IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 July 2024, accepted 19 August 2024, date of publication 26 August 2024, date of current version 3 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3449447

RESEARCH ARTICLE

Contrastive Self-Supervised Learning for Globally Distributed Landslide Detection

OMID GHORBANZADEH[1], HEJAR SHAHABI[2], SEPIDEH TAVAKKOLI PIRALILOU[3], ALESSANDRO CRIVELLARI[4], LAURA ELENA CUÉ LA ROSA[5], CLEMENT ATZBERGER[1], JONATHAN LI[6,7], (Fellow, IEEE), AND PEDRAM GHAMISI[8], (Senior Member, IEEE)

[1]Institute of Geomatics, University of Natural Resources and Life Sciences (BOKU), 1190 Vienna, Austria
[2]Center Eau Terre Environnement, Institut National de la Recherche Scientifique (INRS), Quebec City, QC G1K 9A9, Canada
[3]Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria
[4]Department of Geography, National Taiwan University, Taipei 106319, Taiwan
[5]Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
[6]Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[7]Department of System Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[8]Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany

Corresponding author: Alessandro Crivellari (alessandrocr@ntu.edu.tw)

This work was supported by the National Taiwan University under Grant NTU-NFG-113L7421.

ABSTRACT The Remote Sensing (RS) field continuously grapples with the challenge of transforming satellite data into actionable information. This ongoing issue results in an ever-growing accumulation of unlabeled data, complicating interpretation efforts. The situation becomes even more challenging when satellite data must be used immediately to identify the effects of a natural hazard. Self-supervised learning (SSL) offers a promising approach for learning image representations without labeled data. Once trained, an SSL model can address various tasks with significantly reduced requirements for labeled data. Despite advancements in SSL models, particularly those using contrastive learning methods like MoCo, SimCLR, and SwAV, their potential remains largely unexplored in the context of instance segmentation and semantic segmentation of satellite imagery. This study integrates SwAV within an auto-encoder framework to detect landslides using deca-metric resolution multi-spectral images from the globally-distributed large-scale landslide4sense (L4S) 2022 benchmark dataset, employing only 1% and 10% of the labeled data. Our proposed SSL auto-encoder model features two modules: SwAV, which assigns features to prototype vectors to generate encoder codes, and ResNets, serving as the decoder for the downstream task. With just 1% of labeled data, our SSL model performs comparably to ten state-of-the-art deep learning segmentation models that utilize 100% of the labeled data in a fully supervised manner. With 10% of labeled data, our SSL model outperforms all ten fully supervised counterparts trained with 100% of the labeled data.

INDEX TERMS Deep learning, landslide detection, multispectral imagery, natural hazard, remote sensing.

I. INTRODUCTION

Landslides, severe geohazards in mountainous regions, cause significant life loss and property damage [1]. From 2004 to 2016, landslides resulted in over 55,000 fatalities and yearly economic losses of USD 20 billion globally. These disasters pose immediate threats to humans and infrastructure, potentially causing floods, tsunamis, and deteriorated water quality by disrupting river sediment transport [2].

Landslide inventories are created from historical data or new data using various methods [3]. Field investigations, though reliable, are limited by safety and cost. Remote sensing (RS) allows for identifying landslides in inaccessible areas but isn't efficient for large areas with many landslides [4]. Nonetheless, it cannot be considered an efficient procedure in cases with numerous landslides of very different sizes distributed over large areas. Recent decades have seen

The associate editor coordinating the review of this manuscript and approving it for publication was Wei He.

the development of automated and semi-automated RS image classifications, employing machine learning to quickly detect landslides over extensive areas [5].

Traditional machine learning had challenges in landslide detection from RS images, including intricate spatial relationships and high-dimensional feature handling [6]. Deep learning, especially Convolutional Neural Networks (CNNs), outperforms these methods by learning complex representations and handling large datasets efficiently [4]. The initial use of CNNs for RS image-based landslide detection [4] pioneered the development of automatic techniques, sustained by many subsequent studies [7], [8], [9], [10], [11]. Since 2019, various advanced CNN algorithms with enhanced feature extraction have been developed for detecting landslides in deca-metric to decimetric RS data. Therefore, the challenges of landslide detection can be summarized as follows:

1) Difficulty of capturing the varied and complex nature of landslide patterns: Traditional machine learning methods struggle to effectively capture the complex and varied patterns of landslides, which can differ significantly across different regions and conditions.
2) Limited availability of labeled training data: There is often a scarcity of labeled training data for landslide detection, which hampers the effectiveness of traditional machine learning models.
3) High variability of landslide characteristics across different regions: The characteristics of landslides can vary greatly from one geographic region to another, making it difficult for traditional models to generalize and perform well universally.

Bi-temporal VHR images acquired from QuickBird satellite and a plane equipped with an aerial camera with 0.62 m and 0.5 m resolution, have been used for landslide detection in Lantau Island in Hong Kong, China by [12]. Ju et al. [8] applied YOLO v3, RetinaNet, and Mask R-CNN to automatically detect landslides from Google Earth mosaics images of different satellite images with 1 m resolution. Images with 0.25 m resolution captured by SF-300 Unmanned Air Vehicle (UAV) using a Canon EOS 5D Mark II camera have been used by [7] for training Mask R-CNN with different backbone networks, such as Swin Transformer and ResNet-50, for detecting landslides in Sichuan Province, China.

The landslide4sense (L4S) 2022 competition has provided the first globally distributed landslide benchmark data set, generating the attention of both the computer vision and RS communities for the application of deca-metric resolution multi-spectral images for landslide detection. In [13] the L4S 2022 landslide benchmark data set was introduced and evaluated using 11 state-of-the-art deep learning (DL) segmentation algorithms: DeepLab-v2, DeepLab-v3+, FRRN-A, FRRN-B, PSPNet, U-Net, ResU-Net, FCN-8s, LinkNet, SQNet, and ContextNet. A landslide detection task based on this benchmark data set has been implemented by Zhao et al. [14] using EfficientNetV2, Swin Transformer, and SegFormer. A U-Net-like skip connection structure was demonstrated using L4S 2022 benchmark data set by [15] using Swin Transformer [16] as the encoder part. As a result, they conducted spectral selection analyses on the benchmark data set to establish the best spectral selections for the self-attention mechanism that would allow the Swin Transformer to properly detect landslides. Bai et al. [10] have stated that the application of Deeplabv3, Deeplabv3+, and U-Net for the L4S 2022 benchmark data set led to results with limited accuracy. To segment landslides using deep learning, they proposed a multi-spectral U-Net consisting of two input streams for inputs of different resolutions.

This summary demonstrates that DL algorithms, trained on the L4S 2022 landslide benchmark dataset and similar supervised datasets like iSAID [17] require substantial annotated data [18]. These algorithms, common in RS and computer vision, often use transfer learning or fine-tuning to adapt to specific image types. However, extensive labeling, prone to ambiguity, is crucial, particularly in specialized areas like landslide detection where domain expertise is essential [13]. Furthermore, most big labeled data sets are generated by RGB images, whereas RS data sets generally provide useful information beyond the visible spectrum.

Self-supervised learning (SSL), gaining prominence since 2020, learns data representation without manual labels, rivaling supervised methods in image classifications [19], [20]. SSL models, particularly contrastive learning types like knowledge distillation (e.g., iBOT [21]), negative sampling (e.g., SimCLR [22] MoCo [23]), dual-branch training (e.g., SCL-GCN [24]) and Clustering [25], have gained much attention due to their strong learning ability to mine intrinsic characteristics of data without labels. Recent contrastive learning-based clustering advancements have narrowed the unsupervised-supervised learning gap in visual representation. These models generate cluster assignments, or image codes, to group similar features, yet were historically slow and computationally intensive. Innovations in models like SimCLR and MoCo overcome these limits using efficient data augmentation [25], achieving top performance by comparing features across augmented images [26]. Caron et al. [27] introduced SwAV, which utilizes online code computing [28] for consistent image coding across views. Our study adopts SwAV integrated with ResNet-18 for landslide detection instance segmentation, utilizing the L4S 2022 dataset. Additionally, we provide a comprehensive comparison of the accuracy obtained from SSL and supervised models by building different scenarios. Specifically, we are pursuing the following goals in our research:

1) to extend SSL models to landslide detection and use the publicly available L4S data set to create baseline results for future improvements
2) to adopt the SwAV model for pixel-wise landslide detection tasks by integrating it with an auto-encoder.

