

The Outcome of the 2022 Landslide4Sense Competition: Advanced Landslide Detection from Multi-Source Satellite Imagery

Omid Ghorbanzadeh, Yonghao Xu, *Member, IEEE*, Hengwei Zhao[†], *Student Member, IEEE*, Junjue Wang[†], *Student Member, IEEE*, Yanfei Zhong, *Senior Member, IEEE*, Dong Zhao, Qi Zang, Shuang Wang, *Member, IEEE*, Fahong Zhang, Yilei Shi, *Member, IEEE*, Xiao Xiang Zhu, *Fellow, IEEE*, Lin Bai, Weile Li, Weihang Peng, and Pedram Ghamisi, *Senior Member, IEEE*

Abstract—The scientific outcomes of the 2022 Landslide4Sense (L4S) competition organized by the Institute of Advanced Research in Artificial Intelligence (IARAI) are presented here. The objective of the competition is to automatically detect landslides based on large-scale multiple sources of satellite imagery collected globally. The 2022 L4S aims to foster interdisciplinary research on recent developments in deep learning (DL) models for the semantic segmentation task using satellite imagery. In the past few years, DL-based models have achieved performance that meets expectations on image interpretation, due to the development of convolutional neural networks (CNNs). The main objective of this article is to present the details and the best-performing algorithms featured in this competition. The winning solutions are elaborated with state-of-the-art models like the Swin Transformer, SegFormer, and U-Net. Advanced machine learning techniques and strategies such as hard example mining, self-training, and mix-up data augmentation are also considered. Moreover, we describe the L4S benchmark data set in order to facilitate further comparisons, and report the results of the accuracy assessment online. The data is accessible on *Future Development Leaderboard* for future evaluation at <https://www.iarai.ac.at/landslide4sense/challenge/>, and researchers are invited to submit more prediction results, evaluate the accuracy of their

methods, compare them with those of other users, and, ideally, improve the landslide detection results reported in this article.

Index Terms—Deep learning, landslide detection, multi-spectral imagery, natural hazard, remote sensing.

I. INTRODUCTION

LANDSLIDES are a frequent natural hazard observed in mountainous terrains across the globe [1]. There are several mechanisms by which soil, rock, and objects located on the ground or underground on an unstable hill slope can move downward and create a landslide [2]. Landslides mainly occur in response to natural processes like heavy rainfalls and earthquakes or human-induced activities [3]. The downward movement of the most catastrophic landslides is fast. They can travel large distances and take down everything in their path, creating scars on higher slopes and accumulating to deposition in valleys [4]. Landslides in mountainous areas are a problem, responsible for substantial losses, including damage to buildings and infrastructure and even fatalities [5]. The current climate changes, population growth, and rapid urbanization in areas vulnerable to natural hazards have also increased the occurrence of landslides and their consequences [6]. As a result, in recent years, a considerable amount of attention has been paid to gaining a better understanding of the mechanisms of these catastrophic hazards [7]. The most vital information regarding these catastrophic events is the awareness of past movements and their exact locations and extensions, ideally recorded in a landslide inventory data set [5]. Such a data set is an essential requirement for extracting advanced information, developing knowledge in the field, and predicting the unstable slopes that are prone to landslides [8]–[10]. Prediction maps generated from such a data set can be used for potential mitigation measures for the region under the study [11]. Therefore, a more accurate and detailed landslide inventory data set is a prerequisite for a precise disaster mitigation action [12].

In the past decade, deep learning (DL) has gained a great deal of attention, both in computer vision and remote sensing (RS) image analyses. The application of deep learning and convolutional neural networks (CNNs) to the detection of landslides emerged in early 2019, primarily using very high resolution (VHR) [14] and hyperspectral RS data [15]. The

Manuscript received XX 2021. Corresponding authors: Yonghao Xu; Omid Ghorbanzadeh.

O. Ghorbanzadeh, and Y. Xu are with the Institute of Advanced Research in Artificial Intelligence (IARAI), Landstraßer Hauptstraße 5, 1030 Vienna, Austria (e-mail: omid.ghorbanzadeh@iarai.ac.at; yonghao.xu@iarai.ac.at).

H. Zhao, J. Wang, and Y. Zhong are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 430074, China (e-mail: whu_zhaohw@whu.edu.cn; king-drone@whu.edu.cn; zhongyanfei@whu.edu.cn).

D. Zhao, Q. Zang, and S. Wang are with the School of Artificial Intelligence, Xidian University, Xian, 710071, China (e-mail: shwang@mail.xidian.edu.cn).

F. Zhang is with the Data Science in Earth Observation, Technical University of Munich, 80333 Munich, Germany. Y. Shi is with the Chair of Remote Sensing Technology, Technical University of Munich, 80333 Munich, Germany. X. Zhu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany and Signal Processing in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: xiaoxiang.zhu@dlr.de).

L. Bai, W. Li, and W. Peng are with the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, No.1 East Third Road, Erxianqiao, Chenghua District, Chengdu, 610059, China (e-mail: bailin@cdut.edu.cn; liweile08@cdut.edu.cn; pangdaren@outlook.com).

P. Ghamisi is with (1) Institute of Advanced Research in Artificial Intelligence (IARAI), Landstraßer Hauptstraße 5, 1030 Vienna, Austria; (2) Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Chemnitz Str. 40, 09599 Freiberg, Germany (e-mail: p.ghamisi@gmail.com).

[†]These authors contributed equally.

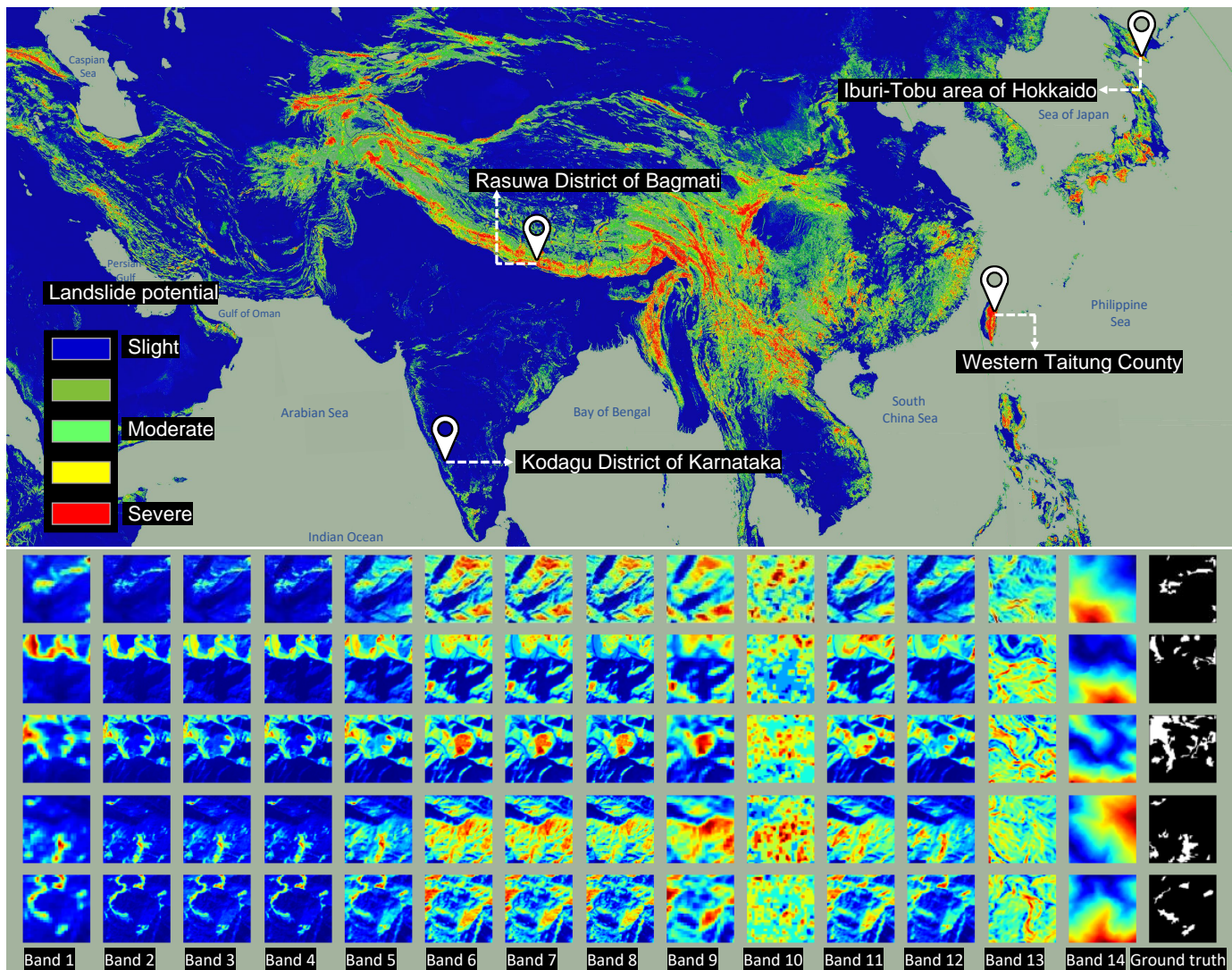


Fig. 1. The locations of the training sites on a global image of landslide susceptibility generated by [13] and the visualization of every image layer in the 128×128 window size patches of the landslide training data set. Multi-spectral Sentinel-2 data is represented by bands 1–12, and slope and DEM data is represented by bands 13–14. The patches in the last column refer to the corresponding ground truth polygons.

prospect of generating landslide maps with more accuracy than can be achieved with traditional methods such as semi-automated [16] and machine learning classifiers [17] has encouraged researchers in this field to develop and apply more sophisticated DL algorithms. To the best of our knowledge, no DL algorithm has been designed specifically for the distinct characteristics of landslide detection. Therefore the application of existing DL models and their variations for this task poses some new concerns, namely their transferability to new geographical areas with different landcovers and morphologies and the lack of any comprehensive open-source benchmark data set [12].

The artificial intelligence for remote sensing (AI4RS) group of the Institute of Advanced Research in Artificial Intelligence (IARAI) is a small international group of scientists working on the development and application of state-of-the-art deep learning (DL) solutions and algorithms for satellite imagery interpretation. This group has organized the Landslide4Sense (L4S) competition to foster ideas and progress in DL al-

gorithms for the specific Earth observation application of landslide detection. The competition provides participants with a landslide benchmark data set with globally distributed multi-source satellite imagery. The benchmark data set is prepared and introduced as an explicit norm for evaluating alternative DL approaches. The training set, which is a subset of the whole benchmark data set is released and thoroughly described by Ghorbanzadeh *et al.* in [6]. The study evaluates this subset of the benchmark data set using 11 different state-of-the-art DL segmentation models.

