



OPEN

InSAR-YOLOv8 for wide-area landslide detection in InSAR measurements

Ruopo Ma¹, Haiyang Yu^{1,2}✉, Xuejie Liu¹, Xinru Yuan¹, Tingting Geng¹ & Pengao Li¹

InSAR monitoring technology is widely used in investigating landslide hazards. Leveraging object detection algorithms to quickly extract landslide information from Wide-Area InSAR measurements is of great significance. Our InSAR-YOLOv8, an algorithm that automatically detects landslides from InSAR measurements, addresses the low accuracy and suboptimal detection performance of existing network models. In this method, we first design and add a detection head specifically targeting small-scale objects. This improvement enhances the model's ability to extract features across different scales and strengthens its capability to detect landslides of varying sizes. We also replace the original C2f module with the lighter C2f_Faster module to process information more efficiently, making the model lighter and more efficient. Finally, the SIoU loss function replaces the ClIoU loss function to improve the bounding box regression and enhance detection accuracy. Our results show that the proposed algorithm achieves a 97.41% mAP50, a 66.47% mAP50:95, and a 92.06% F1 score on the InSAR landslide dataset, while reducing the number of parameters by 25%. Compared with YOLOv8 and other advanced models (YOLOvX, Faster R-CNN, etc.), our model exhibits distinct advantages and possesses a wider range of potential applications in InSAR measurement for landslide detection.

Keywords Improved YOLOv8, InSAR measurements, Automatic detection, Landslide

Landslides, as a prevalent natural disaster, pose significant threats to human life, property, and the environment. Efficiently identifying and continuously monitoring large-scale landslides is crucial for evaluating places that may pose geological hazards and for providing necessary technical assistance in order to prevent and warn against landslide disasters^{1–5}. Landslide detection refers to the process of monitoring, identifying, analyzing, and assessing the occurrence and development of landslides through various methods, such as ground-based monitoring and remote sensing technologies. Traditional methods of landslide detection, such as the GPS measurements and ground levelling^{6,7}, offer high precision in local areas but are limited by geographical constraints, making large-scale landslide detection and analysis challenging⁸. Nowadays, the remote sensing technology has gradually become a research hotspot in landslide disaster prevention and control⁹.

Landslides can be visually interpreted through optical remote sensing or identified by monitoring subtle surface deformations using the microwave remote sensing technology^{10–12}. Although optical remote sensing typically offers a higher spatial resolution, it is limited by weather conditions and the changes in illumination. Optical remote sensing imagery can be used to visually interpret landslides with significant morphological features and deformation markers, making it particularly effective for rapid landslides accompanied by noticeable surface damage. Optical imagery can clearly capture the boundaries and damage characteristics of the landslide area, but it is limited by spatial resolution and weather conditions. Additionally, it is difficult to determine whether the landslide is actively deforming, and it is nearly impossible to detect the subtle deformations of slow-moving landslides. This restricts the application of optical remote sensing in dynamic monitoring and in detecting the deformation of slow-moving landslides^{13–16}. Synthetic Aperture Radar Interferometry (InSAR) detects ground deformation and displacement to capture active landslides, offering advantages such as high precision, high resolution, and all-weather capability. InSAR can continuously track landslide displacements, allowing it to distinguish between stable and unstable landslides within a wide area^{17–20}. InSAR is highly effective in long-term landslide detection, capable of detecting various types of landslides, including potential landslide areas, reactivated landslides, and slow-moving landslides, while also providing early warnings for potentially hazardous areas^{21–25}. Although it may be affected by decorrelation when detecting rapidly deforming landslides, InSAR remains one of the most effective high-precision quantitative techniques^{26,27}. With the advancement of technology and the

¹School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China.

²Key Laboratory of Mine Spatio-Temporal Information and Ecological Restoration, Henan Polytechnic University Ministry of Natural Resources, Jiaozuo 454000, China. ✉email: yuhaiyang@hpu.edu.cn

expansion of application domains, several InSAR technologies, including Differential Interferometric Synthetic Aperture Radar (D-InSAR), Small Baseline Subsets InSAR (SBAS-InSAR) and Permanent Scatterer InSAR (PS-InSAR), have been developed for landslide detection^{28–33}. Among them, the SBAS-InSAR technique effectively addresses many of the limitations of traditional D-InSAR methods. Compared to PS-InSAR, SBAS-InSAR is better suited for monitoring large-scale and continuous deformation, with superior noise resistance. As a result, this technique has been widely applied in the study of landslide-related geological hazards^{34–36}.