Therefore, this study introduces a novel approach by integrating SwAV within an auto-encoder framework for landslide detection, demonstrating how self-supervised learning can significantly reduce the reliance on labeled data. Our contributions extend beyond traditional methods by showcasing the effectiveness of SSL models in practical scenarios with minimal labeled data, thus paving the way for more scalable and efficient remote sensing applications.

II. METHOD

This study focuses on using multi-spectral Sentinel-2 images with different spectral bands for landslide detection. Adopting the framework by Caron et al. [27] it advances landslide detection. The SwAV model's core is training a neural network to distinguish between different views of an image's augmented versions, aiding in learning useful, discriminative features [29]. The ResNet-18 architecture is the encoder in this model. The network processes images through convolutional layers to extract hierarchical features, with the output from the final layer representing these features for SwAV's self-supervised objective (see Fig.1).

Following the training of the feature extractor on unlabeled image patches, the study moves to landslide detection. A U-Net network, based on ResNet-18, is developed for this purpose. The pre-trained ResNet-18 model acts as the encoder, with its weights frozen to preserve learned features. This encoder extracts high-level features from image patches.

Next, we train the decoder part of the U-Net model using labeled image patches representing 1% of the data. The feature fusion of the decoder with labeled data allows our model to produce precise landslide segmentation maps [30], [31]. This method benefits from the pre-trained ResNet-18's capacity to identify key landslide characteristics from unlabeled images. Fine-tuning the decoder with labeled data tailors these features to the specific task, enhancing performance and detection accuracy. This strategy merges knowledge from SwAV with limited labeled data, thus improving our model's landslide detection efficiency. Moreover, for comparison, we trained the U-Net network's encoder and decoder as a supervised model using the entire training dataset (100%), and again with just the 1% used in the SwAV-based SSL model.

A. BRIEF REVIEW OF SwAV

SwAV (Swapping Assignments between multiple Views) by Caron et al. [27] introduces an online method using contrastive learning without needing paired image comparisons. Typically, contrastive loss involves comparing representations from different images or views, essential for integrating diverse transformation-derived representations. These comparisons facilitate discriminative feature extraction, improving various computer vision tasks. Traditional methods, however, face challenges in comparing images pairwise across entire datasets for training. Many estimate loss from a limited image sample [22], unlike SwAV's focus on unsupervised learning. SwAV uses multiple views of the same image, ensuring consistency among codes, representing a swapping form Fig.2.

In SwaV, the loss function is computed using augmented versions of a single image. Given two different views of an image, represented by features z_t and z_s, their associated codes q_t and q_s are calculated by assigning z_t and z_s to a set of K prototypes $\{c_1,c_2,\ldots, c_K\}$. Therefore, the loss function for the swapped problem compares the features using the intermediate codes, as shown in Eq. 1:

$$L(z_t, z_s) = \ell(z_t, q_s) + \ell(z_s, q_t), \tag{1}$$

where ℓ is the loss function that compares the feature z with a code q. This approach bears resemblance to contrastive instance learning [32] but focuses on enforcing consistency within different views of an image, rather than solely targeting the codes. The approach involves a swapping paradigm, where the loss function is computed based on augmented versions of the image. By comparing features using intermediate codes, SwaV aims to ensure consistency within different views of an image. The ''swapped'' prediction problem is established based on this loss function, where the code q acts as the corresponding label (ground truth) for a given image feature z. The softmax operation is then performed by taking the dot product between z and the predicted output, which in this case corresponds to prototype C. The approach involves a swapping paradigm, where the loss function is computed based on augmented versions of the image. By comparing features using intermediate codes, SwaV aims to ensure consistency within different views of an image.

1) ONLINE CLUSTERING

Furthermore, SwaV incorporates an online clustering method to solve the cluster assignment problem. The method treats cluster assignment as an effective transport problem and utilizes the Sinkhorn Knopp algorithm [33]. Unlike standard cluster assignment methods, SwaV performs cluster assignments on a batch-by-batch basis. To allocate N given image features (Z) into K clusters, a matrix Q of codes is created. The mapping of image features to prototype vectors involves passing the features through a linear layer with the same number of neurons as the prototypes $\{c_1,c_2,\ldots, c_K\}$. This ensures similarity between Z and C while avoiding mapping different image features to the same prototype. The prototype vectors are optimized using the Sinkhorn Knopp algorithm.

2) DATA AUGMENTATION FOR CONTRASTIVE LEARNING

In SSL, it is imperative to design appropriate data augmentation procedures in order to acquire effective representations. Augmenting data by random crops is one of the most helpful procedures in contrastive learning. The augmentations provide information regarding the relations between parts of a targeted object or geographical features, like landslides in our case, within the image patch. In contrast, multiplying crops or ''views'' requires considerably more memory and processing power. However, Caron et al. [27] employ a simple

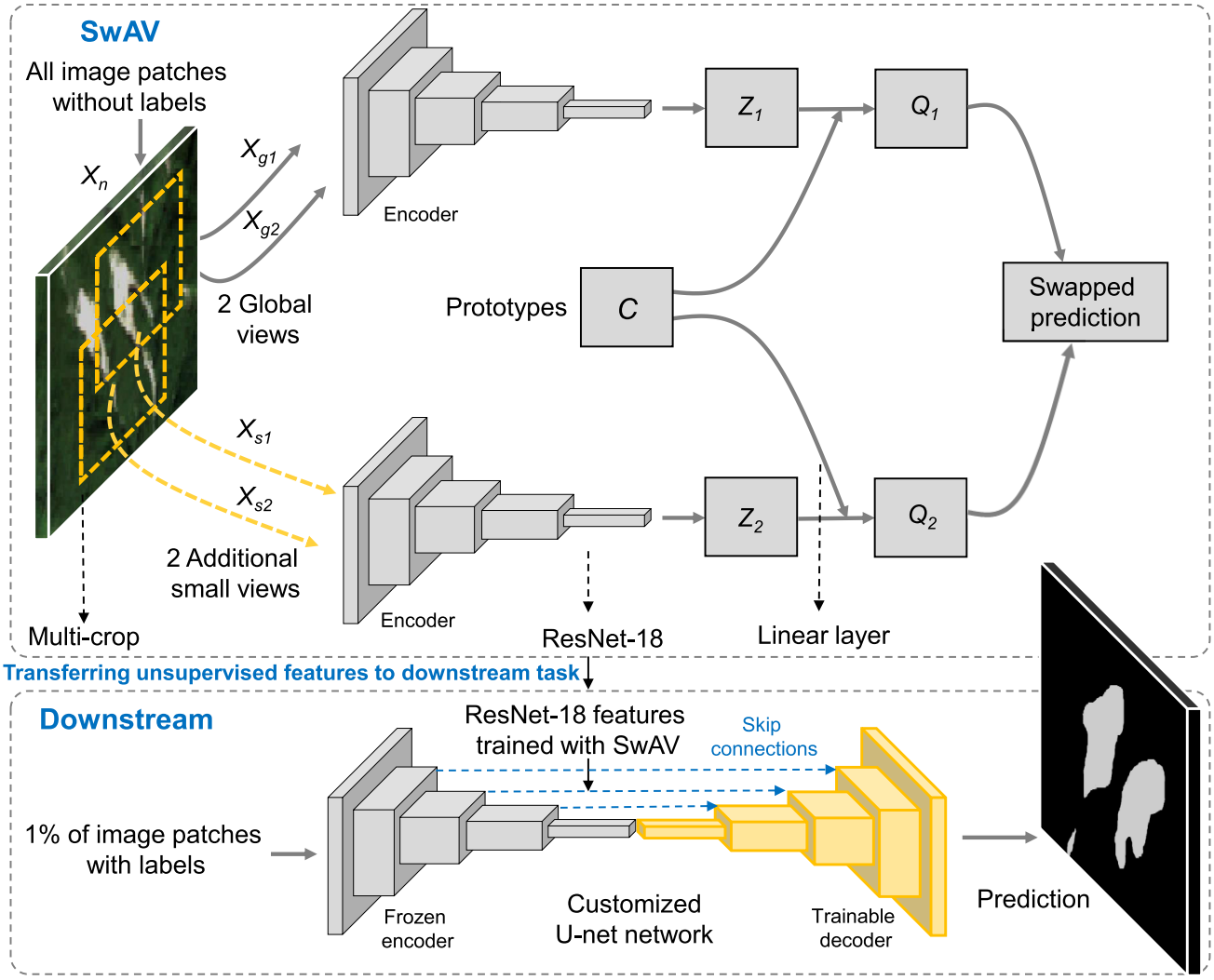

FIGURE 1. The overview of an integrated framework that combines SwAV with a customized U-Net network, built upon the ResNet-18 architecture. This framework is designed for the specific task of landslide detection. Where X_g and X_s refer to global and small views, respectively.

yet intelligent procedure of multi-crop data augmentation that uses two crops with the resolution of the original image and includes additional low-resolution crops covering a smaller part of the image patch. Taking low-resolution crops into account does not significantly increase computation costs. The loss of Eq. 2 is generalized.

$$L(z_{t_1}, z_{t_2}, \ldots, z_{t_{v+2}}) = \sum_{x \in 1,2} \sum_{v=1}^{V+2} 1_{v \neq i} \ell(z_{t_v}, q_{t_i}), \qquad (2)$$

where V is the number of added low-resolution crops. We refer the reader to Caron et al. [27] for results of some SSL models over ImageNet by applying the multi-crop augmentation procedure.