The L4S competition fosters interdisciplinary research in computer vision, artificial intelligence (AI), and RS image analysis for image classification and landslide detection. The global objective is to build DL-based models for understanding the differentiating characteristics of landslides based on the provided optical, digital elevation model (DEM), and slope layers from freely available satellite imagery acquired by Sentinel-2 sensors and ALOS PALSAR. During the L4S competition, along with the highest accuracy assessment results,

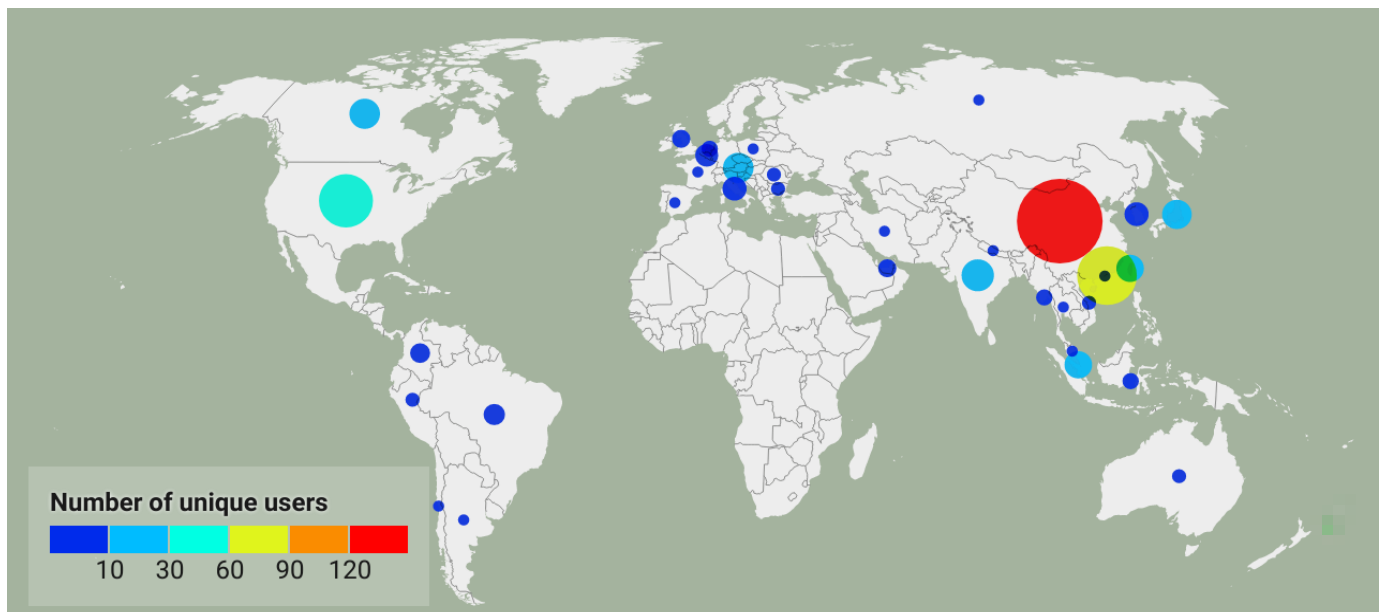


Fig. 2. Global distribution and number of unique users per country or region, created in <https://app.datawrapper.de>.

a special prize was also awarded for the most creative and innovative solution.

The competition is organized by IARAI and aims to improve automatic landslide detection DL algorithms using multi-source satellite imagery. In this competition, the main objective is the creation of landslide inventory maps using only the specified labeled landslide benchmark data set as training data.

The main focus of this article is on the scientific outcomes of the L4S competition. The rest of the paper is organized as follows. Section II describes the L4S benchmark data set used in the competition. Section III provides statistics of submissions and the overall results of the competition. In the next four sections, we discuss the DL algorithms proposed by the first- to third-ranked teams and the team of the special prize. Finally, we summarize our concluding points in Section VIII.

II. THE DATA AND BASELINE OF LANDSLIDE4SENSE COMPETITION 2022

A. Data Set

The benchmark data set for the L4S competition comprises 14 layers of data: multi-spectral data from Sentinel-2 (band1–band12), digital elevation model (DEM), and slope data from ALOS PALSAR. All 14 layers in the landslide benchmark data set are resized to the resolution of about 10m per pixel and are labeled pixel-wise to landslide and non-landslide classes. The landslide benchmark data consists of the training, validation, and test sets that encompass events occurring across a wide range of geographical locations throughout the world's mountainous regions. Specifically, only the training subset is acquired from four different sites: the Ihuri-Tobu area of Hokkaido, the Kodagu district of Karnataka, the Rasuwa district of Bagmati, and western Taitung County. A number of methods have been proposed in various relevant studies for the annotation of landslides, in particular, data preprocessing

workflows [18], manual interpretation methods [14], [19]–[23], conventional machine learning models [24], and object-based methods [25]. Most of the current landslide detection studies typically use manual interpretation of satellite imagery as the basis for creating their image labels. Whatever the case, landslides can be annotated, and therefore the images can be labeled in a variety of ways depending on the expert's opinion. [26]. Consequently, this study used a two-step procedure for landslide annotation and image labeling to address this issue. An object-based method was first developed for detecting the landslide, followed by manual verification and correction of all landslide annotations. We begin by calculating the relevant image difference indexes based on pre- and post-event images. The events may differ for the case study areas considered for generating the training, validation and test data sets. With the resulting indexes (e.g., the red-over-green, the brightness, and vegetation Index differences) and the original post-landslide images, multi-resolution segmentation (MRS) technique was used to generate meaningful image segments. The MRS technique produces the image segments through an iterative process, grouping neighboring pixels until the predefined scale and shape parameters are reached. A rule-based image classification based on several thresholds that vary depending on the study area distinguishes landslides from other features. In [16], [17], [27], [28], similar methodologies and rule-based segmentation and classification have been described. In the second step of our landslide annotation and image labeling procedure, a variety of further data sources were then used to visually correct landslide polygons, specifically Google Earth images, and previous landslide inventory data sets provided in a number of studies like [29], were used to enhance the first step's results. Annotations indicate the precise position and clear boundaries of landslides; we do not include any other information, including the type and volume of mass movement of the landslide. The data collected from these four

TABLE I
BASELINE APPROACH RESULTS IN THE VALIDATION AND TEST PHASES

	Precision (%)	Recall (%)	F1 Score (%)
Validation	51.75	65.50	57.82
Test	52.45	69.87	59.92

sites provide 3,799 image patches with a size of 128×128 pixels (see Fig. 1). The validation and test sets contain 245 and 800 image patches of the same size, respectively, which were acquired from other geographical sites. Details about the 14 layers of the landslide benchmark data set are given below.

- **Sentinel-2.** The multi-spectral Sentinel-2 layers are provided in wavelengths of ultra blue, blue, green, red, visible and near infrared (VNIR), and short wave infrared (SWIR). The bands (B2, B3, B4, B8) have a spatial resolution of 10m, whereas those of (B5, B6, B7, B11, B12) and (B1, B9, B10) have a spatial resolution of 20m and 60m, respectively. This imagery is captured during cloud-free days after the event.
- **ALOS PALSAR.** The ALOS phased array type l-band synthetic aperture radar layers have a spatial resolution of 12.5 m and were acquired from 2006 to 2019. The Alaska satellite facility (ASF) is one of the distributed active archive centers that provides high-resolution DEM from ALOS PALSAR at no cost to the user. The slope layer is derived from the DEM ALOS PALSAR, and both DEM and slope layers are converted to 10m spatial resolution. More details about the landslide benchmark data set can be found in [6].

The task of the L4S competition is to predict landslides from the data set provided. The labels are only provided for 3,799 image patches of the training data set. The landslide detection results are evaluated with the pixel-wise F1 Score on the landslide category in both the validation and test phases. Rankings for the competition were determined using only this accuracy assessment metric. However, competitors also received precision and recall metrics during the validation phase to get more meaningful feedback for their landslide detection results.

B. Baseline

We provided a simple baseline in our public GitHub repository prior to the start of the L4S competition.¹ A state-of-the-art DL model for semantic segmentation was implemented in PyTorch in order to provide this service. This model contains a user-configurable training script for U-Net [30] and the data loader for reading the training and test data sets. U-Net was first applied to biomedical image segmentation, followed by numerous semantic segmentation applications that demonstrated successful results. This model is also common for the landslide detection task and has been applied in a number of studies [12], [31], [32]. U-Net comprises an encoder route capable of capturing low-level representations and a decoder route designed to capture high-level representations.

¹<https://github.com/iarai/Landslide4Sense-2022>

As the decoder route is asymmetrical, where the vanished content of the localization is restored by using an asymmetrical design, the encoder route follows a standard CNN design assembled from consecutive convolution blocks. There is a max-pooling layer with a filter size of 2×2 and a stride of 2, after two convolutional layers with a filter size of 3×3 , leveraging the rectified linear unit (ReLU) activation function in each block [30]. The baseline model implemented in the L4S competition includes 23 convolutional layers, of which 4 are convolutional-transpose layers. The baseline U-Net model is trained using the training data set and tested on 245 and 800 image patches of validation and test data sets, respectively. The resulting baseline accuracy for the validation and test data sets is represented in Table I. We used all 14 bands for training and testing, and no additional measurements were applied (e.g., data augmentation, pre- or post-processing). Adding any external auxiliary data such as very high-resolution images was forbidden, as specified in the L4S competition terms and conditions. The best performance of the baseline model achieves an F1 score of 59.92% on the test set.

III. SUBMISSIONS AND RESULTS

There were 439 unique users within 85 teams that submitted 7775 landslide detection results to the validation phase of the L4S competition website.² The number of total submissions to the test leaderboard was 219 landslide detection results. We limited submissions per team to ten for the test phase. The final ranking was determined based on the highest F1 Score of each team during the test phase. Moreover, a special prize was also considered for the most creative and innovative solution in landslide detection, in the view of the L4S scientific committee. The competitors were from 37 different countries or regions. Most of the competitors were from mainland China, with 134 unique users, followed by 62 from Hong Kong, 50 from the USA, and 42 from Germany. Fig. 2 shows the distribution and the approximate number of unique users per country or region.

The first three ranked teams that were selected based on their highest F1 Scores and the team recipient of the special prize were named winners of the L4S competition and presented their solutions during IJCAI-ECAI 2022, the 31st International Joint Conference on Artificial Intelligence, and the 25th European Conference on Artificial Intelligence at the Workshop on Complex Data Challenges in Earth Observation, CDCEO 2022.

The four winning teams are as follows:

- **1st place:** *Kingdrone* team; Junjue Wang,³ Hengwei Zhao,³ Yang Pan, Ailong Ma, Xinyu Wang, and Yanfei Zhong from Wuhan University, China.
- **2nd place:** *Seek* team; Dong Zhao, Qi Zang, Zining Wang, Dou Quan, and Shuang Wang from Xidian University, China.
- **3rd place:** *Tanmlh* team; Fahong Zhang, Zhitong Xiong, Qingsong Xu, Wei Yao, Yilei Shi, and Xiao Xiang Zhu

²<https://www.iarai.ac.at/landslide4sense/challenge/>

³These authors contributed equally.

