights into the significant effort required to create such a comprehensive resource.

Response 4

Thanks for your insightful advices.

1. We have drawn a flowchart of the dataset preprocessing and added it **Section 3 (Figure.1)**. **(P8L164~171)**

"The creation of the GDCLD dataset can be broadly divided into two main components: landslide data collection and remote sensing data processing. In the first part, we compiled recent landslide events induced by earthquakes worldwide over the past seven years and obtained the corresponding remote sensing imagery. The second part details the process of annotating landslide labels and the methodology used to create the standard dataset. The workflow is illustrated in Figure.1.

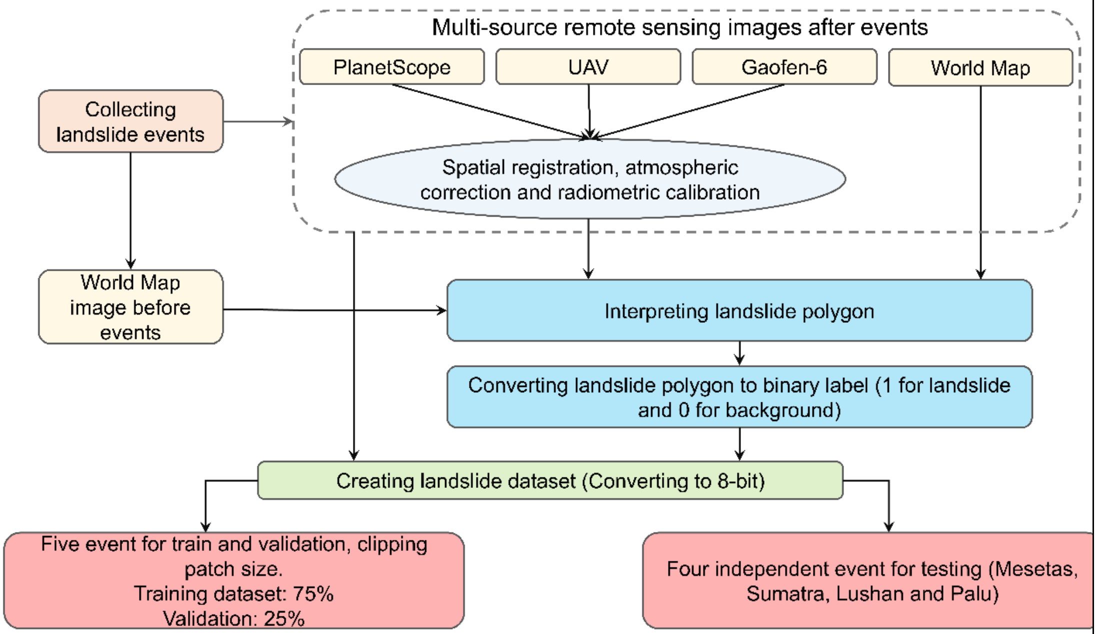

Figure.1 The workflow of producing GDCLD"

2. Regarding the specific timeline and procedures for dataset creation, the landslide data included in the GDCLD were interpreted by our team over one year of research.

*"*The mapping of landslide polygons for these nine events was primarily conducted by a team of five researchers, including the authors. All team members possess expertise in geology or remote sensing and were involved in a year-long process of detailed interpretation.*"* **(P13~14L298~301)**

Moreover, we have acknowledged the efforts of all colleagues involved in the landslide interpretation in the **Acknowledgements** *section with the following statement: "We sincerely thank all colleagues who contributed to the landslide interpretation work."* **(P43L768~769)**

Comment 5

Lastly, the section titled '6.3 Model based on GDCLD performance on existing datasets' necessitates clarification to ensure its content is fully understood.

Response 5

We thank the reviewer for raising the question.