B. DOWNSTREAM TASK

One objective of using a pre-trained self-supervised model such as SwAV is to perform downstream tasks [34]. Deep convolutional networks, specifically ResNets [35], are key for decoding images and feature extraction. We utilize these extracted features for our custom U-Net network, which employs a ResNet-based encoder and symmetrical decoder blocks with skip connections. This consistency in design enhances the SLL model's feature extraction capabilities, crucial for tasks like landslide detection, leveraging convolutions, batch normalizations, activations, and downsampling. In this case, we use the original ResNet-18 architecture [36] as our feature extractor for SwAV. In the supervised scenarios, our custom-designed network is treated as a normal U-shaped network like U-Net. In the SSL model, however, we have removed the average pooling (Avg pool) and fully connected (FC) layers. Moreover, the weights of encoder features are frozen, while the weights of decoder features can be trained. It also reduces considerably the number of parameters to be estimated. This is done with only 1% of the training data set for the SSL model. On the last layer of the decoder, the

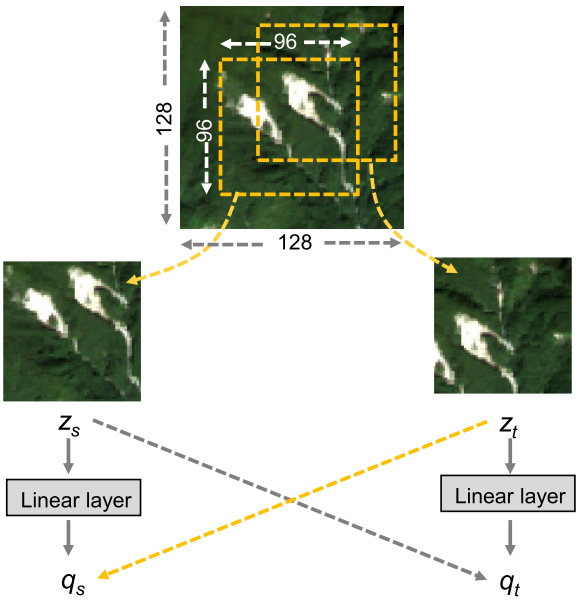

FIGURE 2. Creating randomly multi-crop views of the same landslide image patch in order to set up the swapped prediction problem, where z_t and z_s are image features of the corresponding views, and q_t and q_s are their respective codes.

sigmoid function is applied in order to create a probability map between zero and one for landslides. Unlabelled data for the unsupervised network was transformed and normalized in the same way as labeled data for the supervised network. The visualization of cluster prototypes in the SwAV model for landslide detection task is represented in Fig.3.

III. EXPERIMENTS

Experiments are carried out on the L4S benchmark data set to evaluate our SSL model for landslide detection. We first used the whole training data set, which comprised 3799 image patches for training the SwAV without labels. The generalization of ResNet-18 features trained by SwAV on the training set of the L4S made it possible to detect landslides on the test set of L4S using only 40 image patches from the training set (without augmentation).

In order to conduct a comparison, we trained both the encoder and decoder parts of the same U-Net network as supervised model ones using the whole training data set (100%) and one more time using only the same 1% that was used for the proposed SwAV based SSL model.

A. DATASET DESCRIPTION

Landslide4Sense (L4S) developed a medium-resolution, multi-spectral, multi-source landslide benchmark dataset for machine learning algorithms, with high-accuracy landslide image requirements. It includes 14 data layers: Sentinel-2 multi-spectral layers (bands 1-12), digital elevation model (DEM), and slope layer from ALOS PALSAR. L4S features pixel-wise landslide and non-landslide class labels, with training, validation, and testing sets collected at various times and locations. These sets have diverse characteristics like vegetation and topography, influenced by triggers like rainfalls and earthquakes. The dataset has image patches of 128 × 128 pixels, with 3799 training, 254 validation, and 800 test patches. The geographic location and details on the case study areas are disclosed in Fig.4 and for more details on the training section we refer to [13].

B. DATA SELECTION SCENARIOS

In a multi-spectral image, the sensitivity of different bands to the landslide features is variable, and some bands represent minimal responses to landslides [14]. A comprehensive evaluation of the role that topographic factors of DEM and slope play in landslide detection by [38] also emphasizes that the role of these data is different. Although we do not intend to assess the impact of data on landslide detection in this study, we define some scenarios that are based on different data selections for evaluating our SSL and supervised algorithms. Specifically, seven scenarios are defined: RGB alone; RGBNIR alone; RGB and DEM; RGBNIR and slope; RGB DEM and slope; RGBNIR DEM and slope; and finally all 14 bands provided by the L4S benchmark data set.

C. IMPLEMENTATION DETAILS

This study implements supervised and unsupervised processes using Python and PyTorch on an Ubuntu server with 4 NVIDIA Tesla T4 GPUs. The SwAV model utilizes global and small views sized (2, 128) and (2, 96) respectively, with a consistent batch size and epoch number of 128 and 2000. For the downstream task, limited to 40 samples (1% of training data for SwAV, are represented in Fig, 5), we adjust to a batch size of 16 and 1000 epochs. To effectively select labeled data and exclude non-landslide patches, a function was created to choose patches with at least 50 landslide pixels. The learning rate starts at 0.001, reducing by 0.95 each epoch. Adam optimizer and Binary Cross Entropy loss are used in the optimization process.

D. ACCURACY ASSESSMENT

To evaluate the performance of the models across different scenarios, predictions were generated using the test dataset consisting of 800 image patches from L4S. The accuracy assessment employed widely recognized metrics such as Precision, Recall, and F1 score. These metrics offer detailed insights into the model's ability to accurately classify landslides and non-landslide areas, ensuring a comprehensive evaluation of its performance.

E. RESULTS AND DISCUSSION

The SSL model's performance on the L4S benchmark data set involves seven scenarios, shown in Fig. 6 with various landslides from the test set. It compares three models' predictions, highlighting true positives in red, false negatives in green, and false positives in blue.