TABLE II
THE FOUR WINNING SOLUTIONS WITH F1 SCORES ON THE TEST LEADERBOARD AND TYPE OF BACKBONE NETWORKS AND STRATEGIES USED

Award	Team	Backbone Network	Strategies							F1 Score (%)
			Lovasz loss	Self-training	Band selection	Mix-up	DenseCRF	Ensemble		
1st place	Kingdrone	Swin Transformer, SegFormer, and U-Net	✓	✓					✓	74.54
2nd place	Seek	Swin Transformer	✓	✓	✓					73.99
3rd place	Tanmlh	DeeplabV3+ and HRNet	✓	✓		✓		✓		73.50
Special prize	Sklgp	Enhanced U-Net			✓					71.29

from Technical University of Munich (TUM), Germany; German Aerospace Center (DLR), Germany.

- **Special prize:** *Sklgp* team; Qiang Xu, Weile Li, Lin Bai, Kai Chen, Weihang Peng, Zhenzhen Duan, and Huiyan Lu from Chengdu University of Technology, China.

Table II summarizes the four winning solutions with F1 scores and types of backbone networks and strategies used. It can be observed that all these winning solutions are based on DL models. While the *Tanmlh* team and the *Sklgp* team adopt advanced fully convolutional networks (FCNs) like DeeplabV3+ [33] and U-Net [30] as the main backbone networks in their solutions, the *Kingdrone* team and the *Seek* team further take the recent state-of-the-art Vision Transformer architectures like Swin Transformer [34] and SegFormer [35] into consideration. Besides, it is interesting to find that all top three solutions adopt the Lovasz loss [36] and the self-training technique [37] as auxiliary strategies to train their models in addition to the traditional cross-entropy loss. Considering that different channels may have different effects on landslide detection, the *Seek* team and the *Sklgp* team further adopt the band selection technique in their solutions. Other strategies like the mix-up augmentation [38], dense conditional random field (DenseCRF) [39], and ensemble learning are also considered in the winning teams. Detailed descriptions of the winning solutions can be found in the following sections.

IV. FIRST-PLACE TEAM

A. Analysis of the Characteristics of Landslide

A progressive label refinement-based distribution adaptation landslide detection framework was proposed by the first-place team for large-scale landslide detection. The unique characteristics of landslides create two particular challenges for large-scale landslide detection from remote sensing images: **small objects and class imbalance**, and **distribution inconsistency**.

The first challenge, small objects and class imbalance, is shown in Fig. 3. In remote sensing images, the morphology of landslides is very complex, especially with many small branches, which belong to small objects (Fig. 3a). Furthermore, the landslide is not the dominant ground object in large-scale remote sensing images, as shown in Fig. 3b, which illustrates the statistical result of the training data set in which the proportion of pixels occupied by the landslide is only 2%, and the number of pixels of other ground objects (background) is 49 times that of the landslide. Both of these challenges, of small objects and class imbalances, lead to lower recall scores.

Distribution inconsistency is another difficult challenge for large-scale landslide detection from remote sensing imagery. In real-world large-scale landslide detection applications, images of landslides to be detected come from all over the

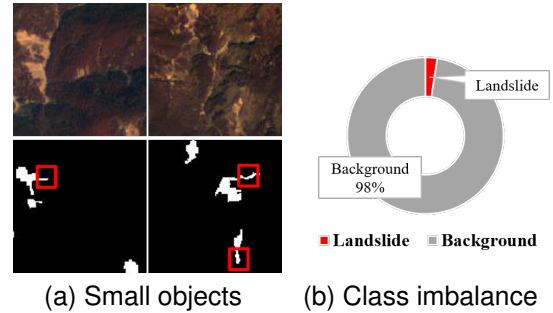


Fig. 3. The problem of small objects and class imbalance. The landslide has some smaller branches and the background has 49 times as many pixels as the landslide.

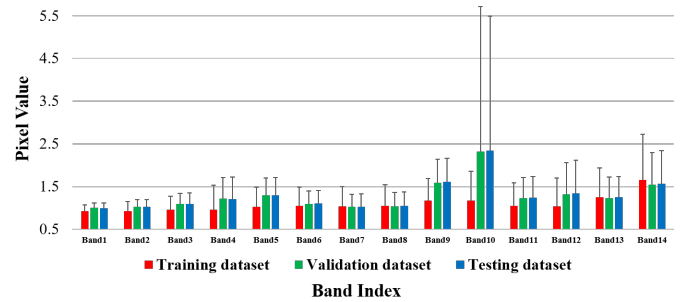


Fig. 4. The problem of distribution inconsistency. The statistical results are calculated band by band, and are significantly different among the training data set, the validation data set, and the testing data set.

world. These images are collected at different times, which leads to different imaging conditions. This spatio-temporal difference leads, in turn, to great differences in radiation values or pixel values of different remote sensing images, especially in the mountains [40], and is characterized by statistical inconsistency. In Fig. 4, the mean and standard deviation of the remote sensing images of the training, validation, and testing datasets are calculated and displayed band by band. The histogram is the mean and the error bar represents the standard deviations. The statistical results of the training data set and the validation data set (testing data set) are quite different, causing the poor generalization performance of the model trained with the training data set data on the validation data set (testing data set).

B. Progressive Label Refinement-based Distribution Adaptation Framework

To address the challenges of large-scale landslide detection, a progressive label refinement-based distribution adaptation framework is proposed by the first-place team for landslide

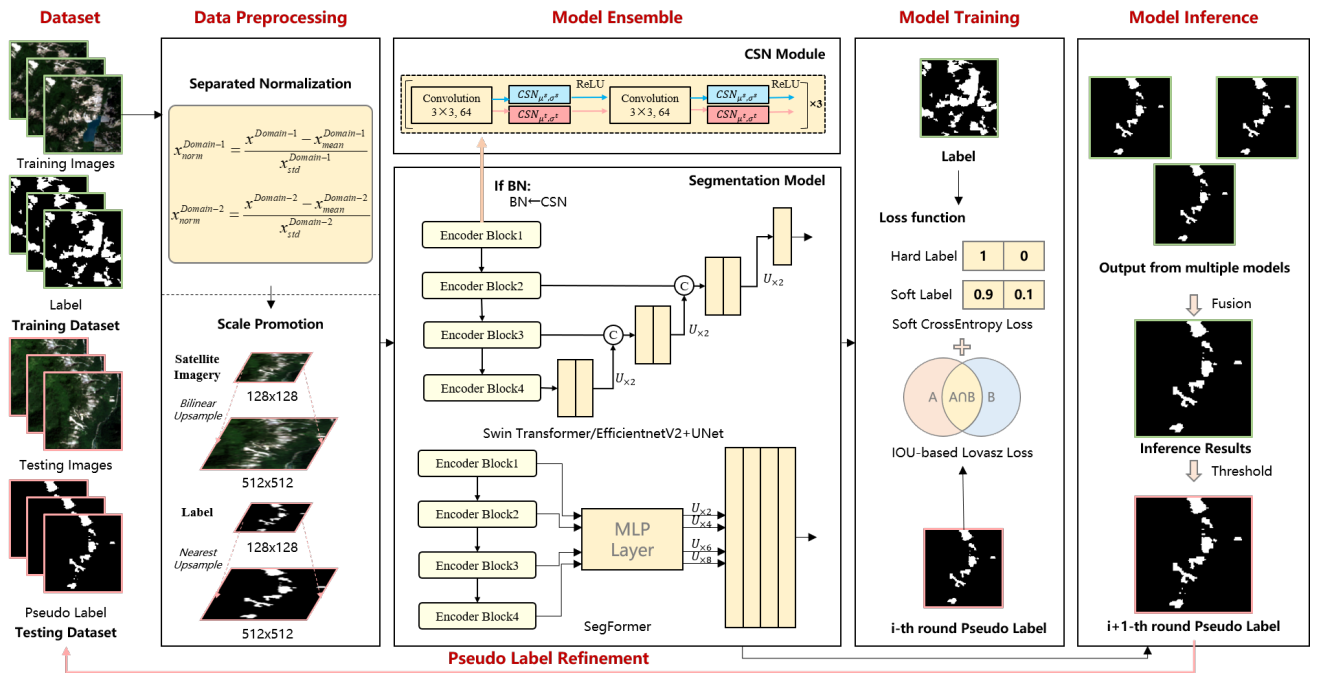


Fig. 5. Progressive label refinement-based distribution adaptation landslide detection framework.

detection. As shown in Fig. 5, the proposed framework includes data preprocessing, model ensemble, model training, model inference, and pseudo-label refinement.

1) *Data preprocessing*: Scale promotion is used to resist the weak representation caused by small landslide branches; the original images are scaled up from 128×128 pixels to 512×512 pixels. Random flip, random rotation, and color perturbation also are adopted for data augmentation. Color perturbation is only used for multi-spectral data, not DEM and slope data.

Separated normalization is proposed to alleviate the distribution inconsistency challenge in the data preprocessing stage, which uses the mean and variance from different domains to normalize the data. For example, two different domains are the training domain from the training data set and the validation domain from the validation data set in the model validation stage. The mean and standard deviation are calculated from the two datasets respectively, and then the data in the two domains are normalized respectively. Separated normalization is similar to the normalization for cross-sensor transfer learning [41], but the operation of domain-specific statistics is performed in the data preprocessing stage.

2) *Model ensemble and training*: In the segmentation model, three models are used to integrate the final landslide detection results. The U-Decoder architecture considering multiple scales is selected as the decoder to further alleviate the small object problem, and Swin Transformer [34] and EfficientNetV2 [42] are selected as encoders to capture complex features of the landslide. This framework also uses SegFormer [35], which utilizes self-attention operations to fit the variant shapes of landslides, and the MLP, which is used to enhance the difficult sample features. To further increase the generalization of the model, the batch normalization in the

three segmentation models is replaced by cross-sensor normalization [41] to encode the statistical consistency between the training data set and the validation (testing) data set.

As for model training, Lovasz loss [36] and an online hard example mining strategy are used to address the problem of class imbalance, and soft cross-entropy loss [43] is used to solve the problem of noisy labels in the pseudo labels.

3) *Model inference and pseudo label refinement*: The probability values output by the above three models are averaged as the final prediction results in the inference stage.

To further alleviate the distribution inconsistency problem, the validation (testing) data set is used in the training process, and progressive pseudo-label refinement is proposed to generate pseudo labels for validation or testing images. Based on the prediction of the i th round, pseudo labels of the $(i + 1)$ th round can be generated using a probability threshold of 0.7. The models of the $(i + 1)$ th round can be trained by training data set and validation (testing) images with pseudo labels. The domain adaptive consistency training and the generation of pseudo labels are performed iteratively, and the pseudo labels are refined progressively.

C. Experimental Results

We conducted a series of experiments to evaluate our proposed method on the Landslide4Sense data set. All bands in the multi-source images were used as inputs during training and testing.

