REGIONAL TRANSFERABILITY OF DEEP LEARNING MODELS FOR LANDSLIDE DETECTION WITH SAR DATA

Ioannis Prapas[1,2]*, Wei Ji Leong*[4]*, Ragini Bal Mahesh*[6]*, Vanessa Boehm*[8]*,*
Ioannis Papoutsis[1]*, Gustau Camps-Valls*[2]*, Siddha Ganju*[9]*,*
Edoardo Nemni[3]*, Freddie Kalaitzis*[7]*, Raul Ramos-Pollan*[5]

[1]National Observatory of Athens, Greece
[2]Image and Signal Processing Lab, Universitat de València, Spain
[3]United Nations Satellite Centre (UNOSAT), Switzerland
[4]The Ohio State University, United States
[5]University of Antioquia, Colombia
[6]German Aerospace Center (DLR)
[7]University of Oxford, United Kingdom
[8]University of California Berkeley, United States
[9]NVIDIA, United States

Rising temperatures and climate change are projected to worsen landslides [1, 2] with the increase of sustained droughts and intense rainfall events [3–5]. Given these predictions, there is a growing need for timely and accurate landslide assessment methods that can inform policymakers and emergency responders. A Synthetic Aperture Radar (SAR), as an active imaging sensor independent of cloud coverage, time of the day, and weather conditions, can be used to map the landslides in the aftermath of a disaster. It has been demonstrated that a simple approach like thresholding can work quite well with polarimetry data. Still, it often requires a lot of manual tuning and long time series of SAR data that are unavailable when the landslide is triggered [6, 7]. Deep Learning (DL) methods are widely used in the geosciences in general, [8, 9], and have been proposed to make sense of the SAR data in a rapid response setting, segmenting the landslides with high skill [10–12]. However, the accuracy of such models is often evaluated in isolation in individual regions, which may overestimate the performance of the models in a real-world setting. Here, we show that conclusions drawn in isolation and the skill of such models do not necessarily transfer between regions. A landslide detector should be evaluated in its ability to detect landslide events that are not present in the training set.

In this study, we explore to what extent a segmentation model trained to detect landslides in one area is transferable to another area. We propose a simple augmentation-based setup to improve transferability. With SAR data, traditional computer vision geometric augmentations, such as rotation, would result in non-plausible imagery that would never be generated by SAR satellites [13]. We are careful to select augmentations that might simulate some semantically meaningless variance between different SAR images as follows. *Gaussian Blurring* instructs the model not to pay attention to the fine details of the imagery but to the more general structures. *Gaussian Noise* instructs the model to be invariant to noise in individual pixels. Since we add this noise in the log-transformed amplitudes, it simulates speckle noise, which characterizes SAR imagery. *Motion Blurring* simulates the effect of suboptimal focusing of the raw radar echoes into single-look complex SAR imagery. *Random cropping with resize* teaches the model a scale invariance for small and big landslides, and to capture landslides that are partially present in the image. For our experiments, we consider the datasets from two earthquake-triggered landslides in Mt Talakmau, Indonesia, and Hokkaido, Japan [14]. The datasets are provided as datacubes that cover the two regions and contain SAR polarimetry, interferometry data, topography data, and landslide labels. For our use case, we consider an emergency response setup where several SAR images are available before a landslide and only one SAR image is available after the landslide.

The machine learning pipeline is set up to solve a segmentation task, using a U-Net [15] model with a ResNet-50 encoder. We stack VV and VH bands before and after the landslide as different input channels of the segmentation model. Notice that regardless of the number of time steps we use, the number of SAR channels remains constant as we take the pixel-wise mean value across all time steps before the event. Landslides occur in particular landscapes, making features derived from Digital Elevation Models (DEM) common for the task. We consider aspect, slope, and curvature as additional input features and dismiss

This work has been supported by the Frontier Development Lab and its partners.

Table 1. Performance in terms of AUPRC for the experiments in the Hokkaido validation set and the Mount Talakmau test set. Best performance in bold, second best underlined. *SAR*: VV, VH before and after the landslides. *SAR & DEM*: SAR, plus DEM-derived features. *AUG*: Augmentations have been applied during training.

	Performance (AUPRC)	
	Hokkaido (Train Region)	Talakmau (Test Region)
SAR	0.618	<u>0.421</u>
SAR & DEM	**0.639**	0.388
SAR (AUG)	0.612	**0.456**
SAR & DEM (AUG)	<u>0.636</u>	0.402
No skill baseline	0.092	0.062

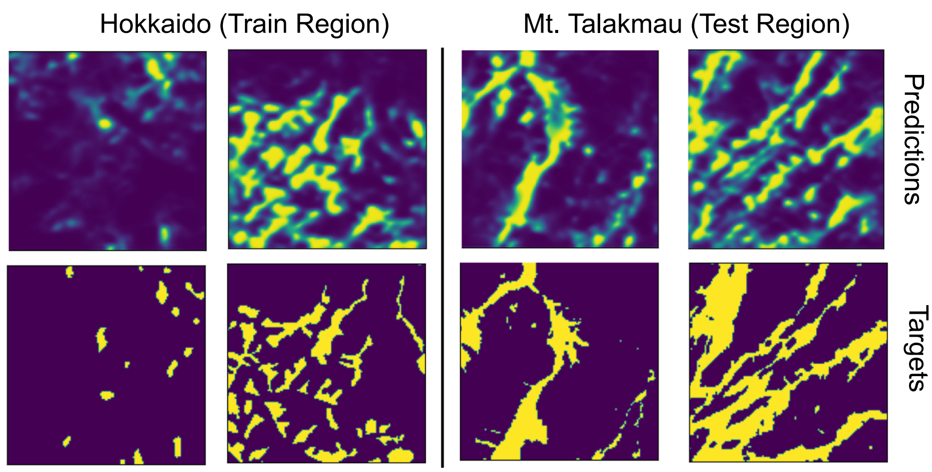

Fig. 1. Examples of the two regions' predicted landslide maps (top row) versus ground-truth target maps (bottom row) from the test set. Background (no landslide) areas are in purple, landslide affected areas are in yellow.