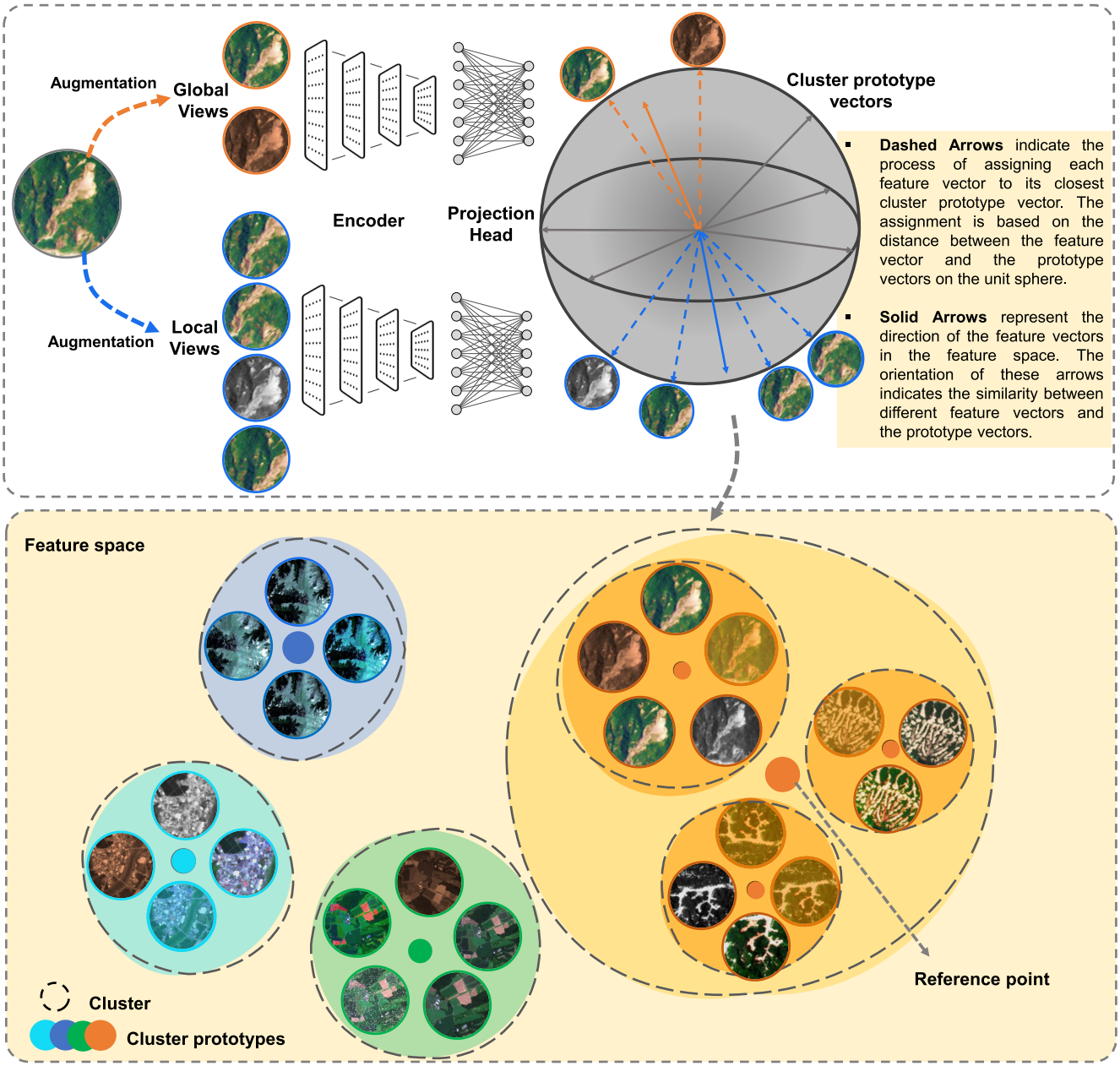

FIGURE 3. Visualization of cluster prototypes in the SwAV Model: This diagram showcases the clustering of image features into distinct groups based on their similarity. Each cluster, represented by a dashed circle, contains multiple image patches that share similar visual characteristics. The colored dots within each cluster indicate the cluster prototypes, serving as reference points for the assignment of image features during the learning process. The different colors of the cluster prototypes highlight the diversity of the clusters formed in the feature space.

Our landslide detection accuracy assessment results are presented in Tables 1- 7. As can be seen, the fully supervised U-Net network (trained with 100% of the labeled data) accomplishes higher performances compared to the SSL model (trained with 1% of the labeled data) in any data selection scenarios. However, in some cases, such as using RGB + Slope and RGB + Slope&DEM bands, the resulting F_1 of the SSL model is only 2 percentage points lower than that of the supervised model (see Tables 2 and 3 and Fig. 7). For the same cases, however, the SSL model achieves 22 and 37 higher F_1 percentage points than the supervised model trained with the same number of labeled data as our SSL model. In general, the average of the resulting F_1 of the fully supervised model based on all applied scenarios is more than 5 percentage points higher than that of the SSL model. The difference, however, is significantly higher in the opposite direction when using 1% of the labeled training data. As a result of using 40 labeled image patches from the training data set, the average F_1 of the SSL model is substantially greater than that of the supervised model.

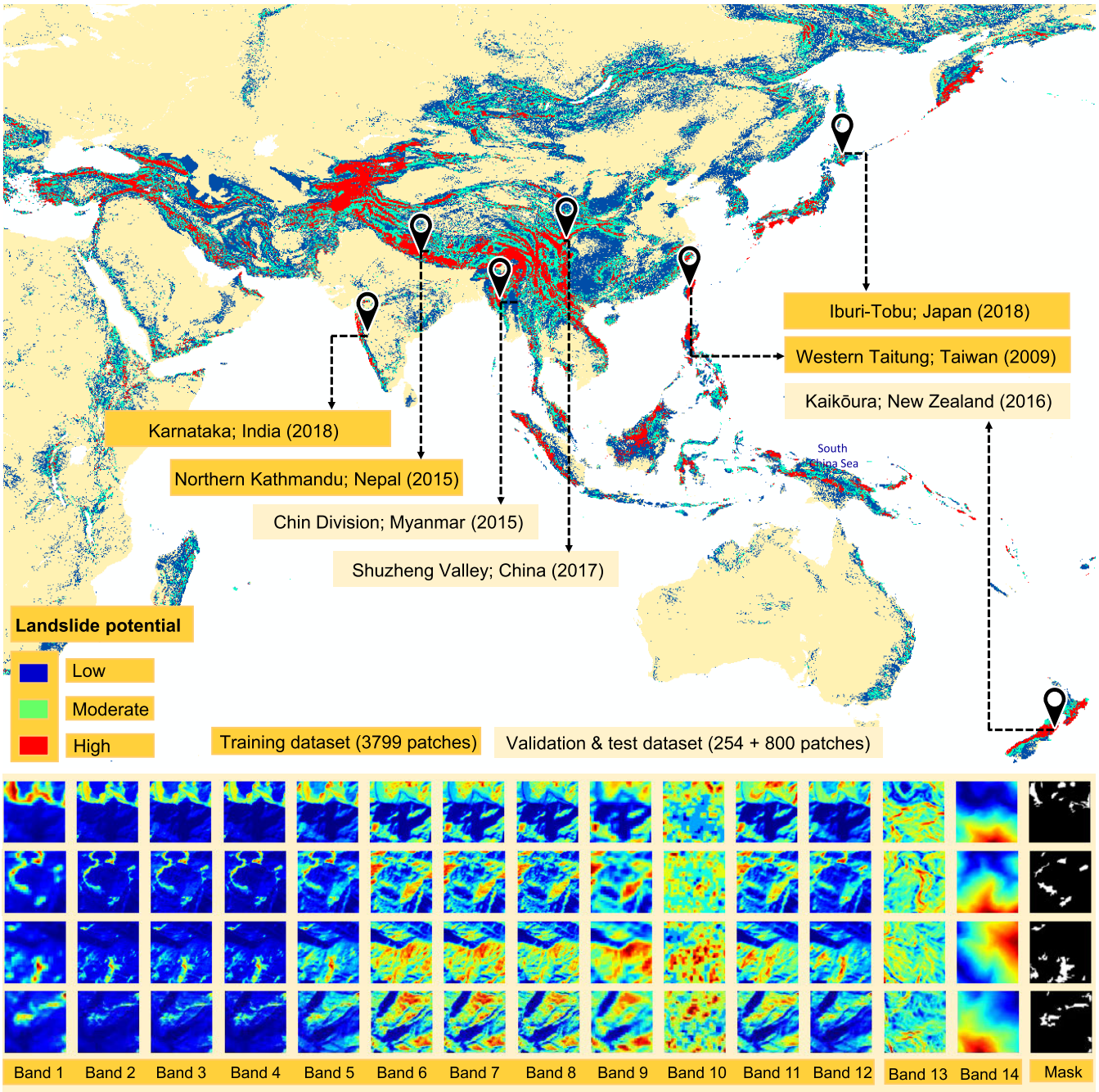

FIGURE 4. The locations of the training, validation, and test sites of landslide4sense (L4S) 2022 competition on a global landslide potential map generated by [37] and the visualization of every image layer in the 128 × 128 window size patches of the landslide dataset. Bands 1–12 represent the multi-spectral Sentinel-2 data, bands 13–14 represent the slope and DEM data, and the final column corresponds to the mask of annotated landslides.