The results in Fig. 6 show that each proposed module improves landslide detection accuracy in the different aspects. In particular, the separate normalization achieves the greatest improvement, addressing the distribution inconsistencies in multi-source data. As the number of label refinements increase,

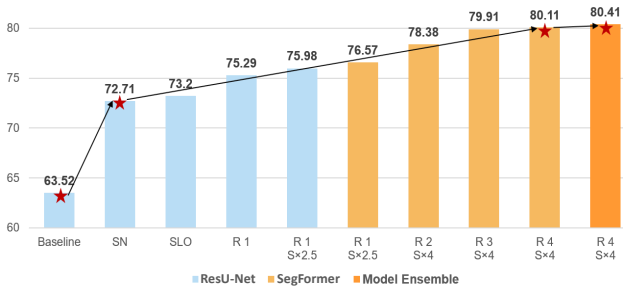


Fig. 6. The experimental results on the validation leaderboard. Each proposed module improves overall performance in the different aspects and is compatible with the other modules. SN – Separate Normalization, SLO – Soft cross-entropy loss+Lovasz loss+OHEM, Sx – Scale promotion, R – refinement round.

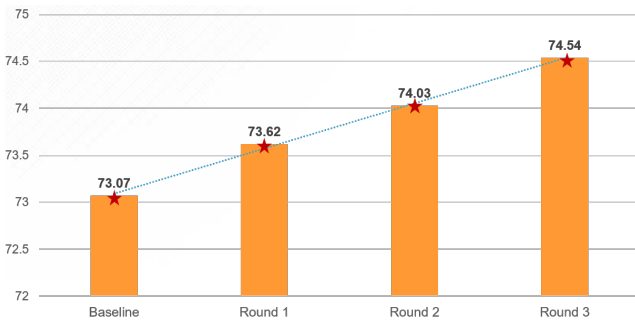


Fig. 7. The experimental results on the test leaderboard.

overall performance improves. After the final round, we combine these advanced models into an ensemble to obtain the highest F1 score of 80.41%.

As for the test leaderboard, the best model in validation experiments is utilized as the baseline and achieves an F1 score of 73.07% (Fig. 7). Consistent with the validation phase, detection accuracy is progressively improved as the number of rounds increase. Finally, the best model obtains the highest F1 score of 74.54%.

V. SECOND-PLACE TEAM

The network structure we propose for the Landslide4Sense competition is shown in Fig. 8. The details of each component of this proposed structure are discussed in detail below.

A. Framework Introduction

The main framework of our model is the encoder-decoder network, which uses an U-Net-like [30] skip connection structure and can better integrate shallow and deep features. Influenced by the rapid development of transformer-based models in the field of computer vision [34], [35], we introduce Swin Transformer [34] as the encoder part in this structure. To enable the Swin Transformer to reasonably capture the associations between landslide regions on multi-spectral data, we performed spectral selection experiments to use the spectra suitable for the self-attention mechanism. Subsequently, in order to alleviate the imbalance problem of positive and negative samples in landslide detection, we design an unbalanced

TABLE III
SPECTRAL SELECTION EXPERIMENTS ON THE VALIDATION SET

Input spectral bands	Input bands	F1 Score (%)		
		Swin Transformer	Deeplabv3	U-Net
RGB	3	65.6	58.0	59.2
SWIR	3	55.6	50.2	52.1
NGB	3	60.8	59.2	58.9
PCA [44]	3	49.5	46.8	52.4
RGB + NIR	4	63.3	57.2	59.4
RGB + SWIR	6	58.2	55.9	59.8
RGB + NIR + SWIR	7	54.8	57.5	60.0
All bands	14	55.8	57.8	61.1

Note: In this table, the RGB denotes the red, green, and blue spectra. SWIR denotes the 3-band far infrared in Sentinel-2. NGB denotes the near-infrared, green, and blue spectra. NIR denotes the near-infrared spectral. PCA refers to the techniques [44] of dimensionality reduction for compressing the original 14 bands into 3 bands.

training strategy that utilizes the unbalanced loss to first train compact feature representations, and then use the feature representations to fine-tune the classifier. Finally, we adopt a self-training strategy to further enhance the generalization of the model in the test domain.

1) *Spectral selection*: The Vision Transformer-based model performs feature aggregations using the self-attention mechanism to capture relations among pixels [34]. If irrelevant spectral information occupies dominant information, it will degrade the performance of the model. However, in the multi-spectral data, the responses of different spectra to the landslide area are quite different, and some spectra are even insensitive to landslides. Therefore, these “unspecific” spectra interfere with the execution of the self-attention mechanism. We performed spectral selection experiments, as shown in Table III and find that the fully convolutional model U-Net performs better with more spectral inputs, while the Transformer model works better when only the RGB spectrum is input. We further visualized the negative effect on self-attention when a spectrum insensitive to landslide responses was fed into the model, as shown in Fig. 9. Finally, we use the RGB spectra as the input to the model.

2) *Balanced training*: We design a two-stage training method to reduce the impact of the imbalance in the proportion of positive and negative samples. In the first stage, both the encoder and the decoder are trained simultaneously. For any input samples $x_i \in R^{w \times h \times 3}$, we use weighted cross-entropy loss \mathcal{L}_{wce} and Lovasz loss \mathcal{L}_{lov} [36] for balanced training as follows:

$$\arg \min_{E,D} \mathcal{L}_{wce} + \mathcal{L}_{lov} + \mathcal{L}_{ice}. \quad (1)$$

The \mathcal{L}_{ice} loss is the image-level loss performed in high-level semantic features in the encoder to assist training, which is defined as follows:

$$\mathcal{L}_{ice} = -\frac{1}{|\mathcal{X}|} \sum_{x_i \in \mathcal{X}} \delta(y_i) \log MP(E(x_i)), \quad (2)$$

where δ is a pointer function. If there is a positive sample (landslide) in y , the value of δ is 1; otherwise, its value is 0. $MP(\cdot)$ is a fully connected layer with a global pooling operation. \mathcal{X} denotes the total data set. Optimizing the \mathcal{L}_{ice} loss can increase the model’s attention to landslides, since the

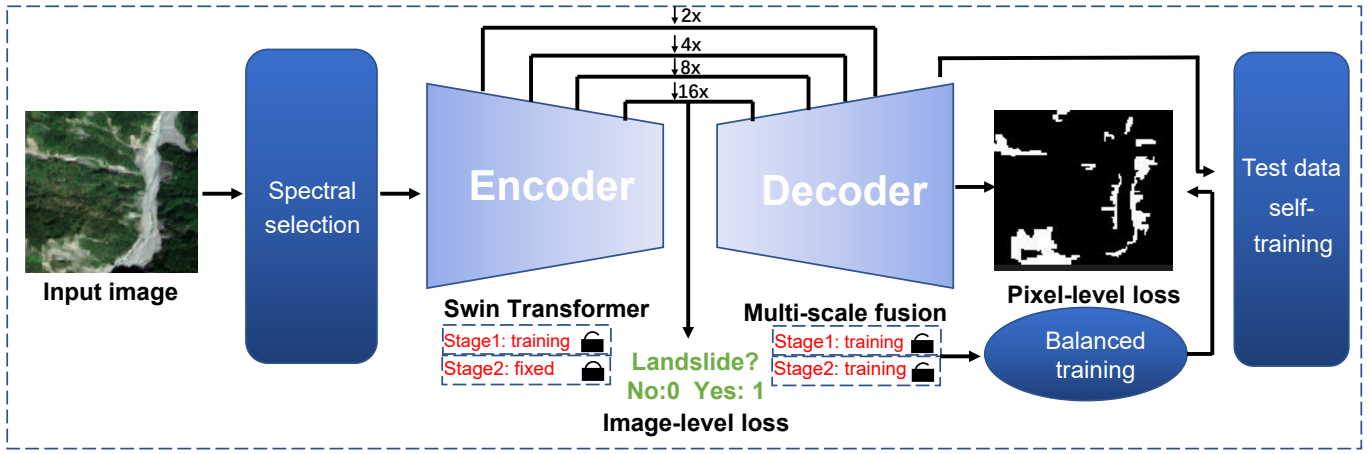


Fig. 8. Model structure for landslide detection proposed by Seek team.

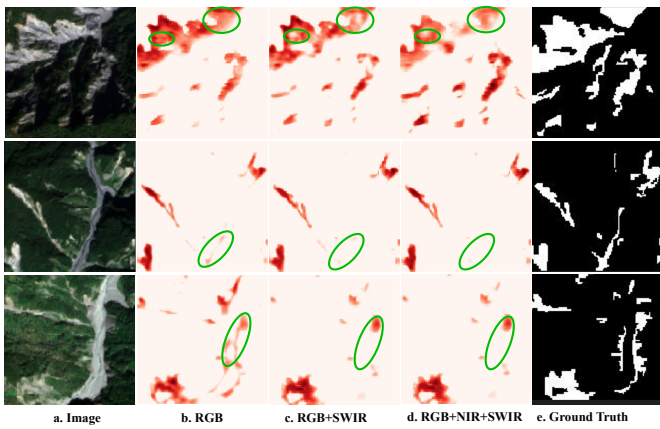


Fig. 9. Visualization of the feature activation map of the Swin Transformer when inputting different spectral bands.

task of finding a landslide in an image is much easier than finding where the landslide is. In order to reweight the learning of negative and positive samples, the \mathcal{L}_{wce} loss is defined as follows:

$$\mathcal{L}_{wce} = -\frac{1}{|\mathcal{X}|} \sum_{x_i \in \mathcal{X}} \frac{N_{neg}}{N_{pos}} y_i \log D(E(x_i)), \quad (3)$$

where N_{neg} denotes the number of negative samples (non-landslides) and N_{pos} denotes the number of positive samples (landslides) in any input image x .

As mentioned in [45], this re-weighting loss \mathcal{L}_{wce} plays a positive role in balancing the feature distribution of positive and negative samples. However, the classifier will still be biased. Therefore, in the second stage, we fix the trained encoder E and use the standard cross-entropy loss \mathcal{L}_{ce} to train the decoder D :

$$\arg \min_D \mathcal{L}_{ce} + \mathcal{L}_{ice}. \quad (4)$$

Once we have balanced feature representations, they can be further exploited to de-bias the classifier.

3) *Test data self-training*: Remote sensing imaging often faces the problem of data distribution shifts due to differences in geography and sampling time. To fully adapt the model

TABLE IV
BALANCE TRAINING EXPERIMENTS ON THE VALIDATION SET

Training methods	F1 Score (%)	
	Swin Transformer	U-Net
Normal training	69.8	63.7
Weighted cross-entropy	70.8	64.9
Focal loss [46]	68.2	61.8
Lovasz loss [36]	72.3	66.4
Balanced training	73.9	67.7

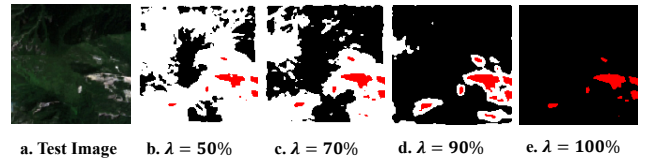


Fig. 10. Visualization of pseudo-labels obtained with different λ values. The black area is the selected negative sample area (landslide), the red area is the positive sample area (non-landslide), and the white area is the filtered-out area.

to the distribution of the test data, we adopt a self-training strategy [37] for enhancing the generalization of the model. We sort the output probabilities predicted in the previous stage, select the top $\lambda\%$ high-confidence pixel-level pseudo-labels, and add them to the training data for self-training.

