Lightweight Attention-Guided YOLO With Level Set Layer for Landslide Detection From Optical Satellite Images

Yueheng Yang, Zelang Miao, *Member, IEEE*, Hua Zhang, Bing Wang, *Member, IEEE*, and Lixin Wu, *Member, IEEE*

***Abstract*—Landslide inventory is significant for landslide disaster reduction. To construct the landslide inventory, deep learning has received growing attention to detect landslides from satellite images. Among various deep learning algorithms, you-only-look-once (YOLO) has a strong ability to detect objects efficiently and has been widely used in landslide extraction. Despite its efficiency, there is no general rule to select the backbone and attention mechanism for YOLO. The selection of these two modules depends on specific application needs. Meanwhile, YOLO output is a series of anchor boxes, not accurate landslide boundaries. A single bounding box may contain many landslides and cannot extract individual landslides, limiting the YOLO applications in constructing landslide inventory. To address these issues, this article presents a lightweight attention-guided YOLO with level set layer (LA-YOLO-LLL) for landslide detection from optical satellite images. First, we introduced the MobileNetv3 to replace the original backbone of YOLO to simultaneously reduce the parameter complexity and improve the model transferability. Then, we presented a light pyramid features reuse fusion attention mechanism to improve landslide detection performance. Finally, we integrated the level set layer into YOLO head to produce accurate landslide boundaries. This article validated the accuracy and transferability of the presented method in two study areas (Bijie and Taiwan) with similar geo-environmental conditions. Experimental results show that the presented LA-YOLO-LLL model outperformed traditional YOLO in landslide detection. Findings in this article are valuable for landslide inventory construction, land use planning and risk control.**

***Index Terms*—Attention mechanism, landslide boundary, landslide inventory, object detection, the level set layer, you-only-look-once (YOLO).**

Manuscript received 24 May 2023; revised 19 August 2023 and 3 November 2023; accepted 2 January 2024. Date of publication 9 January 2024; date of current version 24 January 2024. This work was supported by the National Natural Science Foundation of China under Grant 42171084, Grant 42071256, and Grant 72088101, and in part by the Natural Science Foundation of Hunan Province under Grant 2022JJ30704. *(Corresponding author: Zelang Miao.)*

Yueheng Yang, Zelang Miao, and Lixin Wu are with the School of Geoscience and Info-Physics and Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, Central South University, Changsha 410083, China (e-mail: 225012153@csu.edu.cn; zelang.miao@outlook.com; wulx66@csu.edu.cn).

Bing Wang is with the School of Geosciences & Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China (e-mail: wangbing@student.cumtb.edu.cn).

Hua Zhang is with the School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China (e-mail: zhhua_79@163.com).

Digital Object Identifier 10.1109/JSTARS.2024.3351277

I. Introduction

LANDSLIDE, mainly triggered by rainfall and earthquake, is a common natural hazard that seriously threatens human life and causes significant social and economy losses [1]. Recent studies have shown that climate change and human activities are causing more landslides in mountain regions, such as high mountain Asia [2], [3], [4]. To investigate the type and spatial distribution of landslides, to map the landslide susceptibility and risk, and to study the landscape evolution caused by landslides, it is fundamental to construct a comprehensive landslide inventory. Landslide inventory records key information of landslides, such as location and size [4], [5], [6]. Timely establishing and updating the landslide inventory is essential for a wide range of applications (e.g., risk control and landslide susceptibility mapping) in landslide-prone areas [7], [8], [9].

There are many methods to establish a landslide inventory, including field investigation and satellite image analysis [6]. Field investigation can establish a landslide inventory with high accuracy, but it is a time-consuming and laborious process, and can even be a potentially dangerous task for surveyors [4], [6]. With the advent of modern remote sensing technologies, extracting landslides from satellite images has received growing attention. There are various remote sensing techniques that can be employed to detect landslides from satellite imagery. These techniques include visual interpretation and computational image analysis. Visual interpretation manually extracts landslides from satellite images. Despite its accuracy, it heavily relies on the professionalism of experts and is inefficient, limiting its applications in a large area. Computational image analysis includes traditional image processing technologies, machine learning, and deep learning [10]. Traditional image processing technologies have been developed since the early days of digital imaging, and rely on a variety of mathematical operations to manipulate pixel values within an image. These techniques include change detection, feature extraction, classification, and object-based image analysis, and have been widely used in landslide detection. In recent years, machine learning has emerged as a powerful tool for image processing, particularly in the area of object detection and recognition. Machine learning algorithms use statistical models to learn patterns and relationships within a set of training data, and then apply these models to new data to make predictions or classifications. Common machine learning

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

algorithms used in landslide detection include support vector machines (SVMs), random forests (RFs), and artificial neural networks (ANNs). For instance, Hu et al. [10] used spectrum, geomorphology and normalized vegetation index to train SVM to extract seismic landslides. Hu et al. [11] compared multiple machine learning algorithms for landslide detection. They found that the RF achieved better performance than SVM and ANN. Li et al. [12] proposed a change detection-based Markov random field method for landslide detection. Piralilou et al. [13] combined the object-based image analysis (OBIA) with three kinds of machine learning algorithm to detect landslide, and they found that the OBIA reduced the image noise influence and improved the landslide detection accuracy. Meghanadh et al. [14] used the RF to detect landslide, and result showed that the RF achieved a satisfactory performance in terms of landslide detection. Machine learning typically requires feature engineering, which is the process of selecting and extracting relevant features from raw data for use as input to the machine learning algorithm. Feature engineering can be time-consuming and requires domain knowledge. In addition, machine learning models may struggle with complex and highly variable datasets, as they require a large number of representative features to be manually extracted from the data. These may limit the ability of machine learning to generalize to new problems.

Deep learning, a subset of machine learning, has experienced rapid development in the era of big data. It consists of multiple layers of interconnected artificial neurons that are trained using large amounts of data to automatically learn increasingly complex representations of the input data. Convolutional neural networks (CNNs) are a popular type of deep learning model used in image processing, and have been applied successfully to a wide range of tasks including object detection, segmentation, and image classification. Besides, the attention mechanism, originated from human visual system, has the ability to boost the performance of deep learning algorithms. The development of deep learning has led to significant advances in detecting landslides from satellite images [15]. Deep learning algorithms are widely used for object detection in computer vision. There are three commonly used deep learning algorithms for object detection, including classification, semantic segmentation, and object detection.

1) Classification algorithms are used to classify images into different categories to detect objects. For instance, Ji et al. [16] designed an attention-enhanced CNN to detect landslides in Bijie, China, and experimental results showed that the integration of the designed attention block and ResNet50 achieved a satisfactory landslide recognition performance Bui et al. [17] presented an classification model of combining CNN and image transform to detect landslide in satellite images, the classification results showed that the accuracy can reach 96%. Liu et al. [18] used depth, residual and dense neural networks for training classification models to detect landslides, the experiments showed that three algorithms have higher accuracy. Cai et al. [19] used DenseNet network to detect landslide and validated DenseNet has better detection performance than traditional neural networks and machine learning .Despite its accuracy, it has limited use in object detection as it can only identify the presence or absence of an object in an image, but cannot locate its position or shape.
2) Semantic segmentation algorithms identify the presence, location, and shape of objects in an image. The advantages of semantic segmentation include its high accuracy, its ability to differentiate between different objects that are close together, and its ability to detect objects of different shapes and sizes. Many scholars have presented semantic segmentation algorithms for landslide detection. For instance, Braagnolod et al. [20] compiled Nepal landslides from Landsat-8 images to train the U-net to detect landslides at the pixel level. Fang et al. [21] proposed an attention boosted U-Net to automatically delineate historical landslides from LiDAR dataset, and they validated the joint use effectiveness of LiDAR and deep learning in landslide detection. Zhou et al. [22] proposed a weakly supervised approach which combined CAM and cycle-GAN to reduce the workload of making segmentation labels and realize higher accuracy detection of landslide. Amankwah et al. [23] proposed attention boost SNU-Net segmentation algorithm to detect landslide from optical satellite images and achieve better performances than original U-Net. However, it can be computationally expensive and requires a large dataset for training. Lu et al. [24] proposed a dual U-Net segmentation model to extract landslides using Sentinel-2 and DEM data. Chen et al. [25] combined spectral and context information of landslide, boosting the performance of original U-Net. Liu et al. [26] proposed a feature-fusion based segmentation network to construct landslide inventory from high resolution satellite imagery and DEM data.
3) Object detection algorithms can identify the presence, location, and shape of objects in an image. This algorithm is more complex than classification as it requires identifying the bounding box around each object in an image. There are many object detection algorithms in the field of landslide detection. For example, Ju et al. [27] established 6111 landslide samples in Gansu, China by the visual interpretation of Google Earth images to train RetinaNet, YOLOv3 and Fast R-CNN. They found that fast R-CNN achieved the highest performance. Liu et al. [28] proposed an attention-boosted mask-R CNN to detect landslide from InSAR deformation images. Tanatipuknon et al. [29] combined the faster-R-CNN and classification decision tree for training landslide detection model, the evaluation results showed that the combination method achieved a superior performance. Yang et al. [30] proposed a background-enhancement mask R-CNN method to detect landslide, the result showed that the background-enhancement significantly improved model performance. The advantages of object detection include its ability to detect multiple objects in an image, its accuracy, and its ability to identify objects of different shapes and sizes. Liu et al. [31] integrated YOLOv7 and SENet attention block to detect landslide while using varifocal loss function, experimental results showed that the improved YOLOv7 has strong

performance for landslide detection. Particularly, among various object detection algorithms, YOLO is a popular choice for object detection in a wide range of applications, due to its speed, accuracy, simplicity, and flexibility. By contrast to classification and semantic segmentation based object detection algorithms, YOLO is very fast, capable of processing images in real-time, which makes it ideal for real-time applications. Meanwhile, YOLO can detect objects of different shapes and sizes, and it can also handle multiple objects in an image. Despite its achievements in detection speed and accuracy, there is no general rule to select the backbone and attention mechanism for YOLO. The selection of these modules depends on specific application needs, such as landslide detection in this article. Meanwhile, the output of YOLO is a series of anchor boxes, not accurate landslide boundaries, which are not suitable for constructing landslide inventory.

Based on the aforementioned analysis, this article aims to propose a lightweight attention-guided YOLO with level set layer (LA-YOLO-LLL) for landslide detection. We first conduct a comparative study to find a suitable backbone for YOLO for the landslide detection task. After that, we designed a light-pyramid feature reuse fusion (LPFRF) attention mechanism to further improve landslide detection accuracy. Finally, we integrated the level set layer into YOLO head to extract accurate landslide boundaries.

The rest of this article is organized as follows. In Section II, the dataset used in this article is presented. Section III introduces the proposed method, LA-YOLO-LLL, and provides details of its design. The experimental results are presented and discussed in Section IV. Finally, Section V concludes this article and suggest future research directions.

II. Dataset Used in This Article

Two typical landslide prone regions (i.e., Bijie and Taiwan) were selected as study areas, see Fig. 1. This article trained and tested the performance of the presented method on Bijie case, and the transferability of the presented method was tested on Taiwan case. Bijie is located in the northwest of Guizhou Province, China. The terrain of this area is mainly plateau and mountain with an average elevation of 1400 m. Red soil is the main soil type in Bijie that is susceptible to landslide during heavy rain, particularly in subtropical monsoon. Ji et al. [16] published the landslide dataset of Bijie that can be downloaded.

The sample image size varies from 128×128 to 1024×1024, and each image has 3 spectral channels (i.e., R, G, and B channel) with a spatial resolution of 0.8 m. The dataset contains 770 landslide images and 2100 nonlandslide images. To balance the sample sizes of landslide and nonlandslide images, we used the data augmentation to increase the landslide sample size. In the experiment, we divide the dataset into 80% for training and 20% for testing. Within the training set, only 10% was allocated for validation purposes.