Moreover, the best performance is for the fully supervised U-Net network trained by 5 bands of RGB&NIR + Slope with an F_1 value of more than 74%. Comparing this value with the lowest F_1 of 60% based on the same model trained by 3 bands of RGB illustrates the high sensitivity of our fully supervised U-Net network to the band selection scenarios. Similarly, the SSL model also gets the lowest F_1 value using only RGB bands. The supervised U-Net network trained by the limited number of labeled data also represents a very low F_1 value in this data selection scenario, but its lowest value is for using RGB + Slope&DEM bands with an F_1 of 24% for the landslide detection task. Therefore, regarding the data selection scenarios, RGB shows the lowest accuracy for almost all models, while RGB&NIR + Slope and RGB&NIR + Slope&DEM could considerably help models to represent their highest performance. In addition, the resulting F_1 values using all 14 bands provided by the L4S benchmark data set are the closest to the average F_1 of

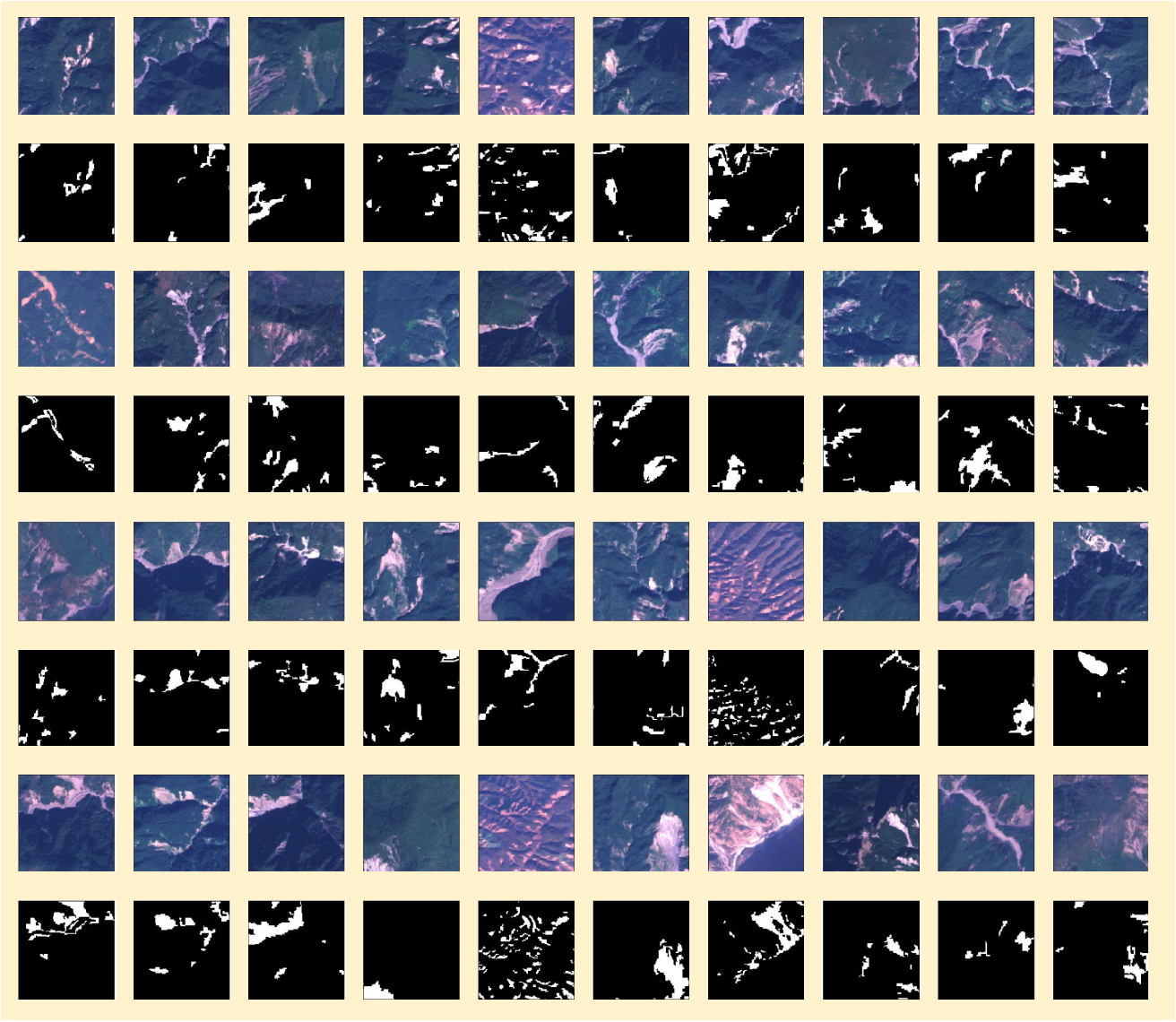

FIGURE 5. The applied 1% of training data, including 40 samples for the SSL model. They are selected randomly from the landslide4sense with the condition of having more than 50 landslide pixels in each patch. Representation is based on the RGB Sentinel-2 image patches and the corresponding pixel-wise labels of landslide and non-landslide.

all data selection scenarios. For the accuracy assessment metric of precision, the supervised U-Net network trained by 1% labeled data illustrates exceptional performance with a precision of 86%, using RGB + Slope&DEM, followed by the model trained with the whole training data set and all 14 bands by an 80% precision value. However, the recall for the former model is only 14%, while, that of the later model is 56%, which leads to very different F_1 values. The precision in all applied models is noticeably higher than recall, except in the scenario using RGB&NIR + Slope for the fully supervised U-Net network. There is the greatest difference between these two metrics within the accuracy assessment results of the supervised model with 1% labeled data. The comparatively higher precision compared to recall in the supervised U-Net network trained by 1% labeled data indicates that while the pixels detected as landslides are reliable enough and there are very fewer pixels that were incorrectly detected as landslides, many other landslides could not be identified by this model. It can be observed in Fig. 8 that the scenario RGB + Slope&DEM for the supervised network with 1% labeled data shows the most pronounced imbalances between precision and recall. In this figure, precision and recall are compared, using dashed yellow triangles to represent each model's accuracy assessment metrics.

F. COMPARISON WITH RELATED WORK

In this subsection, we further compare the proposed method with the state-of-the-art semantic segmentation models that are trained with a fully supervised learning manner using

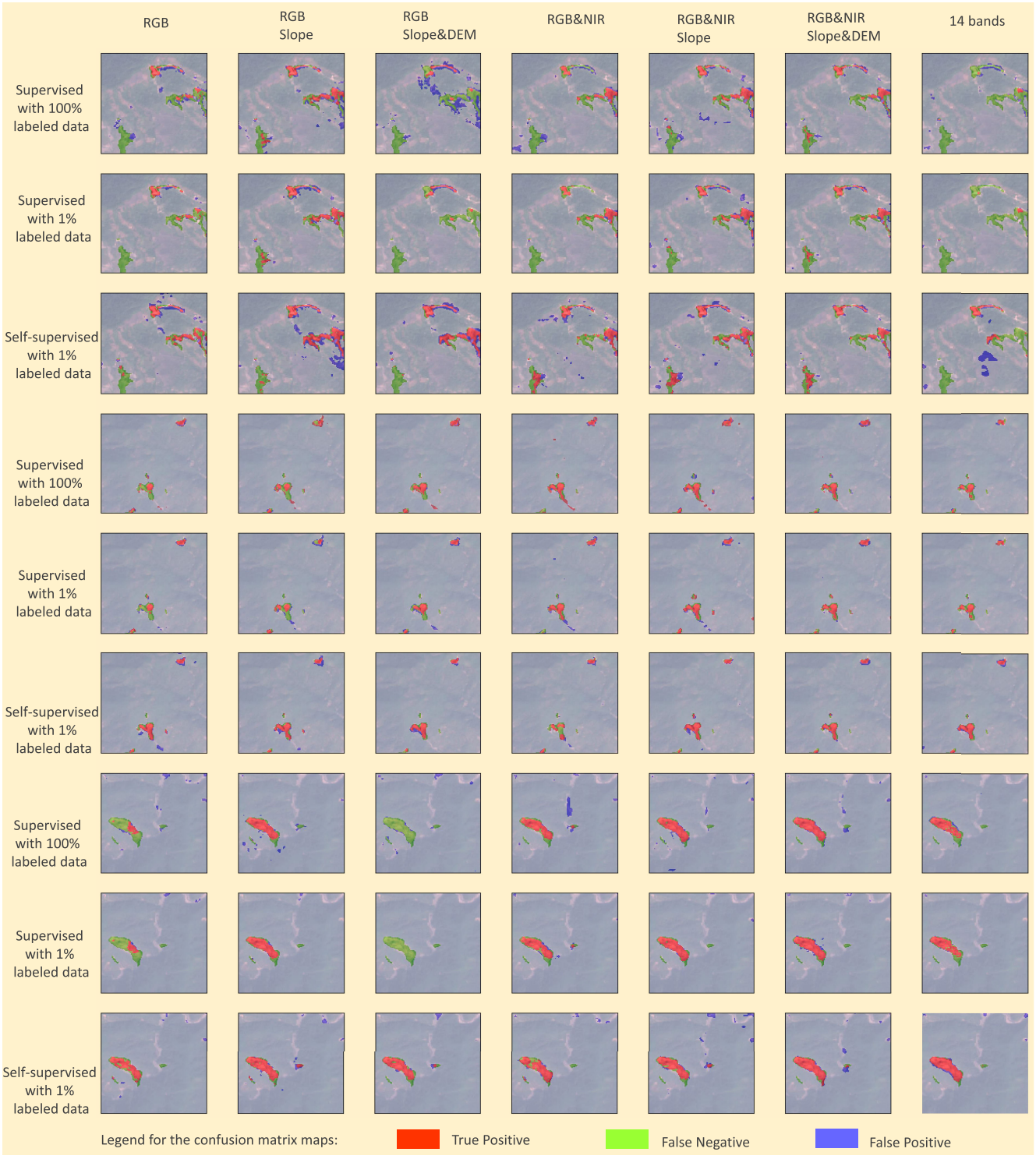

FIGURE 6. The landslide detection maps obtained by the applied supervised model with 1% and 100% of the labeled training data set and the SSL model with that of 1%.