B. Experimental Results

In this subsection, we report the performance of the balanced training and self-training methods.

Table IV shows that the two-stage balanced training method better attenuates the influence of the imbalance problem than focal loss [46] and other common methods [36]. Among them, normal training is a one-stage training method using cross entropy loss. For weighted cross entropy loss, we use the scale coefficients of positive samples and negative samples as the loss weighting coefficient of negative samples. This method has achieved a certain improvement by weighting the positive and negative pixels, but the improvement is relatively limited. Focal loss [46] balances easy and hard samples by

TABLE V
SELF-TRAINING EXPERIMENTS WITH DIFFERENT λ VALUES ON THE VALIDATION SET

λ	Precision (%)	Recall (%)	F1 Score (%)
- (Before ST)	73.4	74.7	73.9
50%	65.2	80.5	72.7
70%	69.3	79.5	73.7
90%	72.4	77.1	74.9
100%	78.2	74.2	76.1

Note: ST denotes self-training.

TABLE VI
THE EXPERIMENTS OF THE PROPOSED METHOD ON THE TEST SET

Training methods	F1 Score (%)
Balanced training (two-stage)	71.89
Self-training ($\lambda = 100$)	73.99

modifying their gradients for back propagation, and is also used in many unbalanced scenarios. But on this task, the performance degrades when this loss is added. Our analysis is that it has a great influence on the gradient, and inappropriate hyperparameters will greatly affect the performance. Lovasz loss [47] is a loss that directly optimizes the IoU coefficients, which is efficient and used as the first stage loss for our balanced training. Based on this, the two-stage balanced training method achieves the better performance, which further corrects the bias of the classifier. Finally, the adopted two-stage balanced training improves the F1 score by 4.1% on the basis of baseline.

Fig. 10 shows how the setting of λ affects pseudo-label selection. It can be observed that as lambda increases, more samples are selected, including both high- and low-confidence samples. Table V shows that the proposed self-training method can enhance the performance of the model, and can also balance precision and recall by adjusting the value of λ . When λ is small, the pseudo-labels corresponding to high-confidence regions is picked. Since the confusing area is easy to appear in the adjacent area of the positive (landslide) and negative (none-landslide) samples, the high-confidence samples mostly appear in the central area. With such pseudo-labels, we found that the model predicts more regions as positive samples, reducing missed detections and increasing recall. However, the boundary area is more blurred, resulting in increased over-detection. As λ increases, more pseudo-labels are picked, including uncertain samples. This part of the sample provides more valuable information about the confusing area, which makes the model not too biased toward the positive sample area. Although these samples contain noise, deep models can reduce the effect of noise by fine-tuning and stopping training early [48]. When λ is set to 100, all pseudo-labels are used for self-training, and experiments show that it achieves a better balance between precision and recall. In conclusion, we think that different λ values can be used in different scenarios.

Finally, the proposed method achieved the 2nd place in the test leaderboard with an F1 score of 73.99% as shown in Table VI, and it differs from the 1st place of 74.54% by only

0.55%. In summary, the transformer-based solution we use can effectively detect landslide areas in multi-spectral remote sensing scenes. In the future, our team argues that adaptive spectral selection or fusion technology is a necessary way to explore the performance of this transformer model further, and will become a follow-up research focus of our team.

VI. THIRD-PLACE TEAM

The solution of the third-place team is illustrated in Fig. 11. The methodology is detailed in the following sections.

A. Problem Formulation

Technically, the landslide detection problem can be formulated as a binary semantic segmentation problem. The training, validation, and test data set can be denoted by $\mathcal{D}_{train} = \{x_{tr}, y_{tr}\}$, $\mathcal{D}_{val} = \{x_{val}\}$, and $\mathcal{D}_{test} = \{x_{te}\}$, where x_{tr} , y_{tr} , x_{val} , and $x_{te} \in \mathbb{R}^{H \times W}$ correspond to the training patch, training label, validation patch, and test patch, respectively. Here H and W denote the data sets' spatial size. The goal of the landslide detection task is to train a semantic segmentation model on \mathcal{D}_{train} and \mathcal{D}_{val} , so that the best performance can be achieved on \mathcal{D}_{test} . Since the data are collected from different regions across the world, improving the exploitation of the unlabeled validation data can be beneficial to mitigate the domain gap between all the labeled and unlabeled data. To this end, we propose to incorporate a mixed supervised loss \mathcal{L}_{sup}^{mix} and a mixed pseudo-label loss \mathcal{L}_{pse}^{mix} to train the network:

$$\mathcal{L} = \mathcal{L}_{sup}^{mix} + \mathcal{L}_{pse}^{mix}. \quad (5)$$

The detailed formulation of \mathcal{L}_{sup}^{mix} and \mathcal{L}_{pse}^{mix} will be given in Section VI-D.

B. Supervised Losses

A combination of the cross-entropy loss \mathcal{L}_{cet} and the Jaccard loss \mathcal{L}_{jac} [49] are used as the supervised losses:

$$\mathcal{L}_{sup} = \mathcal{L}_{cet}(\mathcal{M}_s(x_{tr}), y_{tr}) + \mathcal{L}_{jac}(\mathcal{M}_s(x_{tr}), y_{tr}). \quad (6)$$

Here $\mathcal{M}_s(\cdot)$ denotes the mapping function defined by the student model \mathcal{M}_s .

C. Self-training

The authors propose a self-training strategy [37] to exploit the unlabeled data. First, the teacher model \mathcal{M}_t will be trained solely on the training data \mathcal{D}_{train} . Then it will be used to generate pseudo labels on the unlabeled data \mathcal{D}_{val} to supervise the student model \mathcal{M}_s .

However, the raw predictions from \mathcal{M}_t are likely to be incorrect. To prevent the student model from overfitting to those wrong predictions, a pseudo-label selection strategy is needed to filter out misclassified pixels.

To achieve this, the Monte Carlo dropout strategy [50] is first used to generate an uncertainty map for each unlabeled image patch. More specifically, the unlabeled validation patch x_{val} is input to the teacher model \mathcal{M}_t for 10 different runs. During each run, a dropout layer with 0.3 dropping rate is

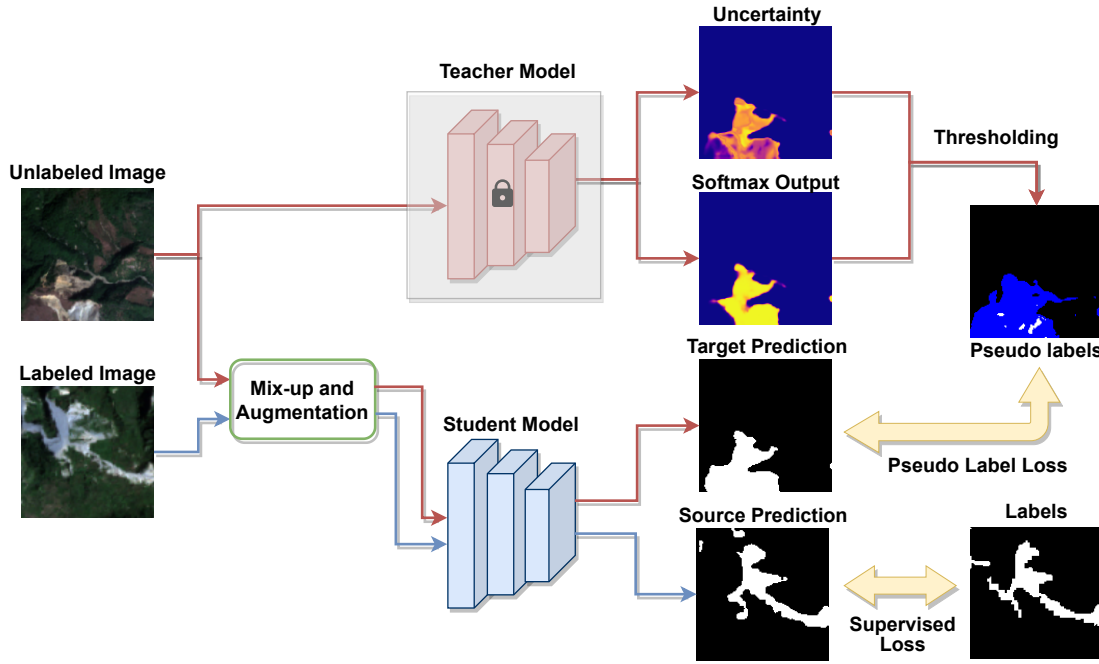


Fig. 11. Network architecture of the landslide detection method proposed by *Tammh* team. The overall architecture follows a self-training scheme, which consists of a teacher model branch and a student model branch. For the teacher model branch, a teacher model pre-trained on the training data is applied to generate pseudo labels based on the unlabeled images, which will later be used to supervise the training of the student model. For the student model branch, both the labeled and the unlabeled images are input to the student model after some data augmentation and mix-up operations. The training losses are then calculated based on both the training labels and the pseudo labels. During the training phase, the teacher model is fixed.

applied after the first convolution layer to disturb the network. The variances of 10 different outputs are then calculated as the uncertainty map.

Next, the uncertainty map is used to mask out those uncertain predictions from the teacher model \mathcal{M}_t . Inspired by class-balanced self-training (CBST) [37], the selection process is conducted in a class-wise manner, which means the top 90% of the background pixels and top 70% of the landslide pixels with the lowest uncertainty will be selected as the pseudo labels. Meanwhile, the other predictions with higher uncertainty will be ignored when calculating the losses.

To this end, the pseudo label loss \mathcal{L}_{pse} can be formulated by:

$$\mathcal{L}_{pse} = \mathcal{L}_{cet}(\mathcal{M}_s(x_{te}), \hat{y}_{te}) + \mathcal{L}_{jac}(\mathcal{M}_s(x_{te}), \hat{y}_{te}). \quad (7)$$

Here \hat{y}_{te} corresponds to the pseudo labels of x_{te} generated by the teacher model and followed by the pseudo label selection process.