Taiwan locates in the southeast of China, and is an island with an area of 36 000 km2. The region is one of the most

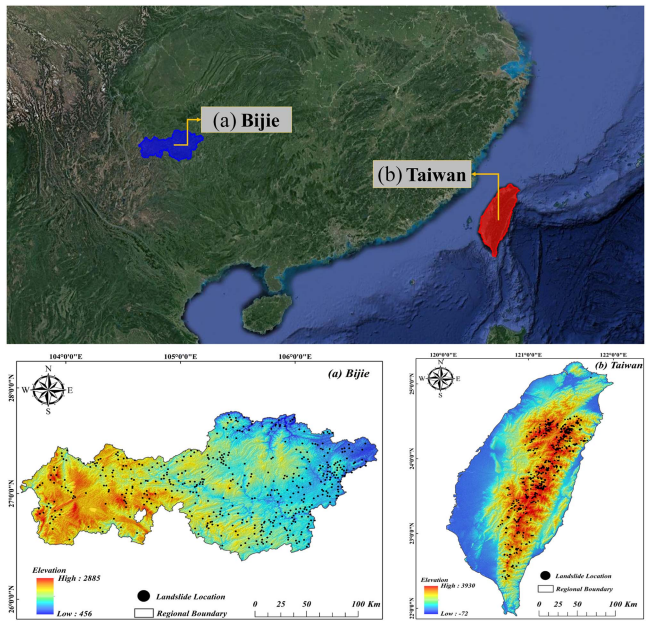


Fig. 1. Locations of the study areas.

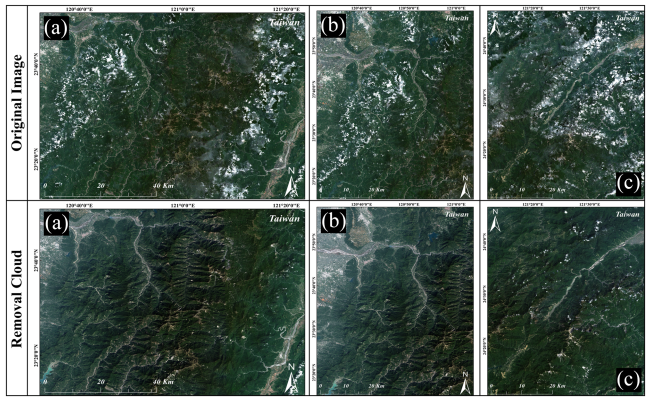


Fig. 2. Upper panel and bottom panel show the original Sentinel-2 image and the image after cloud removal over the study area, respectively.

landslide-prone areas in the world due to many factors, such as heavy rains, steep topography, weak geological formations, strong earthquakes, loose soils, human activities [32], [33]. We used Google Earth Engine to download Sentinel-2 images of Taiwan from January 2022 to January 2023. The data were preprocessed by radiometric calibration and atmospheric correction. We produced a cloud-free Sentinel-2 image mosaic followed the rules presented by [34], see Fig. 2. After that, we produced 630 landslides across Taiwan (see Fig. 3) by the visual interpretation from the Sentinel-2 mosaic image. These landslides were taken as the ground truth to validate the transferability of the presented method.

The landslide characteristics in Bijie and Taiwan are similar since they are both triggered by intensive rain in vegetated areas. In optical satellite images, the landslides in these two areas exhibit similar signatures.

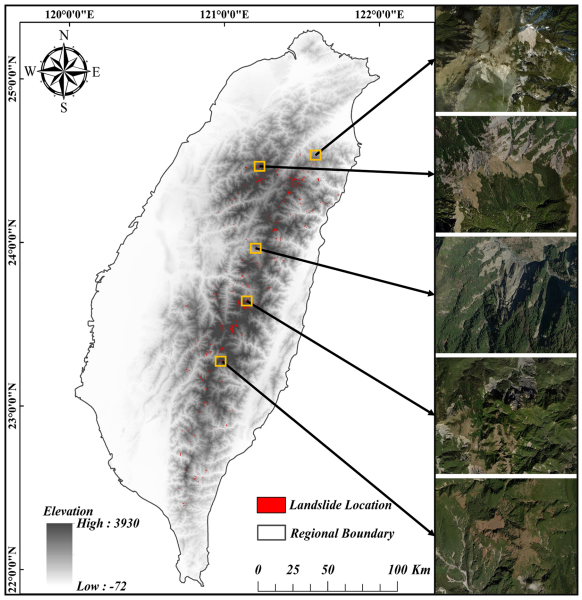


Fig. 3. Ground truth dataset of landslides in Taiwan obtained by the visual interpretation of Sentinel-2 satellite images.

III. Methodology

The presented method mainly consists of three steps. We firstly conduct a comparative study to select an optimal feature extraction backbone from multiple CNNs. Then, we present a new attention mechanism and include it to the end of backbone of YOLO. Lastly, we combine the YOLO and level set to extract precise landslide boundary. Fig. 4 shows the framework of the presented LA-YOLO-LLL-YOLO model.

A. Overview of Original YOLOv4

YOLO is a typical one-stage object detection algorithm that significantly focuses on accelerating the computation speed to realize real time object detection [35], [36], [37], [38]. Up to now, YOLO has updated eight versions [35], [36], [39], [40], [41], [42]. Among these versions, YOLOv4 has high detection accuracy and efficiency [36], [38], [43]. Many object detection algorithms (i.e., YOLOv3 and RetinaNet) need to fix the feature map size that leads to expensive computation and inflexibility. By contrast, the feature map size of YOLOv4 is flexible. When objects of interest appear in the grid, YOLOv4 segments the image into regular grids and generates multiple anchor boxes around objects [36], [43], [44]. Each anchor box has four parameters: the center coordinate; height and width; confidence; and prediction category of the anchor box. YOLOv4 may produce many similar and spatially overlapped anchor boxes surrounding the same object (i.e., landslide in this article). The best bounding box can be selected by the non-maximum suppression algorithm [45]. The structure of YOLOv4 consists of three main parts: backbone, neck and head, see Fig. 5. The backbone (i.e., generating a feature pyramid) aims to learn features on each input image. The original backbone of YOLOv4 is CSPDarkNet53. The neck includes the spatial pyramid pooling (SPP) [46] and the path aggregation network [47] that will boost receptive field and realize information fusion of multiscale feature maps. The head consists of two convolution layers to reduce the information loss and output the prediction results.

B. Structure of the Presented Method

Despite the efficiency of YOLOv4, it suffers from the following drawbacks to hinder the model performance in landslide detection.

1) Original backbone of YOLOv4 (i.e., CSPDarkNet53) has heavy structure that leads to expensive computation cost when the number of anchor boxes and categories is increasing.
2) Although the SPP module can compress feature and boost the receptive field, the pooling function will lose feature information to some extent.
3) YOLOv4 is primarily designed for object detection using a series of bounding boxes. However, the landslide boundary is irregular, which cannot be precisely described by bounding boxes.

To tackle these issues, this article designs a new method called LA-YOLO-LLL (see Fig. 6) to improve the YOLOv4 performance in landslide detection. We, first, conduct a comparative study to find an optimal backbone employed to improve the accuracy as well as the transferability. After that, this article designs a new attention mechanism to enhance the feature extraction ability of YOLOv4 while mitigates information loss due to SPP. Finally, we integrate the YOLOv4 and level set to extract accurate landslide boundary, other than bounding box generated by YOLOv4.

C. Backbone of the Presented Method

The backbone network is an essential component of LA-YOLO-LLL. It is responsible for extracting high-level features from input images, which are then used by subsequent layers of the model for object detection. The choice of backbone has a significant impact on the landslide detection accuracy of the presented method. However, there is no single, general rule for selecting a backbone that will work for all applications, such as the landslide detection in this article. The choice of backbone depends on several factors, including the specific requirements of the task, the size and complexity of the dataset, and the compatibility with other components of LA-YOLO-LLL. Original backbone of YOLOv4 (CSPDarkNet53) is a complex network [43]. While complex networks have higher representation power and detection performance, lightweight backbone networks generally exhibit better transferability compared to complex ones. Lightweight backbone networks are designed to maintain low computational and storage requirements while effectively learning and representing target features in new tasks [44]. This makes them suitable for scenarios with limited resources or when deploying on mobile or edge devices. They strike a balance between performance and efficiency, allowing for efficient adaptation to new detection tasks. Meanwhile, compared to other object detection scenarios, landslide detection scenes are relatively simple.

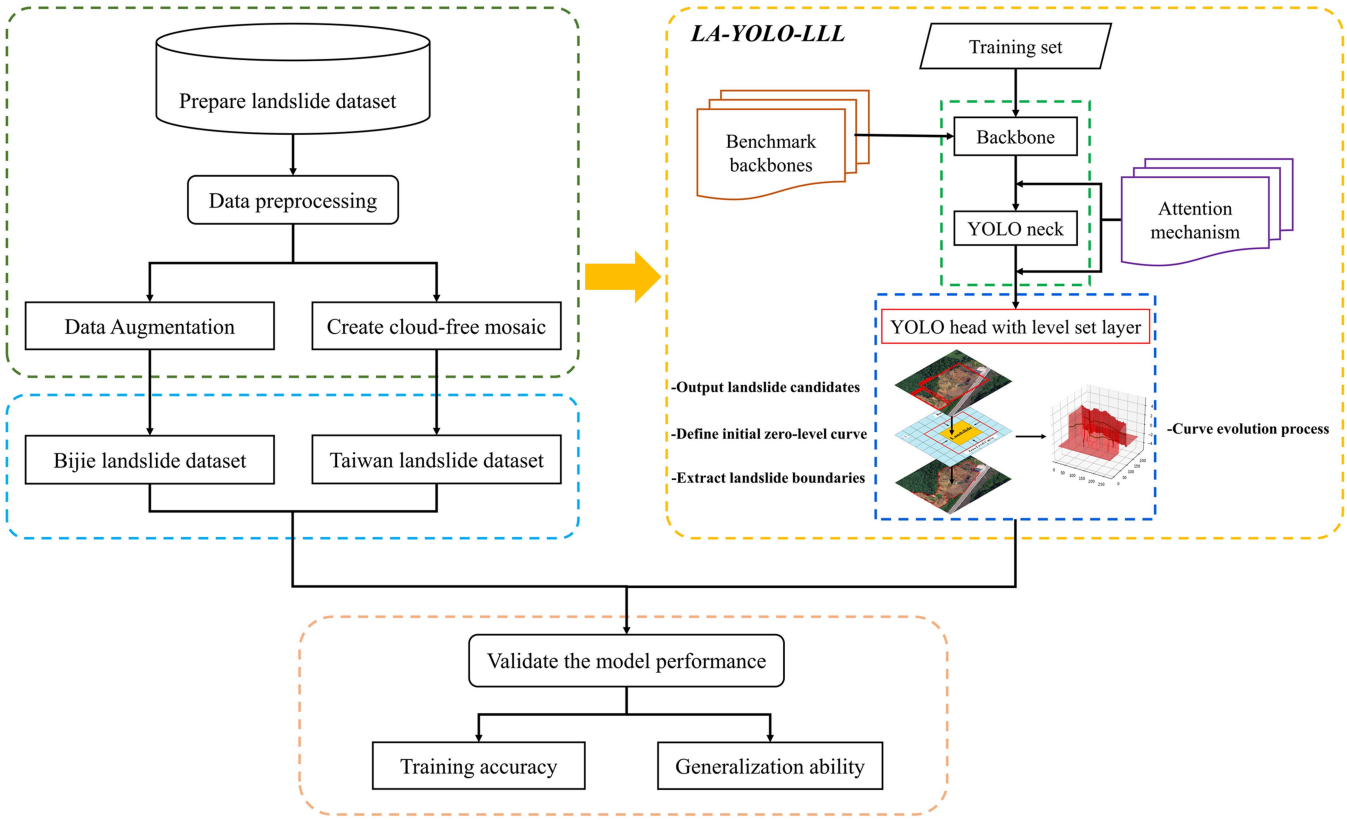


Fig. 4. Flowchart of the presented LA-YOLO-LLL model.