elevation itself as it is highly variant across different regions. We divide the two datacubes into data chips of size 128x128 pixels (each pixel is $10\text{m}\times 10m$ in spatial resolution), keeping only chips that include at least one landslide pixel to reduce the imbalance [10]. From the Hokkaido dataset, we extract 216 chips for training, and 61 chips for validation. We extract the test set containing 61 128×128 chips from the other area, namely Mt Talakmau, Indonesia. In the Hokkaido dataset, about 9% of pixels are labeled as landslides, while in the Mt Talakmau dataset, this percentage is 6%. Model performance is measured using the Area Under the Precision-Recall Curve (AUPRC), which is appropriate for imbalanced datasets.

Table 1 contains the results of the experiments. We see that the DEM-derived features are helpful for the region where the training has occurred, while they are detrimental to the transfer to another region. This is explained as a spatial overfitting of the model, that is avoided when removing the DEM-derived features. Interestingly, when adding the augmentations, the performance in Hokkaido (training region) is not affected, while it is consistently improved in the Mount Talakmau dataset (test region). This shows that training with augmentations increases the generalizability of the representations, improving the transfer of the model to another region. In Figure 1 we can see examples of the model's prediction versus the target image in both regions. As was also revealed by the metrics, the figure shows the model's skill both in the training and test region. While the segmentation is not perfect, the predictions match the target quite well.

In this study, we show how the results obtained in just one region do not necessarily generalize to other regions. While DEM-derived features seem helpful when training and testing in the same region [12], using them lowers the generalizability of the models. Potentially this could change with larger-scale labeled data. In the absence of large-scale data, we find that simple augmentations designed to work with SAR data help us train more robust models across different regions.

1. REFERENCES

[1] Stefano Luigi Gariano and Fausto Guzzetti, "Landslides in a changing climate," *Earth-Science Reviews*, vol. 162, pp. 227–252, Nov. 2016.

[2] Christian Huggel, Nikolay Khabarov, Oliver Korup, and Michael Obersteiner, "Physical impacts of climate change on landslide occurrence and related adaptation," in *Landslides*, John J. Clague and Douglas Stead, Eds., pp. 121–133. Cambridge University Press, 1 edition, Aug. 2012.

[3] Myles R. Allen and William J. Ingram, "Constraints on future changes in climate and the hydrologic cycle," *Nature*, vol. 419, no. 6903, pp. 228–232, Sept. 2002.

[4] William Ingram, "Increases all round," *Nature Climate Change*, vol. 6, no. 5, pp. 443–444, May 2016.

[5] Kieran T. Bhatia, Gabriel A. Vecchi, Thomas R. Knutson, Hiroyuki Murakami, James Kossin, Keith W. Dixon, and Carolyn E. Whitlock, "Recent increases in tropical cyclone intensification rates," *Nature Communications*, vol. 10, no. 1, pp. 635, Dec. 2019.

[6] Alexander L. Handwerger, Mong-Han Huang, Shannan Y. Jones, Pukar Amatya, Hannah R. Kerner, and Dalia B. Kirschbaum, "Generating landslide density heatmaps for rapid detection using open-access satellite radar data in Google Earth Engine," *Natural Hazards and Earth System Sciences*, vol. 22, no. 3, pp. 753–773, Mar. 2022, Publisher: Copernicus GmbH.

[7] Marja Machielse, *Landslide detection and mapping on Synthetic Aperture Radar amplitude satellite imagery*, Delft University of Technology, Delft, Netherlands, 2021.

[8] M. Reichstein, G. Camps-Valls, B. Stevens, J. Denzler, N. Carvalhais, M. Jung, and Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, pp. 195–204, Feb 2019.

[9] G. Camps-Valls, D. Tuia, Xiao Xiang Zhu, and M. (Editors) Reichstein, *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences*, Wiley & Sons, 2021.

[10] Lorenzo Nava, Oriol Monserrat, and Filippo Catani, "Improving Landslide Detection on SAR Data through Deep Learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021, arXiv:2105.00782 [cs].

[11] Lorenzo Nava, Kushanav Bhuyan, Sansar Raj Meena, Oriol Monserrat, and Filippo Catani, "Rapid Mapping of Landslides on SAR Data by Attention U-Net," *Remote Sensing*, vol. 14, no. 6, pp. 1449, Mar. 2022.

[12] Vanessa Boehm, Wei Ji Leong, Ragini Bal Mahesh, Ioannis Prapas, Edoardo Nemni, Freddie Kalaitzis, Siddha Ganju, and Raul Ramos-Pollan, "Deep learning for rapid landslide detection using synthetic aperture radar (sar) datacubes," *arXiv preprint arXiv:2211.02869*, 2022.

[13] Xiao Xiang Zhu, Sina Montazeri, Mohsin Ali, Yuansheng Hua, Yuanyuan Wang, Lichao Mou, Yilei Shi, Feng Xu, and Richard Bamler, "Deep Learning Meets SAR," Jan. 2021, Number: arXiv:2006.10027 arXiv:2006.10027 [cs, eess, stat].

[14] Vanessa Boehm, Wei Ji Leong, Ragini Bal Mahesh, Ioannis Prapas, Raul Ramos-Pollan, Siddha Ganju, Freddie Kalaitzis, and Edoardo Nemni, "Datacubes for landslide detection with sar imagery," Oct. 2022.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, arXiv:1505.04597 [cs].