InSAR technology has great potential for landslide detection, but it confronts many obstacles. Conventional InSAR measurements often necessitate manual intervention and analysis, particularly when dealing with extensive regions and vast image datasets, which can lead to errors and reduced efficiency³⁷. This limitation hinders the ability to perform rapid and accurate large-area identifications. As an emerging technology, deep learning have shown broad application prospects in recent years in the fields of disaster detection, and computer vision³⁸. Therefore, this study aims to introduce deep learning techniques into the field of landslide detection, developing an automated method for detecting landslides in InSAR measurements. This method can quickly detect landslide areas experiencing deformation from large-scale InSAR images, thereby overcoming the inefficiencies associated with the traditional manual InSAR interpretations.

In recent years, some scholars have achieved notable results in identifying geological hazards using InSAR measurements, machine learning and deep learning algorithms. For example, Anantrasiricha et al.³⁹ proposed a method based on large-scale InSAR images to automatically detect surface deformations caused by volcanic activity, demonstrating for the first time the capability of the neural networks in identifying volcanic surface deformations based on InSAR measurements. This method, which directly uses interferometric fringes for detection, is highly susceptible to noise interference, such as atmospheric delays. These noise sources can introduce errors in deformation detection results, thereby affecting the accuracy of the model. Sun et al. created a synthetic volcanic dataset with manually added noise. They then trained an encoder-decoder neural network on this dataset, aiming to mitigate noise interference in InSAR results by learning the noise characteristics⁴⁰. Brengman and Barnhart⁴¹ utilized the SarNet network to identify, locate, and categorize surface deformations in InSAR synthetic interferograms following earthquakes. These methods have shown success in detecting large-scale deformations from volcanic and seismic activities. However, they face challenges when applied to smaller-scale geological hazards like landslides and mine collapses. The smaller spatial scale and increased susceptibility to noise interference make accurate localization and effective detection more difficult. Wu et al. trained the PUNet network on a synthetic mining subsidence dataset, effectively addressing issues related to dense fringes and noise in SAR interferograms. Additionally, by utilizing a deeper model architecture, they improved the detection capability for medium- and small-scale targets⁴². Although these methods have reduced noise interference to some extent, they are all trained on synthetic datasets. As a result, they are unable to fully simulate the deformation characteristics of geological hazards observed in real InSAR measurements. Wu et al.⁴³ utilized the Distributed Scatterers InSAR (DS-InSAR) technology to acquire the deformation rate map of the Greater Bay Area in China, followed by an automatic detection and classification of ground subsidence using a multi-channel R-CNN network. However, this R-CNN network relies on a multi-stage structure and demands significant computational resources, making it difficult to meet the requirements for fast automated detection tasks. To address this issue, Zhao et al.⁴⁴ presented the Convolutional Block Attention Module (CBAM) and developed a convolutional neural network called DefoNet. This network is designed to identify geological hazards resulting from mining activities in Shanxi Province, China, using InSAR deformation measurements. The proposed method achieves a relatively lightweight model while maintaining high detection accuracy and efficiency across extensive monitoring areas. Lin et al.⁴⁵ proposed a network model based on the U2-Net to rapidly and effectively identify large scale subsidence areas using InSAR measurements. By doing so, the authors improved the model's noise resistance in extensive InSAR detection tasks. However, the aforementioned methods are predominantly based on network models using U-net or R-CNN architectures, which rely on complex structures and deep convolutional modules. These approaches demand high computational resources and leave considerable room for improvement in terms of accuracy. Landslide detection tasks require fast and accurate identification of potential hazardous areas over large regions, rendering low-efficiency models inadequate for meeting the demands of such tasks.

Summarizing the above, an analysis of numerous research cases reveals that the InSAR measurement detection tasks face challenges such as noise interference, complex model structures, low efficiency in large-scale recognitions, and poor performance in multi-scale feature extractions. Moreover, there is a paucity of studies on automatic landslide detection based on InSAR measurements and deep learning techniques. Existing models struggle to effectively preserve the multi-scale features and detailed information, failing to meet the requirements for extracting landslide deformations with noticeable differences and in complex morphologies⁴⁶. Any small landslide can potentially serve as a precursor to a larger-scale landslide, and the timely detection of these small landslides is critical for early warning and disaster prevention. Therefore, it is necessary to develop a model with a high detection accuracy, strong feature extraction capability, and a small number of parameters to address these challenges. Object detection algorithms can be categorized into two-stage and single-stage detection algorithms⁴⁷. In 2015, Girshick et al.⁴⁸ proposed the R-CNN, a seminal two-stage object detection algorithm. Building on this foundation, Girshick et al. and Ren et al. subsequently introduced the Fast R-CNN⁴⁹ and Faster R-CNN⁵⁰ algorithms, respectively. Building on these algorithms, scholars have subsequently developed numerous advanced two-stage object detection models^{51–53}. However, due to the relative complexity of two-stage models and their slower inference speeds, they are not well-suited to meet the requirements of this study. The YOLO (You Only Look Once) model, introduced by Joseph Redmon et al. in 2016, markedly differs from the traditional object detection methods⁵⁴. As a single-stage object detection model, YOLO simplifies the detection process by framing it as a single regression problem, enabling direct prediction of bounding boxes and class probabilities from the entire images in a single evaluation. This approach unifies the object detection process,