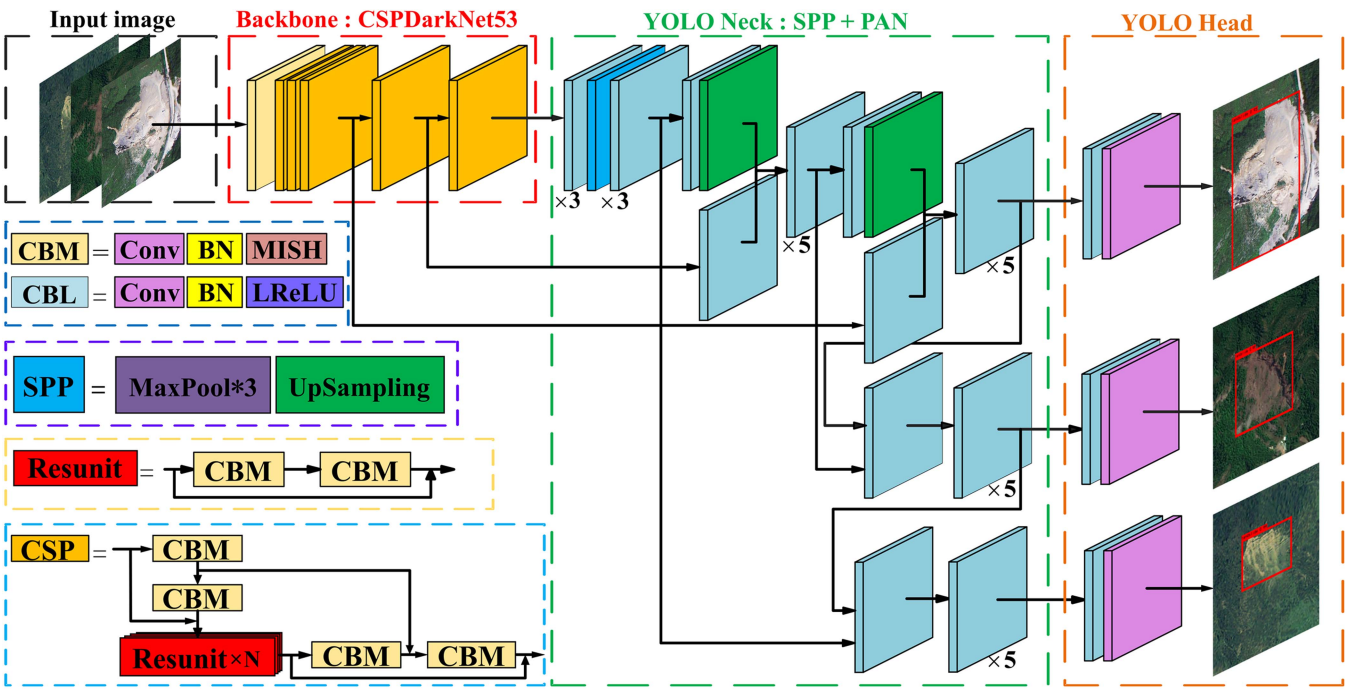


Fig. 5. Network structure of YOLOv4.

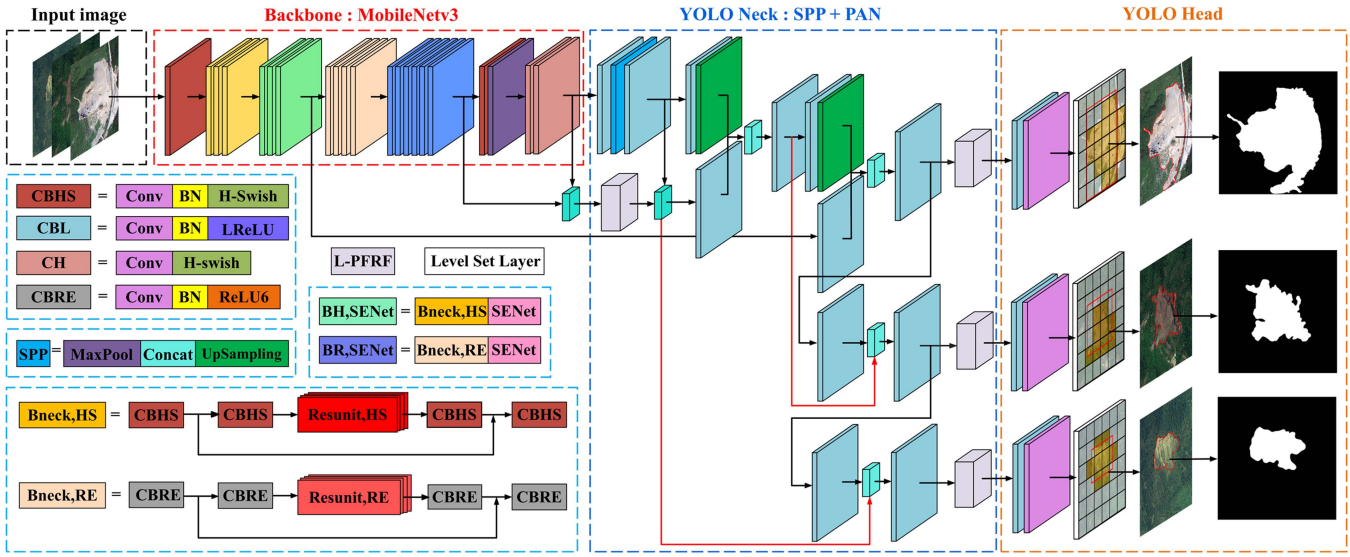


Fig. 6. Structure of the presented method.

Taking all these factors into consideration, MobileNetv3 is introduced as the backbone network in this article. MobileNetv3 is a deep neural network architecture that was developed by Google Research [48]. It is designed to be more efficient than its predecessors, MobileNetv1 and MobileNetv2, while maintaining similar accuracy. MobileNetv3 employs several techniques to reduce the computational cost of the network, such as a streamlined architecture, improved activation functions, and a more efficient implementation of depthwise convolutions. MobileNetv3 has achieved state-of-the-art performance in various image-related tasks and is widely used in mobile and embedded systems. The detail introduction of MobileNetV3 can refer to [48].

D. Light Pyramid Features Reuse Fusion Mechanism

Attention mechanism is a typical method to improve the performance of deep learning algorithms [49], [50]. The attention mechanism learns the weights according to the loss function and then determines the attention features of the network based on the feature map weight. Most attention blocks use a fully connected layer to reduce the output feature map dimension. Despite its improvement in the training accuracy, it results in information loss and affects detection performance. Although the SPP can compress feature map and boost receptive field to reduce the computation cost, it leads to information loss to some extent. To tackle these issues, this article designed a light pyramid features reuse fusion mechanism (LPFRF) that reuses feature information loss from the dimensions of semantic and spatial, see Fig. 7. It can be seen that feature maps of different scales are obtained after processing by the backbone of the presented LA-YOLO-LLL. To improve the accuracy and generalization of the landslide detection, the LPFRF constructs two 1×1 global average pooling and a 3×3 global max pooling to enhance multiscale receptive field. In the spatial attention mechanism, the convolutional window size is 3×3 and 7×7, which can extract features from different spatial scales.

The LPFRF module consists of two submodules: the channel attention submodule and the spatial attention submodule. In the channel attention sub-module, the input feature map F is passed through global average pooling and global max pooling operations with a window size 1×1 of all channels. The resulting features are then processed using a 1×1 convolutional layer and ReLU activation function, followed by another 3×3 convolutional layer with the LuckReLU activation function. Finally, the output is passed through a 1×1 convolutional layer with a sigmoid activation function to obtain the feature map after processing by the LPFRF channel attention sub-module. The mathematical expression for this process is given as, (1) shown at the bottom of the next page.

After processing by the channel attention module of LPFRF, the processed feature map F' will be subsequently processed by the spatial attention module of LPFRF. Specifically, F' is processed by the global max pooling and the global average pooling to obtain two $H \times W \times 1$ feature maps. These two feature maps will be spliced through the broadcast mechanism to generate a new feature map. Specifically, the spliced feature maps are sequentially input into the 3×3 convolutional layers, the ReLU activation function, the 7×7 convolutional layer, and the sigmoid activation function. This process can be implemented as, (2) shown at the bottom of the next page.

The max pooling and average pooling with different sizes enhance the receptive field of the feature map. This will extract more detailed semantic information and spatial location information from multichannel and multispatial dimensions. This will reduce the feature information loss caused by the fully connected layers and the SPP module.

E. Level Set Layer

The output of YOLO consists of a series of bounding boxes that represent the detected landslides in a satellite image. These bounding boxes are not necessarily the exact boundaries of the landslides, but rather rectangles that enclose the detected areas.

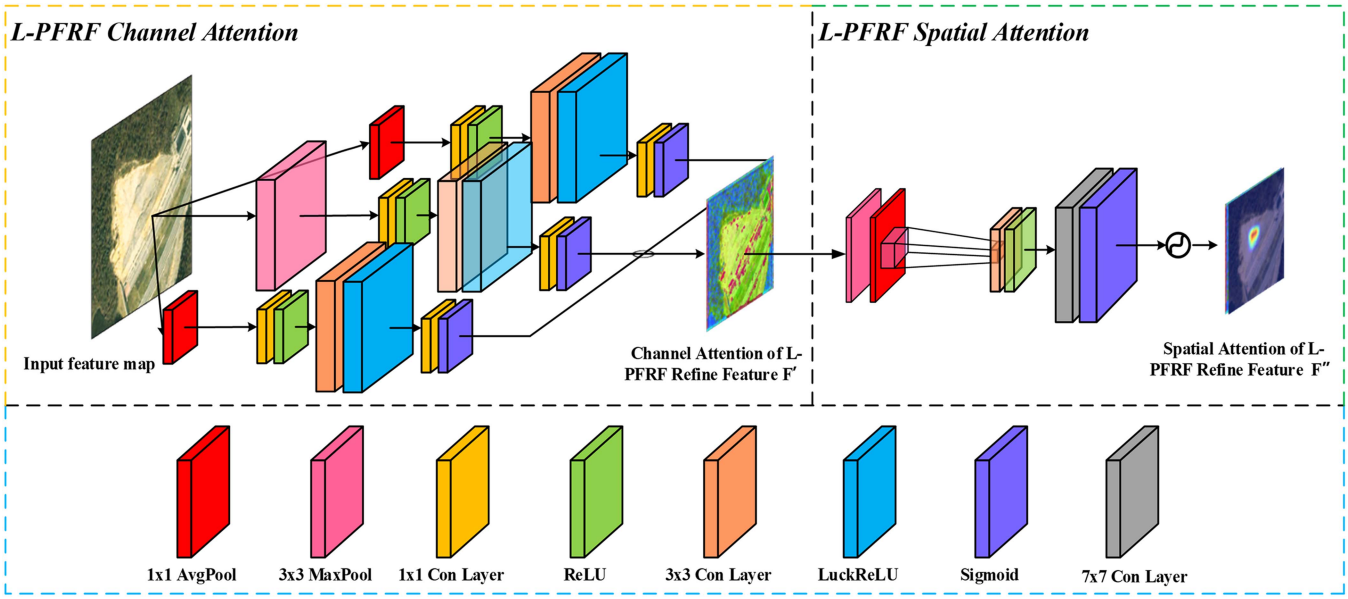


Fig. 7. Structure of the designed LPFRF attention mechanism.