D. Mix-up Strategy

To prevent overfitting and further improve the generalizability of the landslide detection model, a mix-up strategy [38] is applied to both the labeled and the unlabeled data. Given a batch of the training data x_{tr} and the validation data x_{val} , the mixed data can be achieved by their linear mixing:

$$\begin{aligned} \tilde{x}_{tr} &= \lambda x_{tr}^i + (1 - \lambda)x_{tr}^j, \\ \tilde{x}_{val} &= \lambda x_{val}^i + (1 - \lambda)x_{val}^j. \end{aligned} \quad (8)$$

Here x^i and x^j are two image patches from the corresponding data set, and λ is the mixing coefficient randomly sampled

from a beta distribution during each training step. After applying the mix-up strategy, the supervised and pseudo label losses can be reformulated as

$$\begin{aligned} \mathcal{L}_{sup}^{mix} &= \lambda \mathcal{L}_{sup}(\tilde{x}_{tr}, y_{tr}^i) + (1 - \lambda)\mathcal{L}_{sup}(\tilde{x}_{tr}, y_{tr}^j), \\ \mathcal{L}_{pse}^{mix} &= \lambda \mathcal{L}_{pse}(\tilde{x}_{te}, \hat{y}_{te}^i) + (1 - \lambda)\mathcal{L}_{pse}(\tilde{x}_{te}, \hat{y}_{te}^j). \end{aligned} \quad (9)$$

By training on mixed images, the model will be less likely to be overconfident about its predictions, and hence better generalize to the unseen data.

E. Post-processing

The dense conditional random field (DenseCRF) [39] technique is applied to the model's output as post-processing. This step helps to better match the predicted landslide contours with the ground truths. Finally, the best model obtains the highest F1 score of 73.5%.

VII. SPECIAL PRIZE TEAM

Landslide4Sense provides data with 14 bands, while most deep learning semantic segmentation models, such as [30], [33], [51], [52], require an RGB image as the input. This means we cannot utilize pre-trained weights to improve the model performance and shorten training time. On the Landslide4Sense data set, we try three types of models, U-Net, Deeplabv3, and Deeplabv3+, but none of these models yields a very high performance, with F1 scores of only 65%, 66%, and 67%, respectively. So, we explore the use of multi-spectral satellite imagery for the deep learning-based landslide segmentation task.

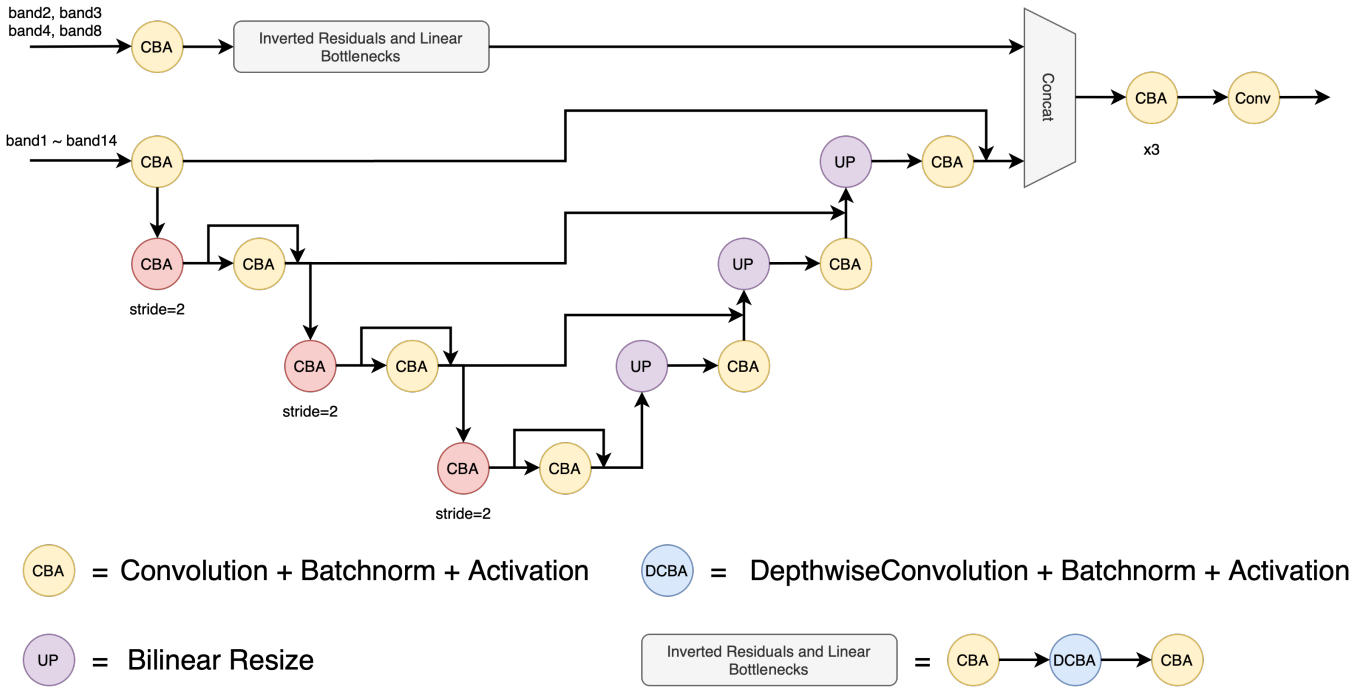


Fig. 12. The structure of the Multi-spectral U-Net. The inputs of the upper branch consist of band2, band3, band4, and band8, which have 10 meters resolution, and the inputs of the lower branch mix all the 14 bands, including 10, 20, and 60 meters resolutions. Specifically, we implement the downsampling layer by using the convolution with stride two and using the bilinear resize for the upsampling layer, and all the skip connection operations are additive. CBA means the sequential block of convolution, batch norm, and activation.

A. Multi-spectral U-Net

Considering the different resolution bands of the imagery in the Landslide4Sense dataset, we introduce a novel model called Multi-spectral U-Net, which has two input branches for the different resolution inputs. The model structure is illustrated in Fig. 12. Multi-spectral U-Net comprises two branches, the High Resolution Branch (upper part) and the General Resolution Branch (lower part), whose features will be merged, and then contribute jointly to the final segmentation prediction.

The High Resolution Branch was used for the data with high resolution, which can yield refined feature maps containing more marginal information. Specifically, we implement this branch by using the Inverted Residuals and Linear Bottlenecks introduced in the MobileNetV2 [53], and consisting of two point-wise convolution layers and one depth-wise convolution layer. To avoid a dramatic increase in the dimensions of the feature maps, from 4 to 128 dimensions, we first apply two simple convolution layers. The feature dimensions will expand then recovered to the original dimension after the depth-wise convolution layer. Additionally, there is no downsampling layer in the branch, as the only aim of this branch is to extract additional marginal information in order to get a better segmentation prediction.

In the General Resolution Branch, we apply some modifications to the original U-Net [30]. U-Net is an expandable segmentation model that has a symmetrical architecture; this kind of architecture has been widely used for other segmentation tasks. It is very convenient to replace some implementations of the U-Net, which is the main reason we chose it for our

model. The specific modifications are as follows. First, we reduce the number of the downsampling layers due to the limited size (128×128 pixels) of the input image. To ensure the smallest feature size is at least 16×16 , we use only three downsampling operations in the U-Net model. Secondly, skip-connection introduced in the ResNet is widely used in the model to mitigate the vanishing gradient problem. Finally, we update the activation function to SMU [54], which can improve model performance without performance loss on inference speed, as shown in Eqs. (10), and (11).

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (10)$$

$$\operatorname{SMU}(x) = \frac{x(1 + \alpha) + x(1 - \alpha)\operatorname{erf}(\mu(1 - \alpha)x)}{2}. \quad (11)$$

In the Multi-spectral U-Net model, we input all 14 bands to the General Resolution Branch and only 10 meters resolution bands (band2, band3, band4, and band8) to the High Resolution Branch. In order to balance the feature dimensions of two branches, we make the High Resolution Branch and the General Resolution Branch have the same output shape, $128 \times 128 \times 128$. The features from two branches will be concatenated to a feature map in the shape of $128 \times 128 \times 256$, which is used for the final pixel-level prediction.

B. Experiments

We trained the model with NVIDIA GeForce RTX 3090 GPU and Intel(R) Core(TM) i7-7800X CPU @3.50GHz. To compare the performance of the three models more clearly, we use a batch size of 8, the Adam optimizer, warmup, and

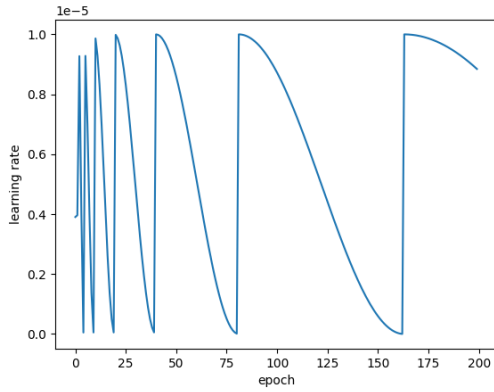


Fig. 13. The cosine learning rate with warmup and restart schedule. In detail the parameters are: warmup epochs 2, restart multiplier 2, init learning rate 0.00001, and minimum learning rate 0.

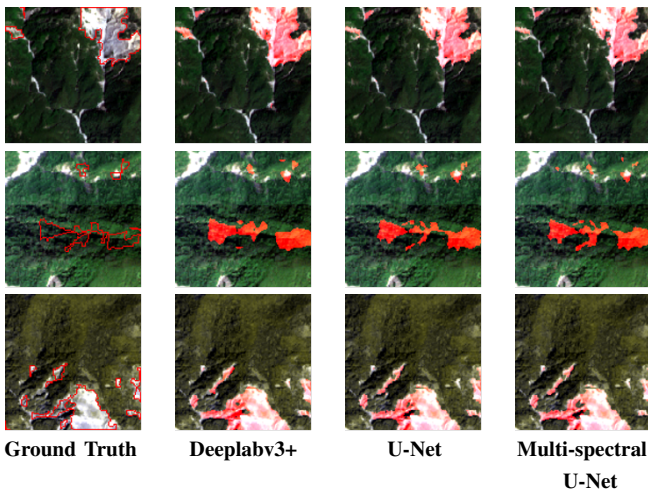


Fig. 14. The landslide segmentation results for some representative data in the validation dataset.

restarted cosine learning rate (shown in Fig. 13) and the cross-entropy loss.

We split the official training dataset into two parts, with 3539 images for training and 260 images for testing. Then, we compared the Multi-spectral U-Net performance with Deeplabv3+ and U-Net on it, after each model is trained with 200 epochs. In terms of recall and F1 score, Multi-spectral U-Net is significantly higher than the other two models, but its precision is lower than that of U-Net. Significantly, the precision of Deeplabv3+ is dramatically lower than Multi-spectral U-Net and U-Net, and we think the potential reason is that a large number of downsampling layers lead to the loss of marginal information.

For a better understanding of the different models' performance on the validation dataset, we analyze the prediction segmentation results and choose three representative examples in Fig. 14. In the first example, the landslide segmentation results of the three models are similar, and it is clear that the Deeplabv3+ tends to predict a wide range of area but does not have very refined edge information about the landslide. In

other words, this may be the reason that the Deeplabv3+ has a higher recall than U-Net, but the precision is significantly lower. We can not directly see the landslide from the image in the second example; however, all three models can predict the landslide very well, which means all the bands besides the RGB bands (band2–band4) also contribute to the final prediction. The third example is a very complex landslide scenario, in which we can clearly view the superior performance of the Multi-spectral U-Net.