During the revision of our manuscript, we have made adjustment to the content of **Section 6.3** *and also revised its title. The overall content of* **Section 6.3** *is outlined as follows:* **(P33~36L601~650)**

"6.3 Comparison with existing landslide datasets and models

To assess the robustness and generalization capabilities of the GDCLD dataset, we employ SegFormer trained on the GDCLD dataset (GDCLD-S model) to identify landslides within three distinct datasets: HR-GLDD, GVLM, and CAS. Initially, we standardize the data from these three datasets into 1024×1024 remote sensing tiles. Subsequently, utilizing the MGDCLD-S model, we conduct landslide identification across all these datasets. Table.8 9 demonstrates favorable performance of the model across these diverse datasets. For instance, in the HR-GLDD dataset, which shares similarities with the PlanetScope image within GDCLD, the model achieves an mIoU of 76.97%, indicating a balance between Precision and Recall metrics. Similarly, when applied to the GVLM dataset, leveraging Map World image, our dataset exhibits robust predictive outcomes, resulting in a comprehensive mIoU of 70.07%. Likewise, for the CAS dataset, GDCLD demonstrates strong generalization capabilities, yielding an outstanding comprehensive metric with mIoU = 76.91%, alongside balanced Recall and Precision metrics.

Although all landslide samples contained in GDCLD are induced by seismic activity, our model demonstrates good detection capabilities for rainfall-induced landslides. These two categories exhibit distinct spectral characteristics from their surrounding environments. Consequently, models trained on GDCLD exhibit proficient detection capabilities for rainfall-induced landslides. We present the identification performance of GDCLD-based model for rainfall-induced landslides from the GVLM dataset in Table.8 9 and Figure.1213. Figure.12 13 underscores the excellent detection performance of the GDCLD-S model on rainfall-induced landslides in the GVLM dataset. Despite occasional misclassifications of small-size targets, the model effectively delineates the majority of rain-induced landslides. the Precision metrics in Table.8 affirm this observation with an mIoU reaching 78.22% and both Recall and Precision exceeding 85%. This highlights the robust generalization capability of the model trained by our dataset, enabling effective identification of rainfall-induced landslides

Table.9 Validation results of other public datasets

Dataset	Precision (%)	Recall (%)	F1 (%)	mIoU (%)
HR-GLDD	84.88	86.81	85.84	76.97
GVLM	72.83	87.54	80.68	70.07
CAS	82.95	86.35	84.62	76.91
GVLM-rainfall	85.88	86.71	86.29	78.22