100% of the labeled training data set and all 14 bands of the L4S benchmark data set. The selected models are FCN-8s [39], DeepLab-v2 [40], DeepLab-v3+ [41], ContextNet [42], SQNet [43], PSPNet [44], U-Net [45], ResU-Net [46], LinkNet [47], and FRRN-B [48]. It can be observed in Table 8 that due to the class imbalance between the landslide and non-landslide regions in the training set, most of the advanced segmentation networks adopted in the experiment can yield an F_1 score of around 60% or even less.

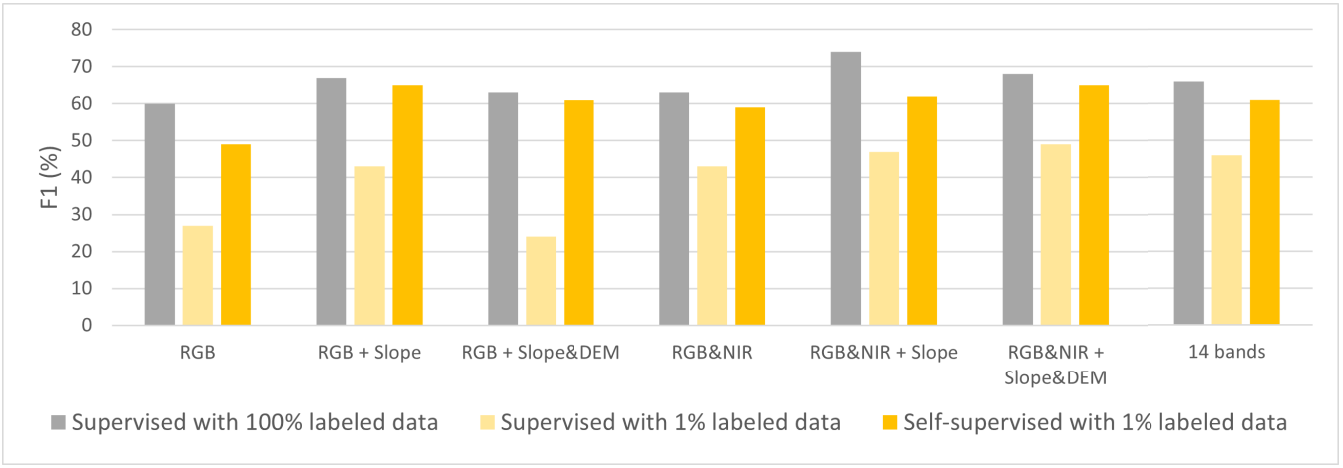

FIGURE 7. Comparison of the resulting F_1 scores of the applied supervised model with 1% and 100% of the labeled training data set and the SSL model with 1%.

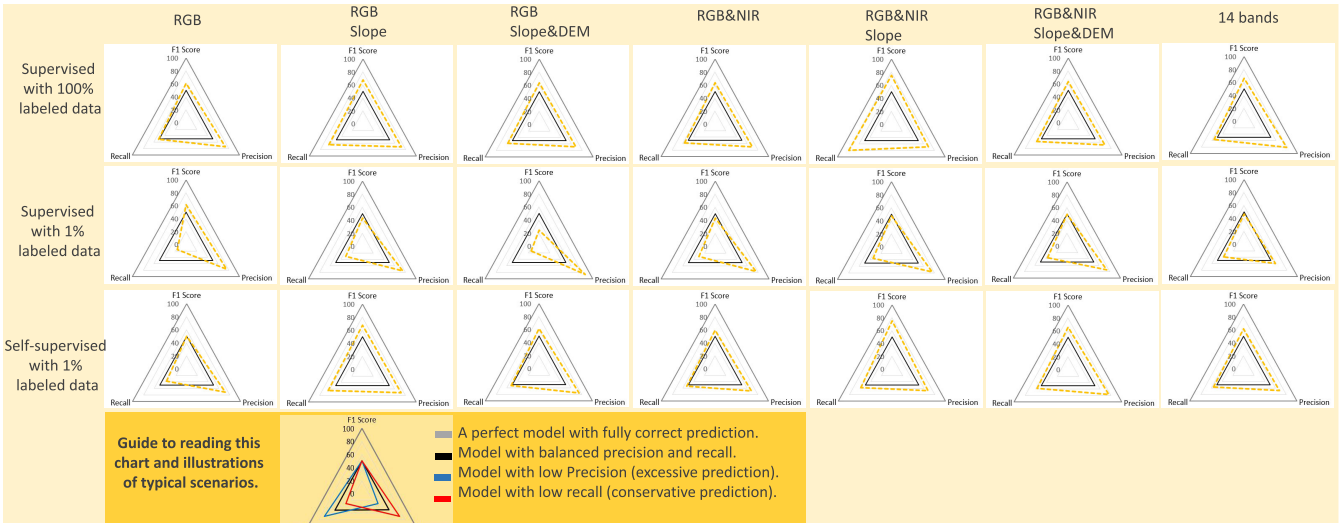

FIGURE 8. Spider plots showing the performance of the applied supervised model with 1% and 100% of the labeled training data set and the SSL model with 1% of that data set in three columns, whereas the rows represent seven data selection scenarios. The precision and recall scores are plotted on the X-axis (left and right angle of three angles), and the F_1 scores are plotted on the Y-axis (top angle).

TABLE 1. Quantitative results of using RGB bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	73	51	60
Supervised with 1% labeled data	76	16	27
Self-supervised with 1% labeled data	71	37	49

TABLE 2. Quantitative results of using RGB + slope bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	71	64	67
Supervised with 1% labeled data	75	31	43
Self-supervised with 1% labeled data	73	59	65

TABLE 3. Quantitative results of using RGB + Slope&DEM bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	68	58	63
Supervised with 1% labeled data	86	14	24
Self-supervised with 1% labeled data	75	51	61

TABLE 4. Quantitative results of using RGB&NIR bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	69	57	63
Supervised with 1% labeled data	76	30	43
Self-supervised with 1% labeled data	67	53	59

By contrast, the proposed architecture can achieve an F_1 score of around 66% in the fully supervised scenario, which outperforms the second-place model (i.e., FRRN-B) by 4 percentage points, (as shown in Table 8). However, if the number of training samples drops to 1% of the entire training set, the performance of the proposed network will also drop significantly with an F_1 score of only 46%. An inspiring phenomenon is that this result can be

TABLE 5. Quantitative results of using RGB&NIR + slope bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	69	79	74
Supervised with 1% labeled data	76	34	47
Self-supervised with 1% labeled data	67	58	62

TABLE 6. Quantitative results of using RGB&NIR + Slope&DEM bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	76	62	68
Supervised with 1% labeled data	74	36	49
Self-supervised with 1% labeled data	76	57	65

TABLE 7. Quantitative results of using all 14 bands (%).

	Precision	Recall	F_1
Supervised with 100% labeled data	80	56	66
Supervised with 1% labeled data	59	38	46
Self-supervised with 1% labeled data	67	56	61

dramatically improved to around 61% with the proposed SSL strategy, which is even competitive with state-of-the-art DL semantic segmentation models like FRRN-B in the fully supervised learning scenario. Therefore, our proposed SSL model could outperform 7 state-of-the-art DL segmentation models from 10 applied fully supervised models using only 1% of the labeled image patches from the training data set. The resulting F_1 score from this model is only 1% less than the other remaining 3 fully supervised models. Moreover, for a better comparison, and to provide benchmarking results based on our introduced contrastive self-supervised method of SwAV, we also applied 10% as the input feeding our auto-encoder for the downstream task. Using 10% of the labeled image patches from the training data set, the SSL model gets the highest F_1 score of more than 63%.