In addition, a single bounding box produced by YOLO may contain multiple landslides and cannot separate them individually. Although YOLO may produce inaccurate bounding boxes at times, it provides valuable information for accurately extracting the boundaries of landslides. Based on this observation, we incorporate the level set layer into YOLO to extract precise landslide boundaries.

For an image I in two-dimensional (2-D) space Ω (i.e., $\Omega \in \mathbb{R}^2$), the boundary C of an open set $\omega \in \Omega$ is defined as $C = \partial\omega$. In the level set layer, the zero-level set ϕ is generally applied to represent the boundary C as follows:

$$\forall (\mathbf{x}) \in \Omega \begin{cases} C: & \{(\mathbf{x}) : \phi(\mathbf{x}) = 0\} \\ \cup C_I: & \{(\mathbf{x}) : \phi(\mathbf{x}) > 0\} \\ \cup C_O: & \{(\mathbf{x}) : \phi(\mathbf{x}) < 0\} \end{cases} \tag{3}$$

where Ω represents the entire spatial domain of an image, $\mathbf{x}$ is the pixel coordinate, $\cup C_I$ and $\cup C_O$ denote the pixel coordinate sets inside and outside the zero level curve, respectively. The size of the landslide is relatively small compared to the entire satellite image, and using the full spatial domain to evolve the zero level set would lead to unnecessary increases in computational cost. To address this issue, this article employed a spatial domain that is twice the size of the bounding box used in YOLO. The function ϕ divides the spatial domain into two parts: region inside ω (landslide) and region outside ω (non-landslide). The level set usually begins with an initial level set ϕ_0. Since we use the bounding box of YOLO as the initial level set, ϕ_0 can be expressed as

$$\phi_0 = \begin{cases} -1, & \mathbf{x} \in \cup C_I \\ +1, & \mathbf{x} \in \cup C_O \end{cases}. \tag{4}$$

The level set is updated through gradient descent by minimizing an energy function. This function is defined based on the difference in image features between the foreground (landslide) and background (non-landslide), such as color and texture. The use of shape and regions in level set segmentation can improve performance. However, when dealing with complex images, level set segmentation is limited since it only uses low-level features. In contrast, YOLO has the ability to learn and encode useful high-level features. Therefore, integrating the level set layer into YOLO can improve landslide boundary detection performance by combining the strengths of both methods. The insertion of the level set layer into YOLO allows for more

$$\begin{aligned} M_c(F) = \text{Sigmoid}(&f^c_{1\times1}(Luck\,\text{ReLU}(f^c_{3\times3})\text{ReLU}(f^c_{1\times1}(\text{AvgPool}^c_{1\times1})))) \\ &+ f^c_{1\times1}(\text{Luck ReLU}(f^c_{3\times3})\text{ReLU}(f^c_{1\times1}(\text{MaxPool}^c_{3\times3})))) \\ &+ f^c_{1\times1}(\text{Luck ReLU}(f^c_{3\times3})\text{ReLU}(f^c_{1\times1}(\text{AvgPool}^c_{1\times1}))))) \\ = \sigma(W_2(W_1)(W_0(F^c_{\text{avg}}))) &+ W_2(W_1)(W_0(F^c_{\max}))) + W_2(W_1)(W_0(F^c_{\text{avg}}))). \end{aligned} \tag{1}$$

$$M_s(F) = \sigma\left(f^{7\times7}\left(f^{3\times3}(\text{AvgPool}(F))\right);(\text{MaxPool})\right) = \sigma\left(f^{7\times7}\left(f^{3\times3}(F^s_{\text{avg}};F^s_{\max})\right)\right). \tag{2}$$

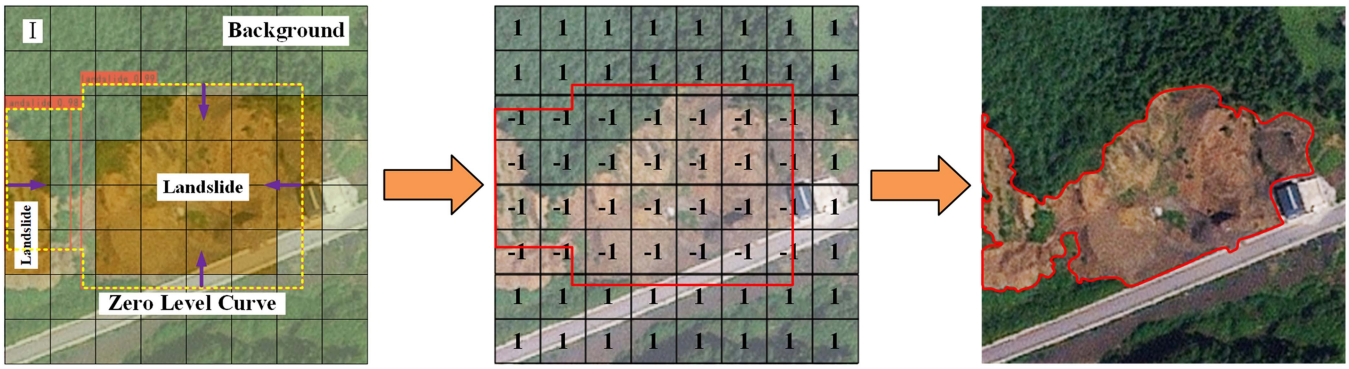


Fig. 8. Level set layer of the presented method for precise landslide boundary detection. (a) Zero level curves derived from the bounding boxes. (b) Evolution of the zero level curve. (c) Final landslide boundaries produced by the presented method with the level set layer.

accurate detection of landslides, leveraging both low-level and high-level features for optimal results.

In the level set, the energy functional typically includes a regularization term that penalizes high curvature of the level set function, a term that enforces the signed distance property of the level set function, and a term that drives the motion of the level set function towards the boundary of the object of interest. Considering these three terms, a level set formula ϕ_t [48] can be defined as follows:

$$\phi_t = \delta_\varepsilon(\phi)\left[\mu \text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \nu - \lambda_1(I - c_{\text{in}})^2 + \lambda_2(I - c_{\text{out}})^2\right] \tag{5}$$

where $\delta_\varepsilon(\bullet)$ is the Dirac delta function, μ, ν, λ_1, andλ_2 are free parameters, $\text{div}(\nabla\phi/|\nabla\phi|)$ denotes the mean curvature of zero level curve, $\text{div}(\bullet)$ represents the divergence operator, c_{in} and c_{out} represent the mean intensities inside and outside the zero level curve, respectively. In image analysis, the Kronecker delta function is commonly used as a discrete approximation of the Dirac delta function, which is a continuous function. However, the Kronecker delta function can be computationally expensive to evaluate, especially for large images or high-dimensional data. Instead of using the exact Kronecker delta function, it is possible to use approximation techniques that have a similar effect on the image but are easier to compute. Thus, (5) can be rewritten as

$$\phi_t = \left[\mu \text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \nu - \lambda_1(I - c_{\text{in}})^2 + \lambda_2(I - c_{\text{out}})^2\right]|\nabla\phi|. \tag{6}$$

Previous studies have demonstrated that $\mu\text{div}(\nabla\phi/|\nabla\phi|)$ and ν only marginally affect the level set evolution, but they do increase the computational cost significantly. To improve the computational efficiency, the first two terms in (6) can be eliminated [52], and we rewrite (6) as

$$\phi_t = \left[-\lambda_1(I - c_{\text{in}})^2 + \lambda_2(I - c_{\text{out}})^2\right]|\nabla\phi|. \tag{7}$$

Suppose that the inside and outside region have the same weights (i.e., $\lambda_1 = \lambda_2 = 1$) in the level set evolution, we can simplify (7) to

$$\phi_t = \left[-(I - c_{\text{in}})^2 + \lambda_2(I - c_{\text{out}})^2\right]|\nabla\phi| = T_f|\nabla\phi| \tag{8}$$

where $T_f = (2I - c_{\text{in}} - c_{\text{out}})(c_{\text{in}} - c_{\text{out}})$. When T_f has a wide range of values or where the level set function is sensitive to small changes in T_f, (8) may become numerically instable. To maintain numerical stability of (8), we applied the data normalization to T_f. Data normalization can also improve the convergence rate and accuracy of the level set by reducing the influence of outliers or extreme pixel values. Fig. 8 illustrated the process of integrating the level set layer into the presented method for detecting precise landslide boundaries.

IV. Experiment and Discussion

This article uses a Windows PC with a NVIDIA GeForce RTX3090 and an Intel Xeon (R) Gold 6230R CPU to execute all the experiments. All models share the same training settings, and are pretrained on the VOC dataset for 300 epochs using the Adam optimizer with a batch size of 64. The initial learning rate started from 10^{-4} with a decreasing rate of 10 after 30 epochs. We set time-step and Gaussian scale of level set to 15 and 3, respectively. In the experiment, we divide the dataset into 80% for training and 20% for testing. Within the training set, 10% was allocated for validation purposes.

This article assesses model performance from two aspects: training accuracy and generalization ability. While training accuracy measures how well a model has learned the patterns in the training data, generalization ability measures how well a model can generalize these patterns to new, unseen data. Both metrics are important to evaluate the performance and reliability of a landslide detection model. Four metrics (precision, recall, F1-score, and mAP) are used to evaluate training accuracy

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

$$\text{F1} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

$$\text{mAP} = \sum_{i=0}^{n} \frac{1}{n} \int_0^1 \text{Precision}(\text{Recall}) d(\text{Recall}) \tag{12}$$

TABLE I
MODEL PERFORMANCE OF LANDSLIDE EXTRACTION USING YOLOV4 DIFFERENT BACKBONES IN BIJIE

Description	Precision (%)	Recall (%)	F1-score (%)	mAP (%)
Baseline + CSPDarkNet53	92.38	89.62	90.98	93.92
Baseline + MobileNetv3	94.05	90.14	92.05	94.20
Baseline + ResNet50	91.82	90.35	91.08	93.61
Baseline + GhostNet	92.56	88.32	90.39	93.10
Baseline + DenseNet121	93.46	78.58	85.38	92.00
Baseline + VGG16	96.17	91.98	94.03	95.40

TABLE II
MODEL PERFORMANCE OF LANDSLIDE EXTRACTION USING YOLOV4 DIFFERENT BACKBONES IN TAIWAN

Description	Overall (%)	Kappa (%)	User's accuracy (%)		Producer's accuracy (%)	
			landside	non-landslide	landslide	non-landslide
Baseline + CSPDarkNet53	67.12	24.29	100.00	17.93	62.76	100.00
Baseline + MobileNetv3	76.74	50.00	81.78	68.15	81.43	68.60
Baseline + ResNet50	58.95	26.28	85.54	36.68	53.08	75.18
Baseline + GhostNet	72.43	42.74	75.71	67.39	78.13	64.32
Baseline + DenseNet121	53.66	20.01	97.85	26.80	44.83	95.35
Baseline + VGG16	67.70	32.30	73.50	58.77	73.30	59.03

where TP, FP, and FN denote true positive, false positive, and false negative, respectively.

Generalization ability is evaluated based on user's accuracy (UA), producer's accuracy (PA), overall accuracy (OA), and Kappa coefficient (κ), which can be computed based on the confusion matrix

$$\mathrm{UA} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{13}$$

$$\mathrm{PA} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{14}$$

$$\mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{15}$$

$$\kappa = \frac{\mathrm{OA}-p_e}{1-p_e} \tag{16}$$

where TN is true negative, and $p_e = \frac{(\mathrm{TP}+\mathrm{FN})\times(\mathrm{TP}+\mathrm{FP})+(\mathrm{FN}+\mathrm{TN})\times(\mathrm{TN}+\mathrm{FP})}{N^2}$, and N is the element number of the confusion matrix.