streamlines computational steps, and significantly enhances the detection efficiency and accuracy. In recent years, the YOLO series has progressively introduced more advanced versions, such as YOLOv5, YOLOv7, and YOLOv8^{55–60}. YOLOv8 builds upon the strengths of previous models by incorporating improved architecture, modules, and other advanced techniques. These enhancements optimize computational efficiency while improving performance and detection accuracy. It has been widely demonstrated to deliver superior results and greater stability across various complex scenarios and multi-scale object detection tasks^{61–64}. Thus, this study builds on the strengths of YOLOv8 to propose the InSAR-YOLOv8 model for the automatic detection of landslide-prone areas using InSAR measurements. The goal is to enhance the application of deep learning in InSAR-based landslide detection, offering a more efficient and accurate method for the rapid identification of large-scale landslides. The main contributions of our research are succinctly outlined as follows.

(1) Applying deep learning object detection algorithms to the field of large scale InSAR landslide detection for rapid identification of landslide hazard areas from large-scale InSAR images. This technique addresses the inefficiencies inherent in the conventional wide-area InSAR-based techniques for landslide detection, which necessitate manual interventions.

(2) To address the considerable variations in landslide sizes observed in InSAR measurements, a detection head is creatively reconstructed to improve its ability to extract information from landslides of different scales. A lightweight C2f_Faster module is developed to replace the original C2f module in the YOLOv8 model, to further improve the detection accuracy and reduce the number of parameters by 25%. The SIoU loss function is introduced as a substitute for the CIoU loss function, bolstering the regression ability of the bounding box and enhancing the detection accuracy.

(3) Compared to the other advanced detection algorithms such as Faster R-CNN, YOLOv3, and YOLOvX, the InSAR-YOLOv8 has fewer parameters, a better detection performance, and a higher recognition accuracy, demonstrating its significant advantages.

Methods

YOLOv8 model

YOLOv8 is the most recent version of the YOLO model introduced by Ultralytics in 2023. It has demonstrated exceptional accuracy in various tasks, including image classification, object detection, and instance segmentation^{65–67}. In this study, the smallest and fastest YOLOv8n configuration model is selected for training. As depicted in Fig. 1, the YOLOv8n architecture comprises three main components: a backbone, a neck, and a head.

YOLOv8 model's backbone network

YOLOv8 utilizes the backbone network primarily for feature extraction, with its model structure drawing inspiration from the CSPDarknet-53 network. The backbone network comprises the CBS (Conv + BatchNorm + SiLU), the C2f, and the SPPF (Spatial Pyramid Pooling-Fast) modules. The structure of the C2f and SPPF modules is illustrated in Fig. 2. The YOLOv8 introduces a more lightweight C2F module to replace the Cross-Scale Part (CSP) module in the original backbone network. The C2f module incorporates the