In the test phase of the LandSlide4Sense competition, we use the well-trained model to predict the validation data set first and get annotations from the prediction result. Then, we get a new training data set by combining the annotated validation data set and the old training data set. Finally, after training the Multi-spectral U-Net on the new training data set, we get an F1 score of 71.29% in the test set.

VIII. CONCLUSIONS

In recent decades, remote sensing techniques have been predominantly used for natural hazard-related applications, i.e., landslide detection. There are many advantages to using Earth observation and remote sensing products in these applications, but the most critical one is their timeliness and objectivity. Early detection is vital for a rapid response and effective management of the consequences of a landslide event. Due to the increasing number and quality of space-borne sensors, the remote sensing community has recently had access to high-quality images with a higher spatial-temporal resolution. In light of the improved availability of data, attention has turned towards the methodologies for retrieving information and knowledge from the data itself [55]. Therefore, there has been a great desire to replace the use of experts' knowledge-based physical methodologies with automatic interpretation methods of remote sensing images.

Although promising results have been obtained by DL models for a wide range of remote sensing applications, the need for solutions to landslide detection challenges such as extracting landslides from remote sensing data has only been brought to the attention of the machine learning and computer vision communities in recent years. The solutions, however, have only been implemented at the local level and have followed a common procedure that includes training the DL model using an annotated data set of landslides covering a relatively small area [12], [21]. The local level is taken into consideration for several reasons related to how model generalization handle high-level issues, such as the impacts caused by different triggers, the types of mass movements, and the geology and morphology of the region, as well as the source of inventory data sets and the method in which they were developed. The landslide inventory data sets that are used for training modern DL models are usually created based on manual or knowledge-based physical semi-automated methods. Thus, implementing such methods for semantic annotations and creating inventory data sets at a large scale is generally a tedious and expensive process. In preparing a precise inventory of landslides, an even greater amount of work is required since it involves not only the analysis of one image but also a comparison of

two images from the pre- and post-event for each case study area. Therefore, it is very unlikely that landslide inventory data sets with highly accurate annotations can be found on a large scale. As a result of the lack of these data sets, serious concerns about the performance of currently available landslide detection DL solutions are warranted, particularly, when applied directly to a new case study area that has not yet been investigated. To address all the above-mentioned issues, the L4S competition has been organized by the IARAI and provides a globally distributed landslide inventory data set. The competition promotes development and demonstration of innovative algorithms for automatic landslide detection using remote sensing images throughout the world, as well as providing fair and objective comparisons of different DL solutions for automatic landslide detection.

This paper presents a summary of the top winners of the 2022 L4S competition. The competition was dedicated to developing DL solutions for solving unsolved challenges in the detection of landslides using remote sensing images collected from various regions around the world. Different strategies and algorithms were brought to light by our winning teams. The first-ranked team identified three main challenges: a large number of small landslides and the huge class imbalance between landslides and non-landslides, as well as the distribution inconsistency of the landslides in the study areas and, consequently, the image patches. In addressing these challenges, they conducted a series of experiments to obtain the competition's highest F1 score value of 73.07%. For the weak representation of small landslides, they applied a scale promotion of original image patches from 128×128 pixels to 512×512 pixels. This team integrated three models of Swin Transformer, EfficientNetV2, and SegFormer by emphasizing self-attention operations. Further landslide detection improvements were effected by the second-place team using the Swin Transformer as the encoder part and the self-attention mechanism. In addition, a self-training strategy was used to enhance the generalization of their proposed model on the competition's test data. To overcome the imbalance between landslide and non-landslide classes, the first-place team adopted and applied the Lovasz loss and online hard example mining strategy. An unbalanced approach to training, however, led to the second team's success. The third-place team proposed an integrated approach of a mixed supervised loss and a self-training consisting of pseudo labels and the Monte Carlo dropout strategy to train their network for landslide detection. Using DenseCRF, this team post-processed the network's outputs to improve the borders of landslides. The special prize team introduced a multi-spectral U-Net inspired by MobileNetV2 to handle the multi-spectral Sentinel-2 and ALOS PALSAR data for landslide detection provided by the competition. As part of the competition's test phase, they generated annotations for the validation data set using the well-trained model, and by adding new labeled data to the training data set, they trained their introduced U-Net. The DL solutions provided by the competition's four winners were presented by the corresponding authors at the CDCEO 2022 workshop as a satellite event at IJCAI-ECAI 2022, the 31st International Joint Conference on Artificial Intelligence, and the 25th European

Conference on Artificial Intelligence.

The data remain accessible after the L4S competition and the *Future Development Leaderboard* for future evaluation at <https://www.iarai.ac.at/landslide4sense/challenge/> is active to allow further research developments and contributions. In this way, anyone can submit landslide detection results on the test data set, make comparisons of their performance to that of other users, and, ideally, enhance the accuracy presented in this outcome paper. It is noteworthy that L4S was the first competition to be based on multi-source satellite imagery for landslide detection and had a significant impact on this field; furthermore, participants agree that the competition is also an extremely interesting challenge from a computer vision and machine learning perspective.

As the consequences of climate change pose an accelerating quantity and range of challenges to the world's scientists, they may not have sufficient time and resources to generate landslide inventory data sets based on fieldwork. Yet modern DL solutions, particularly those based upon such a large source of remote sensing data, must be able to cope with monitoring natural hazards and risk assessment. Therefore, developing innovative DL solutions and training them on a global data set will be crucial to generating timely information from remote sensing data for future landslide events. The L4S 2022 data provide a valuable benchmark data set for evaluating all new DL algorithms developed for landslide detection, and the algorithms developed as part of the L4S competitions will, it is hoped, inspire development of increasingly efficient and accurate algorithms.

ACKNOWLEDGMENTS

The authors would like to thank Alina Mihai, Miriam Vázquez, Gabriel Fratica, Christian Eichenberger, Toma Furdui, Gero Nikolov, Ezgi Özkan. Also, special thanks go to our Organising Committee: Pedro Herruzo, David Kreil, Michael Kopp, and Sepp Hochreiter; Scientific Committee: Cees van Westen, Martin Rutzinger, and Changlin Wang for their valuable help in the organization of Landslide4Sense 2022 competition, and the anonymous referees for their valuable comments/suggestions that have helped us improve an earlier version of the manuscript. This research was funded by the Institute of Advanced Research in Artificial Intelligence (IARAI) GmbH.

REFERENCES

- [1] F. Guzzetti, P. Reichenbach, F. Ardizzone, M. Cardinali, and M. Galli, "Estimating the quality of landslide susceptibility models," *Geomorphology*, vol. 81, no. 1-2, pp. 166–184, 2006.
- [2] D. J. Varnes, *Landslide Types and Processes*. Highway Research Board Special Report, 1958.
- [3] P. Lima, S. Steger, T. Glade, N. Tilch, L. Schwarz, and A. Kociu, "Landslide susceptibility mapping at national scale: A first attempt for Austria," in *Workshop on World Landslide Forum*. Springer, 2017, pp. 943–951.
- [4] S. M. Tayyebi, M. Pastor, M. M. Stickle, Á. Yagüe, D. Manzanal, M. Molinos, and P. Navas, "Two-phase SPH modelling of a real debris avalanche and analysis of its impact on bottom drainage screens," *Landslides*, vol. 19, no. 2, pp. 421–435, 2022.
- [5] J. Branke, T. Zieher, J. Pfeiffer, M. Bremer, M. Rutzinger, B. Gems, M. Keiler, and B. Schneider-Muntau, "Extending the integrated monitoring of deep-seated landslide activity into the past using free and open-source photogrammetry," Copernicus Meetings, Tech. Rep., 2022.