A. Comparison Different Backbones of YOLOv4 for Landslide Detection

The original backbone network of YOLOv4 is CSPDarknet53. To select a suitable backbone network for landslide detection task, this article conducted a comparative study. We compared CSPDarkNet53, Resnet50 [53], GhostNet [54], DenseNet121 [55], MobileNetv3 and VGG16 [56] to replace the original backbone network of YOLOv4. Specially, we compared the training accuracy and generalization ability of YOLOv4 with different backbone networks to select the best one. Tables I and II report the comparison results of YOLOv4 with different backbone networks in Bijie and Taiwan. It can be seen that YOLOv4 with CSPDarkNet53 achieves the lowest training accuracy on Bijie case while YOLOv4 with VGG16 the highest training accuracy. The performance of VGG16, MobileNetV3 and GhostNet are better than the original backbone CSPDrakNet53, and the performance of DenseNet121 and ReseNet50 are relatively lower than CSPDrakNet53. From the viewpoint of the transferability, MobileNetv3, GhostNet, and VGG16 have the better performance than CSPDrakNet53. ResNet50 and DenseNet121 have lower performance than CSPDrakNet53.

Figs. 9 and 10 illustrate examples of landslide detection using YOLOv4 with different backbones in the cases of Bijie (used for model training) and Taiwan (used for model transferability validation), respectively. It can be seen that YOLOv4 with different backbones on Bijie achieved similar results. Both models can successfully detect landslides on eight samples, suggesting that the influence of backbone selection on YOLOv4 training accuracy is small. However, when applied to detect landslides in Taiwan, there was a significant difference in performance among the six backbones. Specifically, YOLOv4 with MobileNetv3 successfully detected seven landslides out of eight examples, while DenseNet121, ResNet50, and VGG16 detected three landslides, and CSPDarkNet53 and GhostNet detected only two landslides. These findings are consistent with the results given in Tables I and II, indicating that YOLOv4 with MobileNetv3 has the highest generalization ability among the six backbones. Besides, comprehensively considering model model size presented in Table III, MobileNetv3 demonstrates lower complexity.

B. Comparison of Multiple Attention Mechanisms for Landslide Detection

Table IV gives the performance of landslide detection by YOLOv4 using different attention mechanisms in Bijie. It

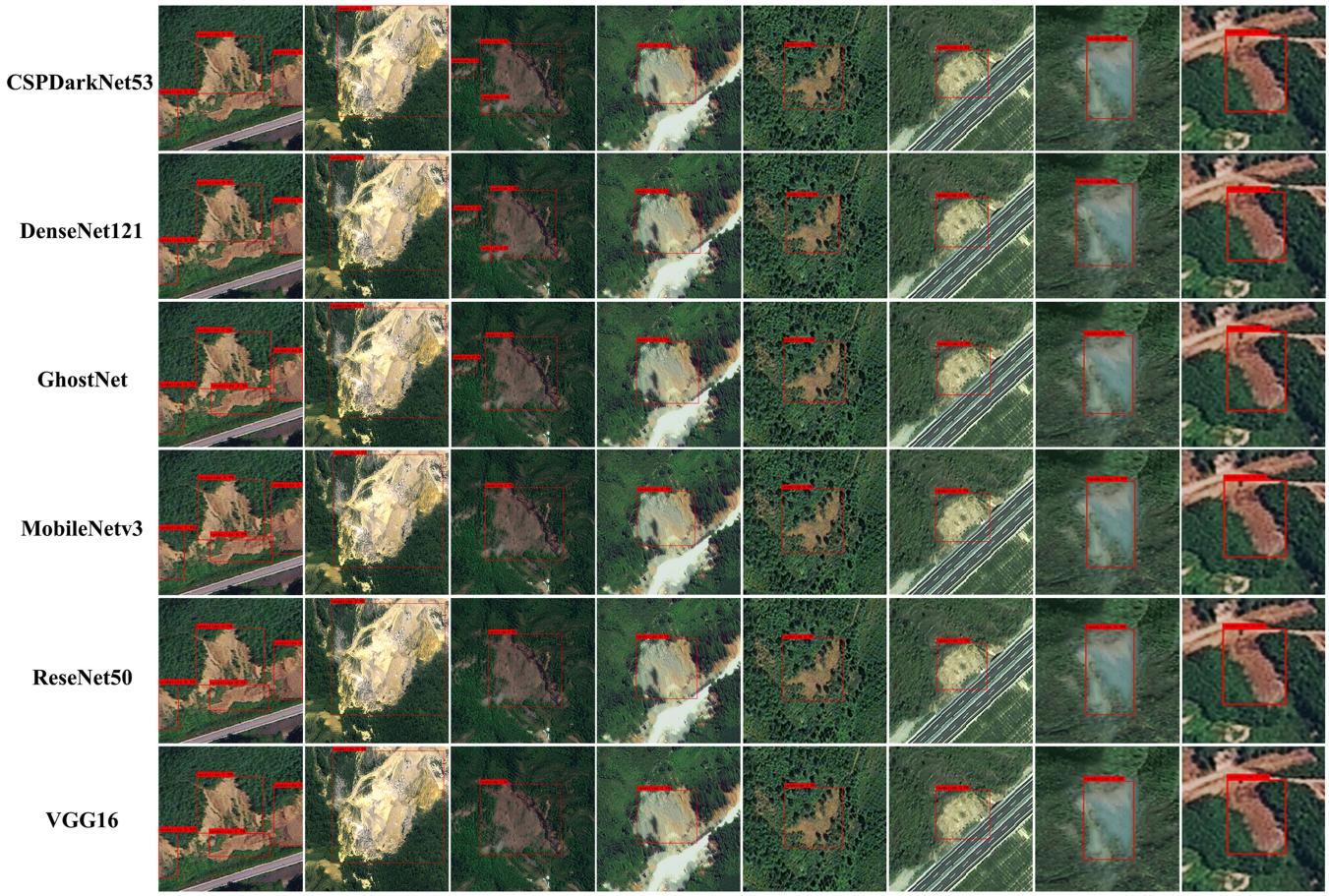


Fig. 9. Landslide extraction results of Bijie dataset by YOLOv4 using different backbones.

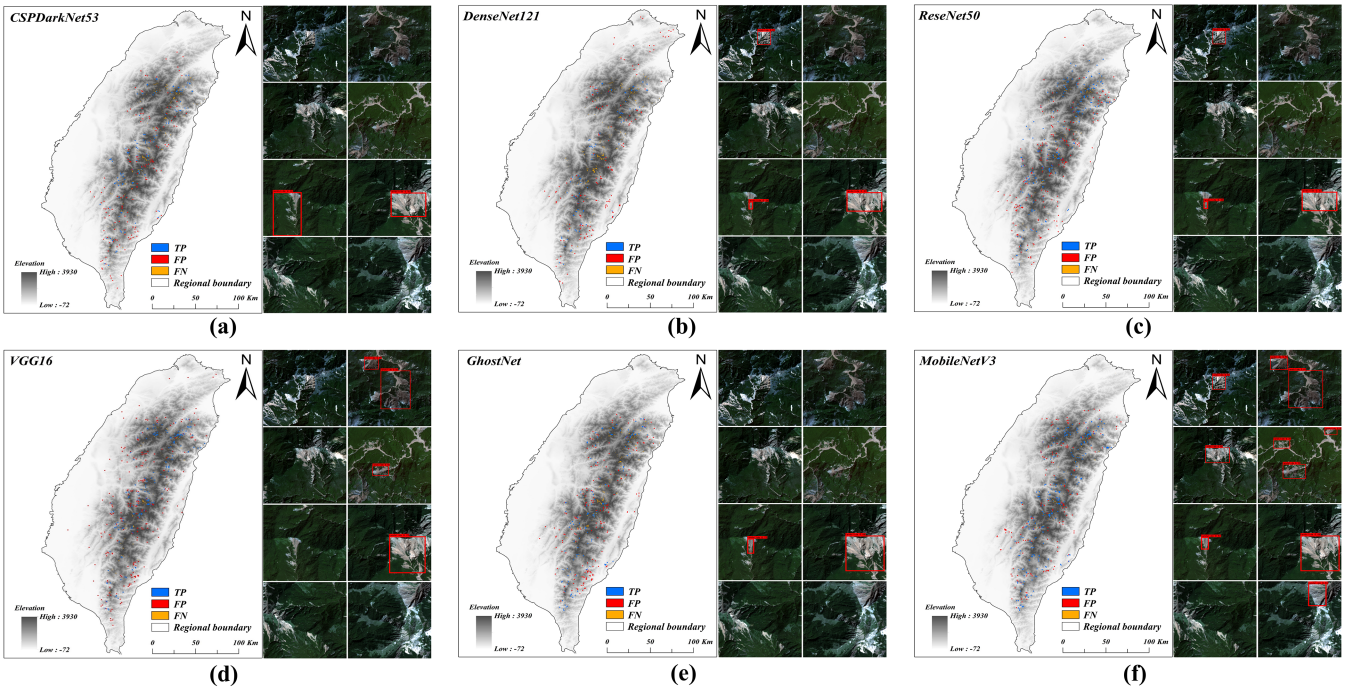
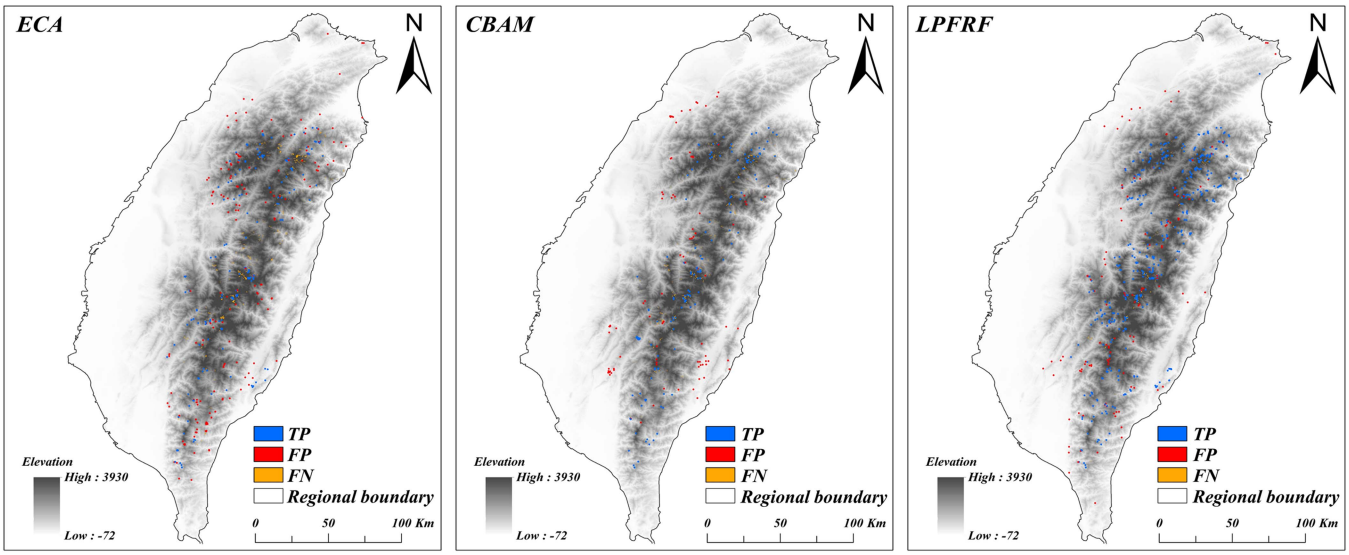


Fig. 10. Landslide extraction results of Taiwan by YOLOv4 using different backbones.