It should be noted that up to now, no other study has used SwAV or any other contrastive self-supervised methods for landslide detection. However, self-learning is applied based on the unlabeled image patches of the validation data set of the L4S by some studies to enhance the generalization of a model that is already trained using 100% of the labeled training data set [9], [14], [15], while in this study the SSL model was trained without using labels of the training data set (only 1% labeled patches were used). Therefore, since we did not use 99% of labels, our results are not directly comparable. Nevertheless, in the case of a fully supervised U-Net network trained by 5 bands of RGB&NIR + Slope, our resulting F_1 is very close to the best landslide detection result based on the L4S reported by Ghorbanzadeh et al. [38]. The main objective of this study is to demonstrate the applicability of SSL models, and contrastive learning in particular, for landslide detection, with results that are in some cases similar to those of fully supervised models trained with a large number of labeled images. Therefore, we compare our proposed framework with similar approaches used in RS. It should be noted that the number of these works is also very limited at the moment.

TABLE 8. Quantitative results of the state-of-the-art semantic segmentation models using all 14 bands (%).

	Model	Precision	Recall	F_1
100% labeled data	FCN-8s [39]	44	74	55
	DeepLab-v2 [40]	54	65	59
	DeepLab-v3+ [41]	49	77	60
	ContextNet [42]	40	70	51
	SQNet [43]	61	63	62
	PSPNet [44]	74	37	50
	U-Net [45]	61	63	62
	ResU-Net [46]	51	64	57
	LinkNet [47]	59	64	61
	FRRN-B [48]	53	74	62
1% labeled data	SSL (ours)	67	56	61
10% labeled data	SSL (ours)	72	55	63

In a very current work, Wang et al. [25] evaluated four prominent contrastive self-supervised procedures of Barlow Twins [49], SwAV, MoCo-v2 [50], and SimSiam [51], which are well-known for their capabilities in redundancy reduction, clustering, negative sampling, and knowledge distillation, respectively. The experiments were based on only Sentinel-2 images of three RS image patch data sets, SEN12MS [52], BigEarthNet [53], and So2SatLCZ42 [54], related to land cover, scene, and local climate zone classification. Similar to our study, they use ResNet-18 as encoder backbones and 128 × 128 image patches, applying frozen features from pre-trained contrastive self-supervised methods. Their results align with ours, showing slightly higher accuracy in fully supervised networks compared to SSL methods across all three RS image patch datasets.

IV. LIMITATIONS AND FUTURE WORK DIRECTIONS

Although the contrastive self-supervised method of SwAV shows very promising results on the downstream task of landslide detection, the performance of this method is not compared to others such as MoCo. Moreover, the applied encoder backbones are usually based on ResNet-50 in such applications, but here the ResNet-18 is used for the sake of simplicity. Regarding the applied benchmark data set, this study did not use the available unlabeled validation data set of L4S and only the training set is used without labels. The images in our dataset are limited to a size of 128 × 128 pixels, which may restrict the demonstration of our method's potential for large-scale scene applications. In addition, there are several well-developed architectures, such as UANet [55], G2LDIE [56], and multi-scale contrast enhancement model [57] designed for specific tasks. These architectures can be adapted and evaluated for their applicability to our future work, particularly for landslide segmentation using PlanetScope imagery. Consequently, the limitations mentioned above will be taken into consideration when determining the direction of our future research.

V. CONCLUSION

In the domain of RS image classification and segmentation tasks, the scarcity of labeled data has led to a rise in

unsupervised and SSL models. This study explores SSL applications for natural hazard analysis, particularly in landslide detection. We introduce an approach employing the contrastive SwAV procedure for effective landslide detection without human annotations. Our framework yields results on par with fully supervised models. To evaluate SSL models' efficiency in landslide detection from RS images, we compared the performance of the same network under fully supervised conditions and when supervised using limited labeled images, as in SSL. Our proposed SSL model achieves the highest accuracy of an F_1 value of 65% in two different data selection scenarios. More intriguingly, the proposed SSL model trained with only 1% of the labeled data can achieve similar or even better performance than ten state-of-the-art DL segmentation models that are trained with 100% of the labeled data. Moreover, the SSL model can easily obtain a higher F_1 score than fully supervised models using 10% of the labeled data. The effectiveness of our SSL method in landslide detection with limited labeled data is highlighted. We focused on the SSL model's performance using the public L4S benchmark dataset, which provides diverse data selection scenarios, facilitating comparisons with future advanced SSL algorithms. Additionally, our findings underscore the importance of appropriate data selection in both supervised and SSL approaches, although the impact on method performance varies. This research opens to feasible adaptations of contrastive self-supervised methodologies in the context of landslide detection, paving the way for future comparative analyses on the benefits of different self-supervised strategies in landslide-related applications.

ACKNOWLEDGMENT

The authors thank all the anonymous reviewers for their insightful comments.

REFERENCES

[1] F. Guzzetti, P. Reichenbach, F. Ardizzone, M. Cardinali, and M. Galli, "Estimating the quality of landslide susceptibility models," *Geomorphology*, vol. 81, nos. 1–2, pp. 166–184, Nov. 2006.

[2] K. B. Sim, M. L. Lee, and S. Y. Wong, "A review of landslide acceptable risk and tolerable risk," *Geoenvironmental Disasters*, vol. 9, no. 1, pp. 1–17, Dec. 2022.

[3] G. Xu, Y. Wang, L. Wang, L. P. Soares, and C. H. Grohmann, "Feature-based constraint deep CNN method for mapping rainfall-induced landslides in remote regions with mountainous terrain: An application to Brazil," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2644–2659, 2022.

[4] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, and J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sens.*, vol. 11, no. 2, p. 196, Jan. 2019.

[5] Y. Huang, J. Zhang, H. He, Y. Jia, R. Chen, Y. Ge, Z. Ming, L. Zhang, and H. Li, "MAST: An earthquake-triggered landslides extraction method combining morphological analysis edge recognition with Swin-transformer deep learning model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2586–2595, 2024.

[6] P. Lu, W. Shi, and Z. Li, "Landslide mapping from PlanetScope images using improved region-based level set evolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[7] R. Fu, J. He, G. Liu, W. Li, J. Mao, M. He, and Y. Lin, "Fast seismic landslide detection based on improved mask R-CNN," *Remote Sens.*, vol. 14, no. 16, p. 3928, Aug. 2022.

[8] Y. Ju, Q. Xu, S. Jin, W. Li, Y. Su, X. Dong, and Q. Guo, "Loess landslide detection using object detection algorithms in Northwest China," *Remote Sens.*, vol. 14, no. 5, p. 1182, Feb. 2022.

[9] F. Zhang, Y. Shi, Q. Xu, Z. Xiong, W. Yao, and X. X. Zhu13, "On the generalization of the semantic segmentation model for landslide detection," in *Proc. CDCEO@ IJCAI*, 2022, pp. 96–100.

[10] L. Bai, W. Li, Q. Xu, W. Peng, K. Chen, Z. Duan, and H. Lu, "Multispectral U-Net: A semantic segmentation model using multispectral bands fusion mechanism for landslide detection," in *Proc. CDCEO@ IJCAI*, 2022, pp. 101–104.

[11] C. Fang, X. Fan, H. Zhong, L. Lombardo, H. Tanyas, and X. Wang, "A novel historical landslide detection approach based on LiDAR and lightweight attention U-Net," *Remote Sens.*, vol. 14, no. 17, p. 4357, Sep. 2022.

[12] Z. Lv, F. Wang, W. Sun, Z. You, N. Falco, and J. A. Benediktsson, "Landslide inventory mapping on VHR images via adaptive region shape similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5630211.

[13] O. Ghorbanzadeh, Y. Xu, P. Ghamisi, M. Kopp, and D. Kreil, "Landslide4Sense: Reference benchmark data and deep learning models for landslide detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5633017.

[14] H. Zhao, J. Wang, Y. Pan, A. Ma, X. Wang, and Y. Zhong, "Progressive label refinement-based distribution adaptation framework for landslide detection," in *Proc. CDCEO@ IJCAI*, 2022, pp. 87–90.

[15] D. Zhao, Q. Zang, Z. Wang, D. Quan, and S. Wang, "Swinls: Adapting Swin transformer to landslide detection," in *Proc. CDCEO@ IJCAI*, 2022, pp. 91–95.

[16] Z. Liu, Y. Tan, Q. He, and Y. Xiao, "SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4486–4497, Jul. 2022.

[17] S. W. Zamir, "iSAID: A large-scale dataset for instance segmentation in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2019, pp. 28–37.

[18] P. Berg, M.-T. Pham, and N. Courty, "Self-supervised learning for scene classification in remote sensing: Current state of the art and perspectives," *Remote Sens.*, vol. 14, no. 16, p. 3995, Aug. 2022.