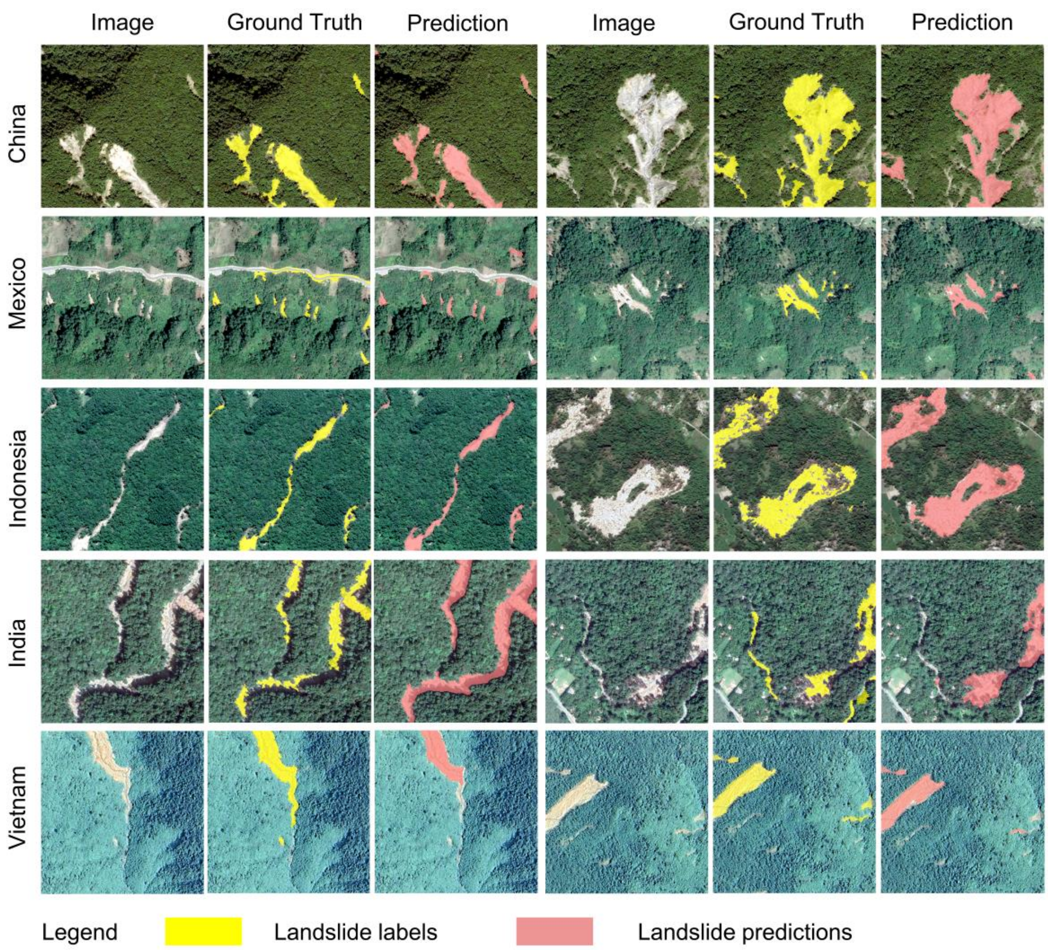

Figure.13 Detection results of rainfall landslides by GDCLD-S model in GVLM dataset

In addition to the aforementioned analyses, we compare the performance of GDCLD with other two datasets, GVLM and CAS. Specifically, we train landslide detection models using the SegFormer algorithm on the GVLM and CAS datasets, denoted as GVLM-S and CAS-S, respectively, with identical training parameters as previously described. Furthermore, we also use the DeepLabV3 to train the CAS-D model based on the CAS dataset and use it for comparison of landslide detection (Xu et al., 2024). Subsequently, the GDCLD-S, CAS-S, CAS-D and GVLM-S models were applied to identify landslides in the Lushan area using three distinct remote sensing data sources: UAV, PlanetScope, and Map World. The results of this comparison are presented in Table 10. From Table 10, it is evident that the GDCLD-S model outperformed CAS-S, CAS-D and GVLM-S across all three remote sensing datasets, achieving mIoU of 72.96%, 69.05%, and 71.92% on UAV, PlanetScope, and Map World. In contrast, CAS-S records mIoU values of 62.03%, 56.86%, and 60.35%

for the same datasets, respectively, which is better than the CAS-D model trained with DeepLabV3, and also illustrates the advantages of the transformer architecture over the CNN architecture. Notably, GDCLD-S exhibited a significantly higher Recall than the other two models and also demonstrated an advantage in Precision. Overall, GDCLD-S, along with CAS-S, exhibited superior performance compared to the single-source data model GVLM-S, particularly in handling multisource remote sensing images. The extensive landslide data and negative samples included in GDCLD-S further contributed to its enhanced robustness against noise and improved Recall in landslide detection.

Table.10 *Performance of the GDCLD-S, GVLM-S, CAS-S, and CAS-D models on the Lushan case*

Model	*Data type*	*Precision (%)*	*Recall (%)*	*F1 (%)*	*mIoU (%)*
CAS-D	UAV	72.73	55.34	62.88	57.91
	PlanetScope	52.07	56.05	53.93	52.86
	Map World	61.79	70.50	64.9	58.11
GVLM-S	UAV	73.03	54.84	57.67	53.41
	PlanetScope	60.13	53.40	54.82	51.52
	Map World	**77.71**	66.40	71.56	63.97
CAS-S	UAV	74.08	67.05	69.95	62.03
	PlanetScope	58.56	76.57	66.40	56.86
	Map World	75.02	64.65	68.37	60.35
GDCLD-S	UAV	**74.72**	**90.35**	**81.80**	**72.96**
	PlanetScope	**81.50**	**82.28**	**81.78**	**69.05**
	Map World	76.18	**87.35**	**81.38**	**71.92**

"

Reference

Xu, Y., Ouyang, C., Xu, Q., Wang, D., Zhao, B., and Luo, Y.: CAS Landslide Dataset: A Large-Scale and Multisensor Dataset for Deep Learning-Based Landslide Detection, Sci Data, 11, 12, 10.1038/s41597-023-02847-z, 2024.