TABLE III
SIZE OF YOLOV4 WITH SIX BACKBONES

Baseline + CSPDarkNet53	Baseline +MobileNetv3	Baseline +ResNet50	Baseline +GhostNet	Baseline +DenseNet121	Baseline +VGG16
334.86MB	44.74MB	128.48 MB	43.60 MB	62.71 MB	91.31 MB



(a) (b) (c)

Fig. 11. Overall extraction of Taiwan by YOLOv4 MobileNetv3 under different attention mechanisms.

TABLE IV
MODEL PERFORMANCE OF LANDSLIDE EXTRACTION USING DIFFERENT ATTENTION MECHANISMS IN BIJIE

Description	Precision (%)	Recall (%)	F1-score (%)	mAP (%)
Baseline	94.05	90.14	92.05	94.20
Baseline + ECA	93.50	91.44	92.46	94.94
Baseline + CBAM	94.05	92.22	93.13	95.15
Baseline + LPFRF	95.54	94.29	94.91	96.02

can be seen that YOLOv4 with LPFRF attention mechanism achieved the highest performance across all evaluation metrics. It had the highest precision (95.54%), recall (94.29%), F1-score (94.91%), and mAP (96.02%). YOLOv4 without any attention mechanism had the lowest mAP (94.20%) compared to the other models. This suggests that attention mechanisms can improve the overall performance of landslide detection by YOLOv4. YOLOv4 with ECA and CBAM attention mechanisms had similar performance, but slightly lower than YOLOv4 with LPFRF attention. Overall, using attention mechanisms can improve the YOLOv4 performance for landslide extraction. The LPFRF attention mechanism appears to be the most effective in improving both precision and recall, resulting in the highest F1-score and mAP. To test the model transferability, we tested YOLOv4 with different attention mechanisms in Taiwan. Table V gives the performance of landslide extraction using different attention mechanisms in Taiwan. It can be seen that the model using LPFRF attention mechanism achieved the highest performance in terms of OA (84.58%). It also had the highest Kappa score (63.24%), which is a statistical measure of interrater agreement that takes into account chance agreement. The models using ECA and CBAM attention mechanisms had similar OA scores (77.89% and 72.21%, respectively), but the ECA model had a higher Kappa score (53.83% compared to 44.26%). The model without any attention mechanism had an OA of 76.74% and a Kappa score of 50.00%. Thus, using attention mechanisms can improve the OA of the model for landslide detection. These results are corresponding to Fig. 11, demonstrating that the designed LPFRF attention mechanism appears to be the most effective in improving both OA and interrater agreement.

Table VI gives the parameter size of YOLOv4 with different backbones. It can be seen that YOLOv4 with no attention mechanism has the smallest size of 44.74 MB. The models with ECA and CBAM attention mechanisms have a slightly larger size of 48.24 MB and 44.74 MB, respectively. The YOLOv4 model with LPFRF attention mechanism has a size of 45.64 MB, which

TABLE V
MODEL PERFORMANCE OF LANDSLIDE EXTRACTION USING DIFFERENT ATTENTION MECHANISMS IN TAIWAN

Description	Overall (%)	Kappa (%)	User's accuracy (%)		Producer's accuracy (%)	
			landside	non-landslide	landslide	non-landslide
Baseline	76.74	50.00	81.78	68.15	81.43	68.60
Baseline + ECA	77.89	53.83	93.12	59.02	73.79	87.38
Baseline + CBAM	72.21	44.26	78.54	65.60	70.45	74.55
Baseline + LPFRF	84.58	63.24	91.13	70.34	86.99	78.46

TABLE VI
SIZE OF YOLOV4 WITH FOUR ATTENTION MECHANISMS

Baseline	Baseline + ECA	Baseline + CBAM	Baseline + LPFRF
44.74 MB	48.24 MB	44.74 MB	45.64 MB

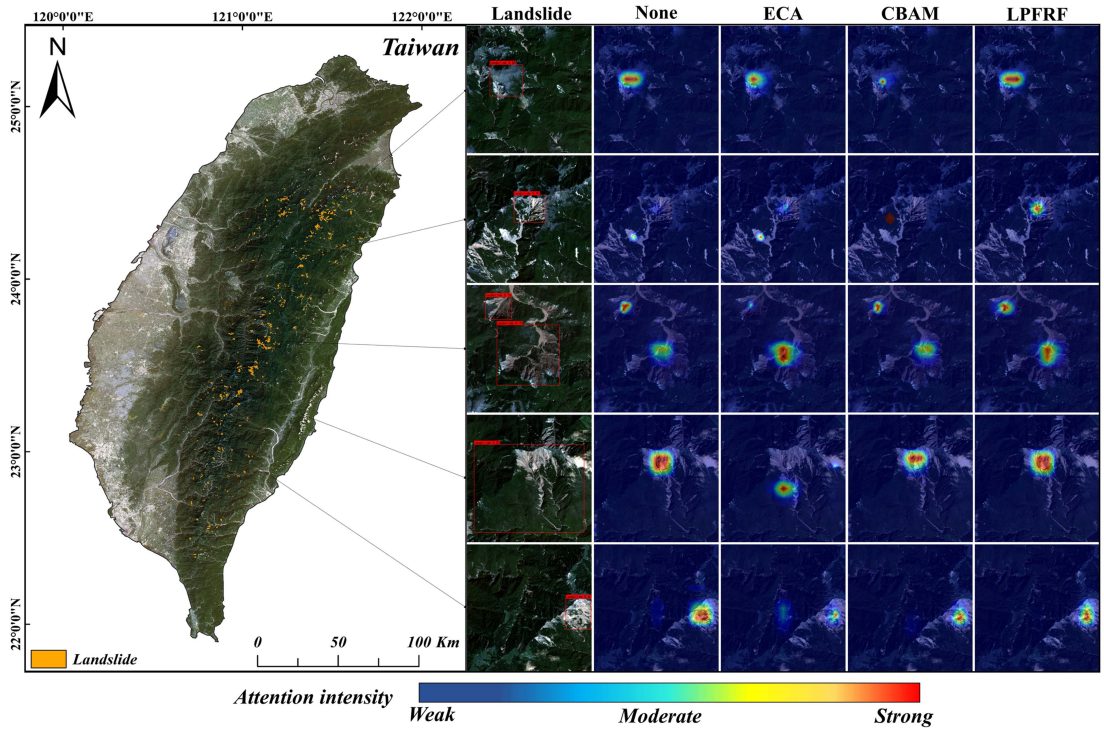


Fig. 12. Visualization of network feature extraction under different attention mechanisms of heat map.

is slightly larger than the model with no attention mechanism. Overall, the differences in model size between the different attention mechanisms are relatively small, with the largest model being only 3.5 MB larger than the smallest model. Thus, the designed LPFRF significantly increase the YOLOv4 detection accuracy and the generalization ability with a slight cost of parameter size increment.

In this article, a heatmap visualization tool [57] was used to visualize the landslide features extracted by different models with and without attention mechanisms. This tool allows us to observe the specific locations where the model is focused. Results are presented in Fig. 12, which shows that the attention mechanisms from the control group (e.g., none, ECA, and CBAM) occasionally focused on nonlandslide areas, rather than the landslide body itself. On the other hand, the designed LPFRF attention mechanism focused on the landslide body in all five scenarios, suggesting its robustness to various background disturbances. Through the joint consideration of training accuracy, generalization ability, parameter complexity and heatmap visualization, the designed LPFRF attention mechanism effectively improved the landslide detection performance and demonstrated strong generalization ability.

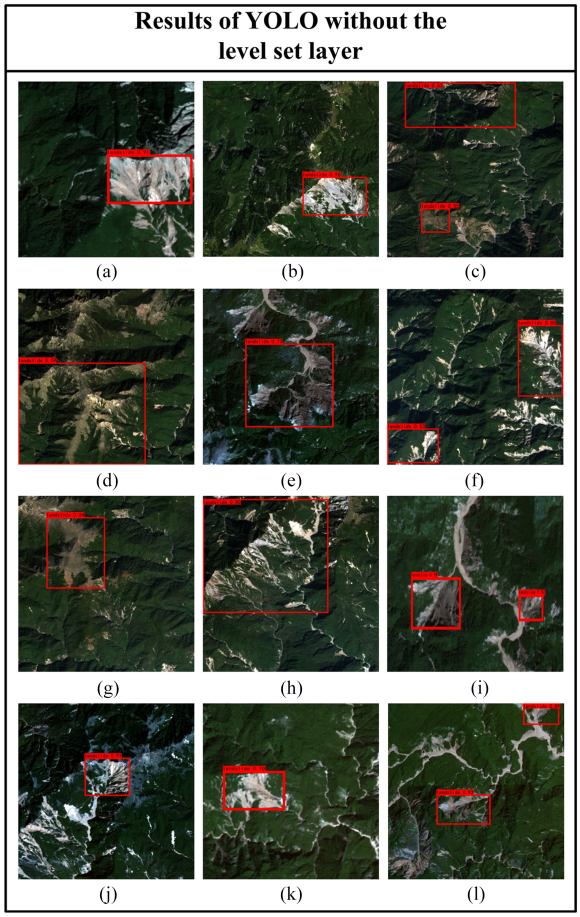


Fig. 13. Landslide detection results of LA-YOLO-LLL without the level set layer in 12 subareas of Taiwan.

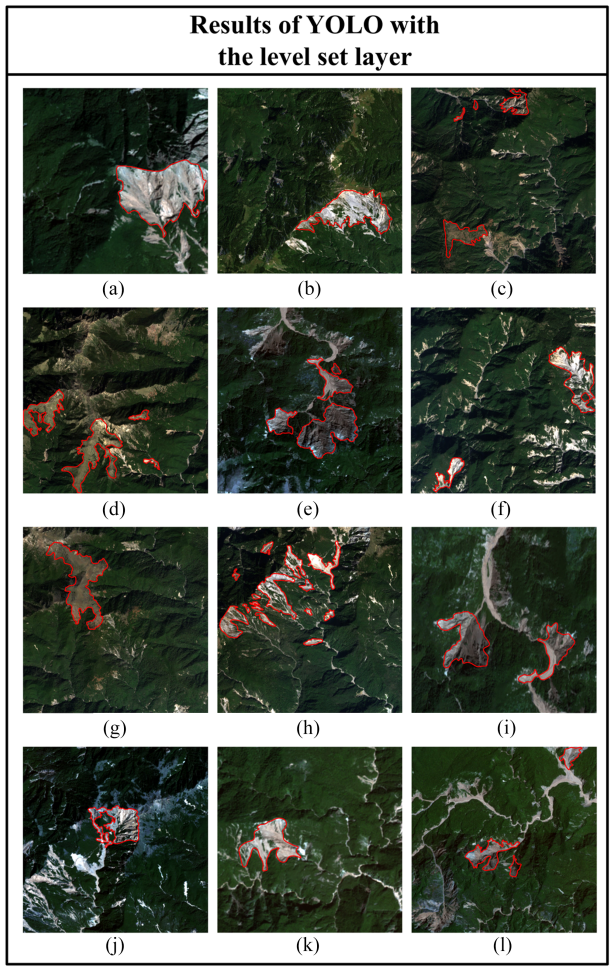


Fig. 14. Landslide detection results of LA-YOLO-LLL with the level set layer in 12 subareas of Taiwan.

C. Evaluation of the Presented Method

To assess the effectiveness of LA-YOLO-LLL with the level set layer, this article conducted a comparison between YOLO with and without the level set layer in detecting landslides in 12 sub-areas of Taiwan, covering a total area of 372.19km^2. Figs. 13 and 14 illustrate the landslide detection results using YOLO without and with the level set layer, respectively. The primary difference between the two is that Fig. 13 represents landslides using bounding boxes, while Fig. 14 represents landslides using accurate boundaries. The use of accurate boundaries would facilitate the calculation of several critical landslide parameters, including width, length, and size. In addition, in Fig. 13, the bounding box may contain multiple landslides [e.g., Fig. 13(c)–(i)], while each boundary in Fig. 14 represents only one landslide. In rare cases, YOLO without the level set layer may underestimate the landslide area, see Figs. 13(i) and 14(i).