[19] V. Stojnic and V. Risojevic, "Self-supervised learning of remote sensing scene representations using contrastive multiview coding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1182–1191.

[20] S. Zhang, X. Zhang, S. Wan, W. Ren, L. Zhao, and L. Shen, "Generative adversarial and self-supervised dehazing network," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 4187–4197, Mar. 2024.

[21] J. Zhou, C. Wei, H. Wang, W. Shen, C. Xie, A. Yuille, and T. Kong, "IBOT: Image BERT pre-training with online tokenizer," 2021, *arXiv:2111.07832*.

[22] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[23] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[24] H. Feng, L. Ma, Y. Yu, Y. Chen, and J. Li, "SCL-GCN: Stratified contrastive learning graph convolution network for pavement crack detection from mobile LiDAR point clouds," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, Apr. 2023, Art. no. 103248.

[25] Y. Wang, C. M Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," 2022, *arXiv:2206.13188*.

[26] T. Si, F. He, Z. Zhang, and Y. Duan, "Hybrid contrastive learning for unsupervised person re-identification," *IEEE Trans. Multimedia*, vol. 25, pp. 4323–4334, 2022.

[27] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," in *Proc. NIPS*, Dec. 2020, pp. 9912–9924.

[28] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 132–149.

[29] J. Zhu, K. Yang, N. Guan, X. Yi, and C. Qiu, "HCPNet: Learning discriminative prototypes for few-shot remote sensing image scene classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 123, Sep. 2023, Art. no. 103447.

[30] C. Yang, T. Liu, G. Chen, and W. Li, "ICSFF: Information constraint on self-supervised feature fusion for few-shot remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5800312.

[31] S. Zhang, W. Ren, X. Tan, Z.-J. Wang, Y. Liu, J. Zhang, X. Zhang, and X. Cao, "Semantic-aware dehazing network with adaptive feature fusion," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 454–467, Jan. 2023.

[32] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3733–3742.

[33] Y. M. Asano, C. Rupprecht, and A. Vedaldi, "Self-labelling via simultaneous clustering and representation learning," 2019, *arXiv:1911.05371*.

[34] M. Patacchiola and A. J. Storkey, "Self-supervised relational reasoning for representation learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 4003–4014.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[36] F. Ramzan, M. U. G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, and Z. Mehmood, "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," *J. Med. Syst.*, vol. 44, no. 2, pp. 1–16, Feb. 2020.

[37] T. Stanley and D. B. Kirschbaum, "A heuristic approach to global landslide susceptibility mapping," *Natural Hazards*, vol. 87, no. 1, pp. 145–164, May 2017.

[38] O. Ghorbanzadeh, Y. Xu, H. Zhao, J. Wang, Y. Zhong, D. Zhao, Q. Zang, S. Wang, F. Zhang, Y. Shi, X. X. Zhu, L. Bai, W. Li, W. Peng, and P. Ghamisi, "The outcome of the 2022 Landslide4Sense competition: Advanced landslide detection from multisource satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9927–9942, 2022.

[39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[40] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[41] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[42] R. P K Poudel, U. Bonde, S. Liwicki, and C. Zach, "ContextNet: Exploring context and detail for semantic segmentation in real-time," 2018, *arXiv:1805.04554*.

[43] M. Treml et al., "Speeding up semantic segmentation for autonomous driving," in *Proc. MLITS, NIPS Workshop*, vol. 2, no. 7, 2016.

[44] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[46] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[47] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[48] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3309–3318.

[49] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12310–12320.

[50] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," 2020, *arXiv:2003.04297*.

[51] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15745–15753.

[52] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS—A curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion," 2019, *arXiv:1906.07789*.

[53] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "BigearthNet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5901–5904.

[54] X. Xiang Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Häberle, Y. Hua, R. Huang, L. Hughes, H. Li, Y. Sun, G. Zhang, S. Han, M. Schmitt, and Y. Wang, "So2Sat LCZ42: A benchmark dataset for global local climate zones classification," 2019, *arXiv:1912.12171*.

[55] W. He, J. Li, W. Cao, L. Zhang, and H. Zhang, "Building extraction from remote sensing images via an uncertainty-aware network," 2023, *arXiv:2307.12309*.

[56] J. Sun, W. He, and H. Zhang, "G2LDIE: Global-to-local dynamic information enhancement framework for weakly supervised building extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5406714.

[57] Y. Liu, Z. Yan, T. Ye, A. Wu, and Y. Li, "Single nighttime image dehazing based on unified variational decomposition model and multi-scale contrast enhancement," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105373.

OMID GHORBANZADEH received the Ph.D. degree in applied geo-informatics from the University of Salzburg, Austria, in 2021. He is a Machine Learning Researcher (Postdoctoral) with the AI4RS Group, Institute of Advanced Research in Artificial Intelligence, Vienna, Austria. He is a GIS and Remote Sensing Expert with a strong background in research and development of machine (deep) learning models to monitor land cover dynamics and natural hazards. Currently, he is leading the EO4CerealStress Project with the University of Natural Resources and Life Sciences (BOKU), Vienna, funded by ESA.

HEJAR SHAHABI is currently pursuing the Ph.D. degree with INRS, Quebec City, Canada. He is also a Data Scientist with Beneva. He carry out research in natural hazard mapping and risk assessment using multi-temporal/multi-source remote sensing data and machine learning algorithms. His core research interests are object-based image analysis (OBIA) and its integration with deep learning for mapping and object detection and developing early warning systems for natural hazards.

SEPIDEH TAVAKKOLI PIRALILOU received the Ph.D. degree in applied geo-informatics from the University of Salzburg, Austria, in 2022. She is a Researcher with the Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria. Her primary research interests include satellite image processing and its integration with deep learning techniques for mapping and feature extraction. Additionally, she focuses on developing spatial analysis methods and geospatial data visualization.

ALESSANDRO CRIVELLARI received the B.Sc. and M.Sc. degrees from the Politecnico di Milano, and the Ph.D. degree in applied geo-informatics from the University of Salzburg. He is an Assistant Professor with the Department of Geography, National Taiwan University. His past experiences include a Postdoctoral Fellowship with the Southern University of Science and Technology and a Research Internship with the Prosus AI Team. His professional interests revolve around the domains of geoAI, spatial data science, and geo-informatics, with a main focus on spatial–temporal predictive analytics and satellite image processing.

LAURA ELENA CUÉ LA ROSA received the Ph.D. degree from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, in 2022. During the Ph.D. degree, she was an Intern with IBM Research Brazil, Rio de Janeiro, and a Ph.D. Guest with the Helmholtz Institute Freiberg for Resource Technology (HZDR), Freiberg, Germany. She currently holds a postdoctoral position with the Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research (WUR), Wageningen, The Netherlands. Her professional interests primarily revolve around the application of deep learning methods (DL) in the field of RS image analysis.

CLEMENT ATZBERGER received the Ph.D. degree in crop growth modeling and remote sensing data assimilation from Trier University, Germany, in 1997. He is a Full Professor and the Head of the Institute for Surveying, Remote Sensing and Land Information and the Institute for Surveying, Remote Sensing and Land Information, University of Natural Resources and Applied Life Sciences, Vienna. He is also the Head of research with Mantle Labs, London, U.K.

JONATHAN LI (Fellow, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, Cape Town, South Africa, in 2000. He is currently a Professor of geomatics and systems design engineering with the University of Waterloo, Waterloo, ON, Canada. He has supervised more than 120 master's and Ph.D. students and postdoctoral fellows to completion. He has co-authored more than 490 publications, more than 320 of which were published in refereed journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, ISPRS-JPRS, RSE, and JAG. He has also published papers in flagship conferences in computer vision and AI, including CVPR, AAAI, and IJCAI. His main research interests include AI-based information extraction from Earth observation images and LiDAR point clouds, and 3-D vision and GeoAI. He is a fellow of the Royal Society of Canada (RSC) Academy of Science, the Canadian Academy of Engineering (CAE), and the Engineering Institute of Canada (EIC). He is the Editor-in-Chief of *International Journal of Applied Earth Observation and Geoinformation* (JAG) and an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

PEDRAM GHAMISI (Senior Member, IEEE) received the M.Sc. degree (Hons.) in remote sensing from the K. N. Toosi University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015. He is the Head of the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology (HZDR-HIF), Germany; a Co-Founder of VasoGnosis Inc., with two branches in the USA; and a Visiting Professor and a Leader of AI4RS, Institute of Advanced Research in Artificial Intelligence, Austria. He was the Vice-Chair of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee, from 2019 to 2021.