- [6] O. Ghorbanzadeh, Y. Xu, P. Ghamisi, M. Kopp, and D. Kreil, "Landslide4sense: Reference benchmark data and deep learning models for landslide detection," *arXiv preprint arXiv:2206.00515*, 2022.
- [7] B. van den Bout, C. Tang, C. van Westen, and V. Jetten, "Physically-based modelling of co-seismic landslide, debris flow and flood cascade," *Natural Hazards and Earth System Sciences Discussions*, pp. 1–56, 2022.
- [8] B. Yang, T. Xiao, L. Wang, and W. Huang, "Using complementary ensemble empirical mode decomposition and gated recurrent unit to predict landslide displacements in dam reservoir," *Sensors*, vol. 22, no. 4, p. 1320, 2022.
- [9] T. Liu, T. Chen, R. Niu, and A. Plaza, "Landslide detection mapping employing CNN, ResNet, and DenseNet in the three gorges reservoir, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11 417–11 428, 2021.
- [10] 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.
- [11] G. Titti, C. van Westen, L. Borgatti, A. Pasuto, and L. Lombardo, "When enough is really enough? On the minimum number of landslides to build reliable susceptibility models," *Geosciences*, vol. 11, no. 11, p. 469, 2021.
- [12] O. Ghorbanzadeh, A. Crivellari, P. Ghamisi, H. Shahabi, and T. Blaschke, "A comprehensive transferability evaluation of U-Net and ResU-Net for landslide detection from Sentinel-2 data (case study areas from Taiwan, China, and Japan)," *Scientific Reports*, vol. 11, no. 1, pp. 1–20, 2021.
- [13] T. Stanley and D. B. Kirschbaum, "A heuristic approach to global landslide susceptibility mapping," *Natural Hazards*, vol. 87, no. 1, pp. 145–164, 2017.
- [14] 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 Sensing*, vol. 11, no. 2, p. 196, 2019.
- [15] C. Ye, Y. Li, P. Cui, L. Liang, S. Pirasteh, J. Marcato, W. N. Gonçalves, and J. Li, "Landslide detection of hyperspectral remote sensing data based on deep learning with constrains," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5047–5060, 2019.
- [16] D. Hölbling, P. Füreder, F. Antolini, F. Cigna, N. Casagli, and S. Lang, "A semi-automated object-based approach for landslide detection validated by persistent scatterer interferometry measures and landslide inventories," *Remote Sensing*, vol. 4, no. 5, pp. 1310–1336, 2012.
- [17] S. Tavakkoli Piralilou, H. Shahabi, B. Jarihani, O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, and J. Aryal, "Landslide detection using multi-scale image segmentation and different machine learning models in the higher himalayas," *Remote Sensing*, vol. 11, no. 21, p. 2575, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/21/2575>
- [18] Y. Yi and W. Zhang, "A New Deep-Learning-Based Approach for Earthquake-Triggered Landslide Detection from Single-or RapidEye Satellite Imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6166–6176, 2020.
- [19] F. Catani, "Landslide detection by deep learning of non-nadir and crowdsourced optical images," *Landslides*, vol. 18, no. 3, pp. 1025–1044, 2021.
- [20] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 982–986, 2019.
- [21] N. Prakash, A. Manconi, and S. Loew, "A new strategy to map landslides with a generalized convolutional neural network," *Scientific Reports*, vol. 11, no. 1, pp. 1–15, 2021.
- [22] W. Qi, M. Wei, W. Yang, C. Xu, and C. Ma, "Automatic mapping of landslides by the ResU-net," *Remote Sensing*, vol. 12, no. 15, p. 2487, 2020.
- [23] S. L. Ullo, A. Mohan, A. Sebastianelli, S. E. Ahamed, B. Kumar, R. Dwivedi, and G. R. Sinha, "A new mask R-CNN-based method for improved landslide detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3799–3810, 2021.
- [24] A. C. Mondini, "Measures of spatial autocorrelation changes in multitemporal sar images for event landslides detection," *Remote Sensing*, vol. 9, no. 6, p. 554, 2017.
- [25] W. Shi, M. Zhang, H. Ke, X. Fang, Z. Zhan, and S. Chen, "Landslide recognition by deep convolutional neural network and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4654–4672, 2020.
- [26] H. Tanyaş, C. J. van Westen, K. E. Allstadt, and R. W. Jibson, "Factors controlling landslide frequency–area distributions," *Earth surface processes and landforms*, vol. 44, no. 4, pp. 900–917, 2019.
- [27] O. Ghorbanzadeh, K. Gholamnia, and P. Ghamisi, "The application of ResU-net and OBIA for landslide detection from multi-temporal sentinel-2 images," *Big Earth Data*, pp. 1–26, feb 2022. [Online]. Available: <https://doi.org/10.1080/20964471.2022.2031544>
- [28] O. Ghorbanzadeh, H. Shahabi, A. Crivellari, S. Homayouni, T. Blaschke, and P. Ghamisi, "Landslide detection using deep learning and object-based image analysis," *Landslides*, pp. 1–11, 2022.
- [29] S. Zhang, R. Li, F. Wang, and A. Iio, "Characteristics of landslides triggered by the 2018 Hokkaido Eastern Ibari earthquake, Northern Japan," *Landslides*, vol. 16, no. 9, pp. 1691–1708, 2019.
- [30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
- [31] P. Liu, Y. Wei, Q. Wang, Y. Chen, and J. Xie, "Research on post-earthquake landslide extraction algorithm based on improved U-Net model," *Remote Sensing*, vol. 12, no. 5, p. 894, 2020.
- [32] X. Tang, Z. Tu, Y. Wang, M. Liu, D. Li, and X. Fan, "Automatic detection of coseismic landslides using a new transformer method," *Remote Sensing*, vol. 14, no. 12, p. 2884, 2022.
- [33] 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.*, 2018, pp. 801–818.
- [34] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 10 012–10 022.
- [35] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Proc. Neural Inf. Process. Syst.*, vol. 34, 2021.
- [36] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4413–4421.
- [37] Y. Zou, Z. Yu, B. Kumar, and J. Wang, "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 289–305.
- [38] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.
- [39] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with Gaussian edge potentials," *Proc. Neural Inf. Process. Syst.*, vol. 24, 2011.
- [40] H. Zhao, Y. Zhong, X. Wang, X. Hu, C. Luo, M. Boitt, R. Piironen, L. Zhang, J. Heiskanen, and P. Pellikka, "Mapping the distribution of invasive tree species using deep one-class classification in the tropical montane landscape of Kenya," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 328–344, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271622000715>
- [41] J. Wang, A. Ma, Y. Zhong, Z. Zheng, and L. Zhang, "Cross-sensor domain adaptation for high spatial resolution urban land-cover mapping: From airborne to spaceborne imagery," *Remote Sens. Environ.*, vol. 277, p. 113058, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425722001729>
- [42] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2021, pp. 10 096–10 106.
- [43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *arXiv e-prints*, p. arXiv:1512.00567, Dec. 2015.
- [44] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2001.
- [45] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9719–9728.
- [46] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [47] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4413–4421.
- [48] Q. Wang, W. Li, and L. V. Gool, "Semi-supervised learning by augmented distribution alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, October 2019.

- [49] M. Polak, H. Zhang, and M. Pi, "An evaluation metric for image segmentation of multiple objects," *Image and Vision Computing*, vol. 27, no. 8, pp. 1223–1227, 2009.
- [50] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2016, pp. 1050–1059.
- [51] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [52] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv*, vol. abs/1706.05587, 2017.
- [53] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
- [54] K. Biswas, S. Kumar, S. Banerjee, and A. K. Pandey, "SMU: Smooth activation function for deep networks using smoothing maximum technique," *arXiv*, vol. abs/2111.04682, 2021.
- [55] S. Lang, L. Wendt, D. Tiede, Y. Gao, V. Streifender, H. Zafar, A. Adebayo, G. Schwendemann, and P. Jeremias, "Multi-feature sample database for enhancing deep learning tasks in operational humanitarian applications," *GI_Forum*, vol. 9, no. 1, pp. 209–219, 2021.



Omid Ghorbanzadeh works as a machine learning researcher (postdoctoral) in the AI4RS group at the Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria. He received the Ph.D. degree in applied Geoinformatics from University of Salzburg, Austria, in 2021.

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. He is awarded best papers at the GISTAM 2019 Conf., and Journals of Mathematics 2020, Spatial Science 2021, and Remote Sensing 2021.



Yonghao Xu (S'16, M'21) received the B.S. degree and the Ph.D. degree in Photogrammetry and Remote Sensing from Wuhan University, Wuhan, China, in 2016 and 2021, respectively. He is currently a post-doctoral researcher at the Institute of Advanced Research in Artificial Intelligence (IARAI), Austria. His research interests include remote sensing, computer vision, and machine learning.



Hengwei Zhao received the B.S. degree in surveying and mapping engineering from the School of Resources and Civil Engineering, Northeastern University, ShenYang, China, in 2019. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan. His major research interests are weakly supervised remote sensing image processing and computer vision.



Junjue Wang received the B.S. degree from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, in 2019. He is pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan. His major research interests are remote sensing imagery semantic segmentation and relational reasoning.



Yanfei Zhong received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2002 and 2007, respectively. Since 2010, He has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China. Dr. Zhong is a Fellow of the Institution of Engineering and Technology (IET). He organized the Intelligent Data Extraction, Analysis and Applications of Remote Sensing (RSIDEA)

research group. His research interests include hyperspectral remote sensing information processing, high-resolution remote sensing image understanding, and geoscience interpretation for multisource remote sensing data and applications. For detailed info, see <http://rsidea.whu.edu.cn/e-index.html>.



Dong Zhao received the B.S. degree in electronic information engineering from the University of Science and Technology, Liaoning, China, in 2018. He is currently pursuing the Ph.D. degree in computer science with the School of Electronic Engineering, Xidian University, Xi'an, China. His research interests include semantic segmentation and domain adaptation.



Qi Zang received the B.S. degree in electronic information engineering from the Northwest Normal University, Lanzhou, China, in 2019. She is currently pursuing the Ph.D. degree in computer science and Technology with the School of Artificial Intelligence, Xidian University, Xi'an, China. Her research interests include semantic segmentation, unsupervised domain adaptation and domain generalization.



Shuang Wang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in circuits and systems from Xidian University, Xi'an, China, in 2000, 2003, and 2007, respectively. She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. Her research interests include sparse representation, image processing, synthetic aperture radar (SAR) automatic target recognition, remote sensing image captioning, and polarimetric SAR data analysis and interpretation.



Fahong Zhang received the B.E. degree in software engineering from Northwestern Polytechnical University, Xi'an, China, in 2017, the M.S. degree in computer science with the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. He is now pursuing the Ph.D. degree with the department of Aerospace and Geodesy, Data Science in Earth Observation, Technical University of Munich. His research interests include computer vision and satellite image processing.



Lin Bai Lin Bai works as a researcher in the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu (CDUT), China. He received the Ph.D. degree in Earth Exploration and Information Technology from CDUT in 2022. He has done much research on the application of artificial intelligence in the earth sciences.



Yilei Shi (Member, IEEE) received the Dipl.-Ing. degree in mechanical engineering and the Dr.-Ing. degree in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2010 and 2019, respectively. He is a Senior Scientist with the Chair of Remote Sensing Technology, TUM. His research interests include fast solver and parallel computing for large-scale problems, high-performance computing and computational intelligence, advanced methods on synthetic aperture radar (SAR) and InSAR processing, machine learning, and

deep learning for variety of data sources, such as SAR, optical images, and medical images, and partial differential equation (PDE)-related numerical modeling and computing.



Weile Li received the B.S. degree in surveying and mapping engineering, the M.S. degree in GIS RS, and the Ph.D. degree in engineering geology from Chengdu University of Technology, China, in 2005, 2008 and 2019, respectively. Since 2021, He has been a Full Professor with the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (SKLGP), Chengdu University of Technology, China. His major research interests are landslide detection and monitoring.



Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.-Ing., and Habilitation degrees in the field of signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is currently a Professor for Data Science in Earth Observation (former: Signal Processing in Earth Observation) with TUM, and the Head of the Department "EO Data Science," Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany. Since 2019, she has

been a Co-Coordinator of the Munich Data Science Research School, Munich, (www.mu-ds.de). Since 2019, she has been heads the Helmholtz Artificial Intelligence-Research Field "Aeronautics, Space and Transport." Since May 2020, she has been the Director of the International Future AI Laboratory "Artificial Intelligence for Earth Observation (AI4EO): Reasoning, Uncertainties, Ethics and Beyond," Munich. Since October 2020, she has been serves as a Co-Director of the Munich Data Science Institute (MDSI), TUM. She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, The University of Tokyo, Tokyo, Japan, and University of California, Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently a Visiting AI Professor with the ESA'S Phi-Laboratory. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with their applications in tackling societal grand challenges, e.g., global urbanization, United Nations (UN) sustainable development goals (SDGs), and climate change.

Dr. Zhu is a member of Young Academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She serves in the scientific advisory board in several research organizations, among others the German Research Center for Geosciences (GFZ) and Potsdam Institute for Climate Impact Research (PIK). She is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and serves as the area editor responsible for special issues of the IEEE Signal Processing Magazine.



Weihang Peng received the B.S. degree in computer science and technology from Chengdu University of Technology (CDUT), China, in 2020. He is currently working as research assistant in State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (SKLGP). His major research interests include object detection and point cloud segmentation on deep learning and the AI application in the landslide detection.



Pedram Ghamisi (S'12, M'15, SM'18) graduated with a Ph.D. in electrical and computer engineering at the University of Iceland in 2015. He works as (1) the head of the machine learning group at Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany and (2) visiting research professor and group leader of AI4RS at the Institute of Advanced Research in Artificial Intelligence (IARAI), Austria. He is a co-founder of VasoGnosis Inc. with two branches in San Jose and Milwaukee, the USA. His research interests include deep learning, with

a sharp focus on remote sensing applications. For detailed info, see <http://www.ai4rs.com/>.