The quantitative comparison results are given in Table VII, which shows that YOLO with the level set layer detected a significantly larger number of landslides (40) compared to YOLO without the level set layer (16). Moreover, the total area of detected landslides was much smaller for YOLO with the level set layer (22.50 km^2) than for YOLO without the level set layer (68.51km^2). This indicates that YOLO without the level set layer detected both landslide and nonlandslide areas, leading to overestimation of landslide areas. In contrast, the level set layer integrated into YOLO improved the accuracy of individual landslide boundary detection, resulting in an improved performance in detecting both the number and area of landslides. Thus, our analysis indicate that the level set layer is an effective approach to enhance the performance of YOLO in landslide detection.

D. Comparison With Benchmark Landslide Detection Algorithms

To validate the performance of the presented method, we compared with two benchmark deep learning based landslide detection algorithms (U-Net [58], and PSPNet [59]). Fig. 15 shows the landslide detection results of three models in Nantou, Taiwan. To provide a clearer illustration of the details, only the extraction results in Nantou are presented, rather than the entire region of Taiwan. The results demonstrate that the presented method can extract more landslides accurately compared to U-Net and PSPNet. Moreover, the false negative and false positive rates of the presented method are lower than those of U-Net and PSPNet. However, the presented method also falsely detects

TABLE VII
COMPARISON OF YOLOV4 WITH AND WITHOUT THE LEVEL SET LAYER

ID	Image area (km^2)	Number of detected landslides		Area of detected landslides (km^2)		Area differences between two models (km^2)
		Model #1[1]	Model #2[2]	Model #1	Model #2	
a	12.18	1	1	1.42	1.30	0.13
b	39.44	1	2	3.02	2.46	0.55
c	39.44	2	4	6.79	0.44	6.34
d	39.44	1	4	15.85	4.36	11.49
e	33.01	1	4	7.59	2.62	4.97
f	39.44	2	3	5.97	1.43	4.54
g	39.44	1	1	4.99	4.18	0.81
h	39.44	1	14	17.23	2.05	15.19

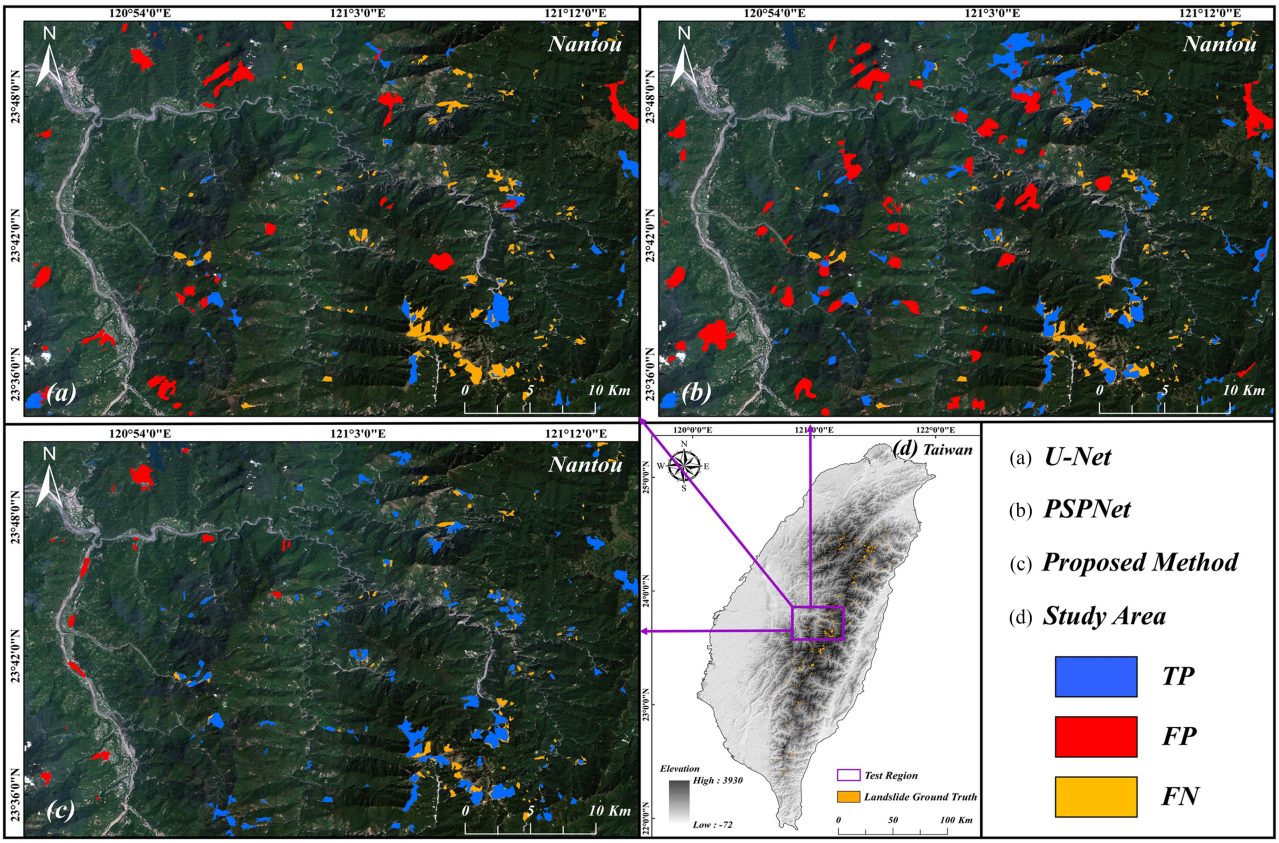


Fig. 15. Landslide detection results of three models in Nantou, Taiwan. (a)–(c) landslide detection results by U-Net, PSPNet, and LA-YOLO-LLL, respectively. (d) Study area.

terraced fields and estuarine shoal as landslides. This suggests that we can add negative samples in these soil types to further improve the performance of the presented method.

Table VIII gives the comparison results of three different models for landslide detection in Bijie. The proposed method achieved the highest scores for all four metrics, with a precision of 95.54%, recall of 94.29%, F1-score of 94.91%, and mAP of 96.02%. U-Net and PSPNet also performed well, but the proposed method outperformed both models with a significant margin. U-Net achieved a precision of 90.42%, recall of 87.58%, F1-score of 88.98%, and mAP of 93.64%. PSPNet achieved a precision of 89.52%, recall of 87.33%, F1-score of 88.41%, and mAP of 93.50%.

To compare the transferability performance, we applied three models to detect landslides in Taiwan. Table IX and Fig. 13 provide a comparison of three different models for landslide detection in Taiwan, including U-Net, PSPNet, and the proposed method. The results demonstrate that the proposed method achieved the highest OA and Kappa scores of 84.58% and 63.24%, respectively. The U-Net and PSPNet models also achieved reasonable accuracy, but their performance was lower than that of the proposed method. Specifically, the U-Net model

TABLE VIII
COMPARISON WITH TWO EXISTING METHODS OF LANDSLIDE DETECTION IN BIJIE

Model	Precision (%)	Recall (%)	F1-score (%)	mAP (%)
U-Net	90.42	87.58	88.98	93.64
PSPNet	89.52	87.33	88.41	93.50
Proposed	95.54	94.29	94.91	96.02

TABLE IX
COMPARISON WITH TWO EXISTING METHODS OF LANDSLIDE DETECTION IN TAIWAN

Model	Overall (%)	Kappa (%)	User's accuracy (%)		Producer's accuracy (%)	
			landside	non landslide	landslide	non landslide
U-Net	65.78	35.03	62.23	67.19	43.03	81.71
PSPNet	62.59	26.80	51.18	69.89	52.11	69.10
Proposed	84.58	63.24	91.13	70.34	86.99	78.46

TABLE X
COMPARISON OF THE PARAMETER SIZE OF THREE MODELS

U-Net	PSPNet	Proposed
167.59 MB	178.17 MB	44.74 MB

achieved an OA of 65.78% and a Kappa score of 35.03%, while the PSPNet model achieved an OA of 62.59% and a Kappa score of 26.80%. This suggests that the presented method archives the best generalization ability among three models. The findings in Table IX are consistent with Fig. 15.

Table X gives a comparison of the parameter size of three models (U-Net, PSPNet, and the proposed method) for landslide detection. The results demonstrate that the proposed method has a significantly smaller parameter size of 44.74 MB, compared to U-Net and PSPNet, which have parameter sizes of 167.59 and 178.17 MB, respectively. This suggests that the proposed method may be more efficient and faster to run, making it a promising solution for real-time landslide detection applications. Thus, the proposed method demonstrated superior performance compared to U-Net and PSPNet in terms of detection accuracy, transferability, and parameter size.

V. DISCUSSION AND CONCLUSION

Within the domain of deep learning-oriented landslide detection, the performance of the suggested model is subject to various determinants, such as experimental situations and dataset disparities. Evaluating model performance under diverse conditions is essential for ensuring adaptability and robustness. Our experimental findings indicate that, although different scenarios and datasets can cause performance variations, the suggested model exhibits considerable resilience. Future research could benefit from employing a wider range of diverse landslide image patches from various study areas and multisensor data sources, as well as considering additional spectral features and terrain factors for model training to enhance landslide detection performance [24], [60].

Optimal model performance necessitates careful parameter selection. Our experiments reveal that the temporal step and Gaussian kernel scale within the Level Set Layer significantly affect performance. Configuring these parameters to 15 and 3, respectively, corresponds to optimal settings for feature extraction from satellite imagery [12], [52], [61], [62], [63], [64]. With respect to the optimizer and learning rate, the Adam optimizer with a learning rate ranging from 0.001 to 0.0001 is considered effective [67], [68]. Our research introduces a dynamic learning rate strategy that enables autonomous adjustment during training within this range, adapting the learning rate according to model performance and accommodating dataset variations.

In this article, we proposed a LA-YOLO-LLL for accurate individual landslide boundary detection from optical satellite images. Our study revealed that the choice of backbone greatly influences the accuracy and generalization ability of the landslide detection models, and we introduced MobileNetv3 as the backbone of LA-YOLO-LLL due to its high transferability and low complexity. In addition, the designed attention mechanism outperformed benchmark attention mechanisms by focusing on the landslide body, and thus improved the landslide detection accuracy. Integration of level set layer into the head of the presented method proved effective in accurately delineating the landslide boundary, making it a suitable method for extracting landslides from optical satellite images. While this article focused on landslide detection from Sentinel-2 optical satellite images, small landslides remain a challenge to detect at this spatial resolution. Therefore, joint use of crowdsourcing data, including satellite SAR images, aerial photographs, and social media, would be an interesting research direction.

ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous reviewers for their valuable time on this manuscript.

REFERENCES

[1] V. Svalova, V. Zaalishvili, G. Ganapathy, A. Nikolaev, and D. Melkov, "Landslide risk in mountain areas," *Geol. South Russia*, vol. 2, 2019, Art. no. 012009.

[2] T. Davies, "Landslide hazards, risks, and disasters: Introduction," in *Landslide Hazards, Risks, and Disasters*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 1–16.

[3] D. Cruden, *Landslide Risk Assessment*. London, U.K.: Routledge, 2018.

[4] K. Sassa, H. Fukuoka, F. Wang, and G. Wang, *Landslides: Risk analysis and Sustainable Disaster Management*. Berlin, Germany: Springer, 2006.

[5] N. Casagli, E. Intrieri, V. Tofani, G. Gigli, and F. Raspini, "Landslide detection, monitoring and prediction with remote-sensing techniques," *Nature Rev. Earth Environ.*, vol. 4, no. 1, pp. 51–64, 2023.

[6] F. Guzzetti, A. C. Mondini, M. Cardinali, F. Fiorucci, M. Santangelo, and K.-T. Chang, "Landslide inventory maps: New tools for an old problem," *Earth-Sci. Rev.*, vol. 112, no. 1-2, pp. 42–66, 2012.

[7] J. R. Mantovani et al., "Novel landslide susceptibility mapping based on multi-criteria decision-making in Ouro Preto," *Brazil J. Geovisualization Spatial Anal.*, vol. 7, no. 1, 2023, Art. no. 7.

[8] A. V. Thomas et al., "Landslide susceptibility zonation of Idukki district using gis in the aftermath of 2018 Kerala floods and landslides: A comparison of AHP and frequency ratio methods," *J. Geovisualization Spatial Anal.*, vol. 5, no. 2, 2021, Art. no. 21.

[9] Z. Z. YiHe et al., "A unified network of information considering superimposed landslide factors sequence and pixel spatial neighbourhood for landslide susceptibility mapping," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 104, no. 2021, 2021, Art. no. 102508.

[10] Z.-X. Hu, C.-L. Wang, Z.-G. Zhou, and C.-R. Li, "Using recovery rate and SVM to detect landslides in medium satellite images time series: A case study in Ludian, China," in *Proc. 2017 Int. Conf. Wireless Commun., Netw. Appl.*, 2017, pp. 240–244.

[11] Q. Hu, Y. Zhou, S. Wang, F. Wang, and H. Wang, "Improving the accuracy of landslide detection in 'off-site' area by machine learning model portability comparison: A case study of Jiuzhaigou earthquake, China," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2530.

[12] Z. Li, W. Shi, P. Lu, L. Yan, Q. Wang, and Z. Miao, "Landslide mapping from aerial photographs using change detection-based Markov random field," *Remote Sens. Environ.*, vol. 187, pp. 76–90, 2016.

[13] S. Tavakkoli Piralilou et al., "Landslide detection using multi-scale image segmentation and different machine learning models in the higher Himalayas," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2575.

[14] D. Meghanadh, V. K. Maurya, M. Kumar, and R. Dwivedi, "Automatic detection of landslides based on machine learning framework," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 8460–8463.

[15] A. Farhadi and J. Redmon, "Yolov3: An incremental improvement," in *Computer Vision and Pattern Recognition*, vol. 1804, Berlin, Germany: Springer, pp. 1–6, 2018.

[16] S. Ji, D. Yu, C. Shen, W. Li, and Q. Xu, "Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks," *Landslides*, vol. 17, pp. 1337–1352, 2020.

[17] T.-A. Bui, P.-J. Lee, K.-Y. Lum, C. Loh, and K. Tan, "Deep learning for landslide recognition in satellite architecture," *IEEE Access*, vol. 8, pp. 143665–143678, 2020.

[18] 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. 11417–11428, Oct. 2021.

[19] H. Cai, T. Chen, R. Niu, and A. Plaza, "Landslide detection using densely connected convolutional networks and environmental conditions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5235–5247, May 2021.

[20] L. Bragagnolo, L. Rezende, R. da Silva, and J. Grzybowski, "Convolutional neural networks applied to semantic segmentation of landslide scars," *Catena*, vol. 201, 2021, Art. no. 105189.

[21] 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, 2022, Art. no. 4357.

[22] Y. Zhou, H. Wang, R. Yang, G. Yao, Q. Xu, and X. Zhang, "A novel weakly supervised remote sensing landslide semantic segmentation method: Combining cam and Cyclegan algorithms," *Remote Sens.*, vol. 14, no. 15, 2022, Art. no. 3650.

[23] S. O. Y. Amankwah et al., "Landslide detection from bitemporal satellite imagery using attention-based deep neural networks," *Landslides*, vol. 19, no. 10, pp. 2459–2471, 2022.

[24] W. Lu, Y. Hu, Z. Zhang, and W. Cao, "A dual-encoder u-net for landslide detection using sentinel-2 and DEM data," *Landslides*, vol. 31, pp. 1–13, 2023.

[25] H. Chen et al., "A landslide extraction method of channel attention mechanism u-net network based on sentinel-2a remote sensing images," *Int. J. Digit. Earth*, vol. 16, no. 1, pp. 552–577, 2023.

[26] X. Liu et al., "Feature-fusion segmentation network for landslide detection using high-resolution remote sensing images and digital elevation model data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 4500314.

[27] Y. Ju et al., "Loess landslide detection using object detection algorithms in Northwest China," *Remote Sens.*, vol. 14, no. 5, 2022, Art. no. 1182.

[28] Y. Liu et al., "Study of the automatic recognition of landslides by using InSAR images and the improved mask R-CNN model in the eastern Tibet plateau," *Remote Sens.*, vol. 14, no. 14, 2022, Art. no. 3362.

[29] A. Tanatipuknon et al., "Study on combining two faster R-CNN models for landslide detection with a classification decision tree to improve the detection performance," *J. Disaster Res.*, vol. 16, no. 4, pp. 588–595, 2021.

[30] R. Yang, F. Zhang, J. Xia, and C. Wu, "Landslide extraction using mask r-CNN with background-enhancement method," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2206.

[31] Q. Liu, T. Wu, Y. Deng, and Z. Liu, "Se-yolov7 landslide detection algorithm based on attention mechanism and improved loss function," *Land*, vol. 12, no. 8, 2023, Art. no. 1522.

[32] C. Wu, "Certainty factor analyses and spatiotemporal characteristics of landslide evolution: Case studies in the Chishan river watershed in Taiwan," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 7, 2022, Art. no. 382.

[33] C.-Y. Wu and S.-Y. Lin, "Event-based landslide susceptibility models in Shihmen watershed, Taiwan: Accounting for the characteristics of rainfall events," *Environ. Monit. Assessment*, vol. 194, no. 6, 2022, Art. no. 405.

[34] C. Corbane et al., "A global cloud free pixel-based image composite from sentinel-2 data," *Data Brief*, vol. 31, 2020, Art. no. 105737.

[35] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[36] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," Accessed: Jan. 15, 2024.

[37] L. Wu, R. Liu, G. Li, J. Gou, and Y. Lei, "Landslide detection methods based on deep learning in remote sensing images," in *Proc. 29th Int. Con. Geoinformatics*, Beijing, China, 2022, pp. 1–4.

[38] P. Hurtik, V. Molek, J. Hula, M. Vajgl, P. Vlasanek, and T. Nejezchleba, "Poly-yolo: Higher speed, more precise detection and instance segmentation for yolov3," *Neural Comput. Appl.*, vol. 34, no. 10, pp. 8275–8290, 2022.

[39] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series," *Comput. Vis. Pattern Recognit.*, 2021.

[40] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, Accessed: Jan. 15, 2024, pp. 7464–7475.

[41] C. Li et al., "Yolov6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.

[42] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, Art. no. 25125.

[43] X. Yang et al., "Remote sensing image detection based on yolov4 improvements," *IEEE Access*, vol. 10, pp. 95527–95538, 2022.

[44] M. Zha, W. Qian, W. Yi, and J. Hua, "A lightweight yolov4-based forestry pest detection method using coordinate attention and feature fusion," *Entropy*, vol. 23, no. 12, 2021, Art. no. 1587.

[45] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4507–4515.

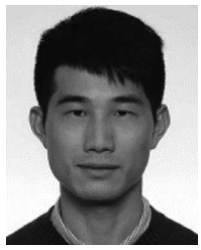
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[47] W. Wang et al., "Efficient and accurate arbitrary-shaped text detection with pixel aggregation network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8440–8449.

[48] A. Howard et al., "Searching for mobilenetv3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.
[49] C. Wickens, "Attention: Theory, principles, models and applications," *Int. J. Hum.–Comput. Interaction*, vol. 37, no. 5, pp. 403–417, 2021.
[50] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.
[51] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[52] Z. Li, W. Shi, Q. Wang, and Z. Miao, "Extracting man-made objects from high spatial resolution remote sensing images via fast level set evolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 883–899, Feb. 2014.
[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[54] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1580–1589.
[55] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
[56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations, Comput. Biol. Learn. Soc.*, 2015. Accessed: Jan. 15, 2024.
[57] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.
[58] 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. Intervention*, 2015, pp. 234–241.
[59] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.
[60] 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, Oct. 2022, Art. no. 5633017.
[61] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
[62] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM J. Numer. Anal.*, vol. 29, no. 1, pp. 182–193, 1992.
[63] Y. Shi and W. C. Karl, "A real-time algorithm for the approximation of level-set-based curve evolution," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 645–656, May 2008.
[64] Y. Bazi, F. Melgani, and H. D. Al-Sharari, "Unsupervised change detection in multispectral remotely sensed imagery with level set methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3178–3187, Aug. 2010.
[65] G. Cao, Y. Li, Y. Liu, and Y. Shang, "Automatic change detection in high-resolution remote-sensing images by means of level set evolution and support vector machine classification," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6255–6270, 2014.
[66] Z. Li, W. Shi, S. W. Myint, P. Lu, and Q. Wang, "Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method," *Remote Sens. Environ.*, vol. 175, pp. 215–230, 2016.
[67] P. Hu, B. Shuai, J. Liu, and G. Wang, "Deep level sets for salient object detection," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2017, pp. 2300–2309.
[68] Z. Wang, D. Acuna, H. Ling, A. Kar, and S. Fidler, "Object instance annotation with deep extreme level set evolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7500–7508.

Yueheng Yang received the B.S. degree in geomatics engineering from the Inner Mongolia Agricultural University, Hohhot, China, in 2022. He is currently working toward the M.S. degree in geomatics engineering with the Central South University, Changsha, China.

His research interest includes landslide extraction, landslide mapping, image segmentation, and computer vision.



Zelang Miao (Member, IEEE) received the B.S. degree in surveying engineering and the M.S. degree in geodesy and surveying engineering from the China University of Mining and Technology, Xuzhou, China, in 2009 and 2014, respectively, and the Ph.D. degree in satellite image processing from The Hong Kong Polytechnic University, Hong Kong, in 2016.

He was a visiting Ph.D. Student with the Department of Industrial and Information Engineering, University of Pavia, Pavia, Italy, in 2015. He is currently an Associate Professor with the School of Geoscience and Info-Physics, Central South University, Changsha, China. He has authored or co-authored more than 30 papers on international and peer-reviewed scientific journals. His research interests include pattern recognition, land cover and land use mapping, and global/regional urbanization.

Hua Zhang received the Doctoral degree in cartography and geographical information engineering from the China University of Mining and Technology, Xuzhou, China, in 2012.

He is currently an Associate Professor of GIS and Remote Sensing with the School of Environment and Spatial Informatics, China University of Mining and Technology. He has authored or coauthored more than 40 peer-reviewed articles in international journals, such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *ISPRS Journal of Photogrammetry and Remote Sensing*. His current research interests include multi/hyperspectral and high-resolution remotely sensed images processing.

Bing Wang (Member, IEEE) was born in Longnan, Gansu Province, China in 1999. He received the B.S. degree in engineering from Inner Mongolia Agricultural University, Hohhot, China, in 2021. He is currently working toward the master's degree in surveying and mapping engineering with the China University of Mining and Technology, Beijing, China.

His interest includes remote sensing environmental monitoring and phonological analysis.

Lixin Wu (Member, IEEE) received the B.S. degree in mining survey from the China University of Mining and Technology, Xuzhou, China, in 1988, and the M.S. and Ph.D. degrees in geomatics from the China University of Mining and Technology, Beijing, China, in 1991 and 1997, respectively.

He is currently a Professor with the School of Geoscience and Info-Physics, Central South University, Changsha, China. His research interests include geohazards synergic observation, remote sensing rock mechanic, and geospatial informatics.