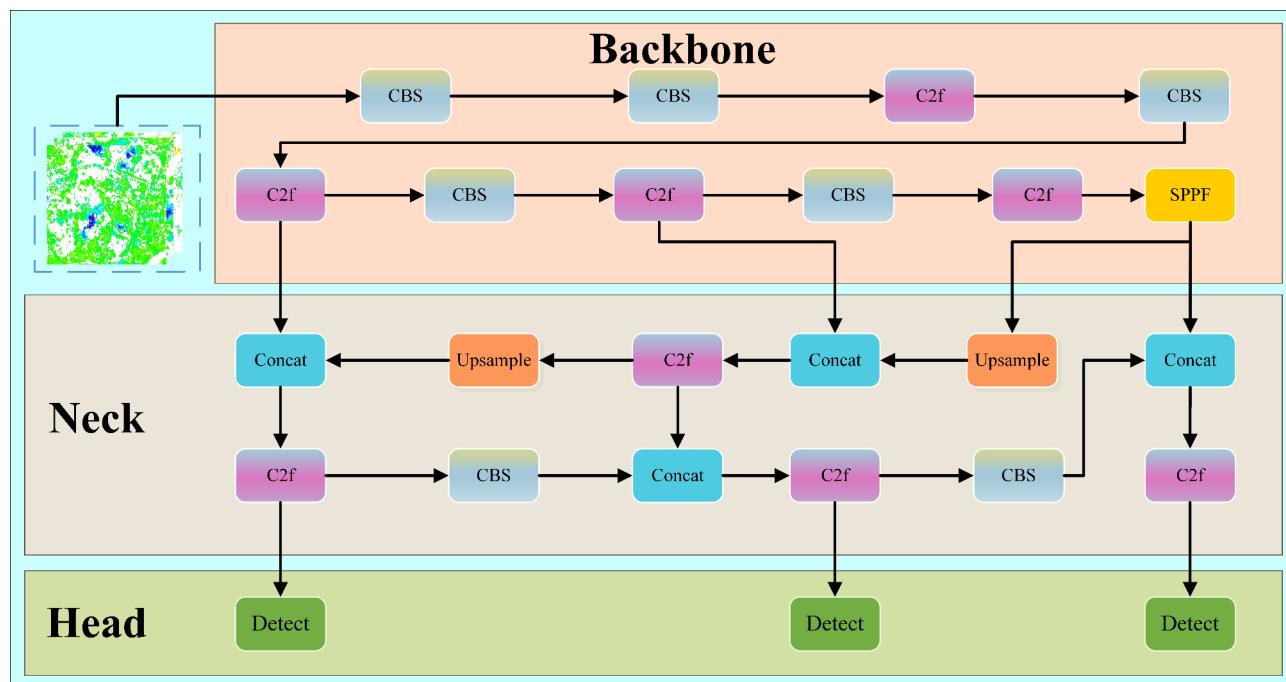


Fig. 1. Structure of the yolov8 network.

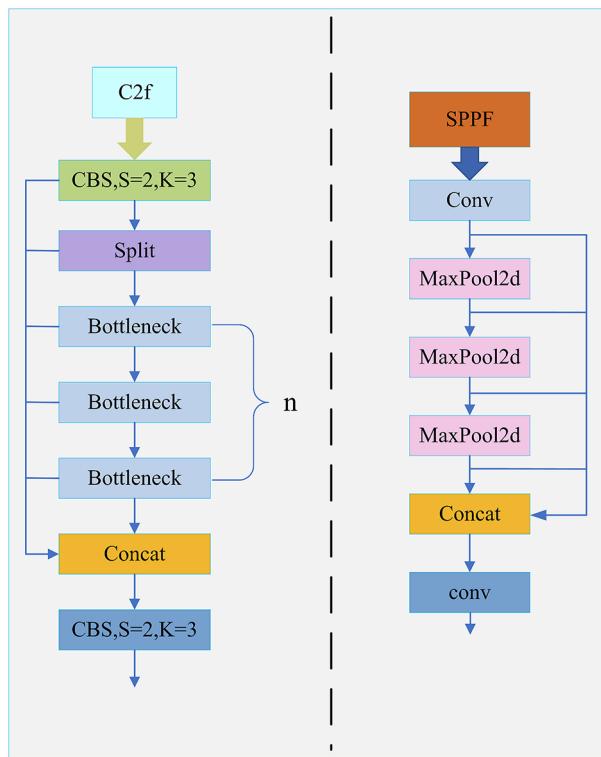


Fig. 2. C2f module structure (left) and SPPF module structure (right).

Bottleneck units and utilizes the gradient shunt connections, effectively enriching the information flow of the feature extraction network. This enhancement maintains its lightweight nature while significantly improving convergence speed and effectiveness.

The CBS module integrates convolution, batch normalization, and the SiLU activation function to construct an efficient convolutional neural network unit. First, input data passes through the convolutional layer to extract features information. Subsequently, batch normalization normalizes the feature distributions within mini-batches, accelerating model training and enhancing stability. Finally, the SiLU activation function is employed to enhance the model's non-linear capabilities, improving the network's ability to represent complex relationships. The SPPF layer is positioned at the tail of the network structure, serving as a pivotal component for multi-level feature aggregation. By performing multi-scale pooling operations, it effectively captures fine-grained local features and extracts global contextual information, thereby facilitating a thorough analysis of the spatial structures within feature maps. This seamless integration of local and global information significantly enhances the model's scalability and adaptability to varying target sizes, making it highly effective in multi-scale object detection scenarios. Furthermore, the SPPF layer expands the receptive field by simulating pooling effects over larger regions, thereby improving the network's capacity to capture rich contextual dependencies and elevating its detection performance in complex environments.

YOLOv8 model neck module

Inspired by PANet, the YOLOv8 neck region adopts a PAN-FPN structure to build the feature pyramid for the YOLO. The upsampling layer restores the spatial resolution of feature maps through specific methods, gradually reconstructing the feature maps from low resolution to a resolution close to that of the original input image. This compensates for the loss of spatial detail caused by downsampling, ensuring that the model captures more local features. The function of the concatenation layer is to merge feature maps from different sources or scales along the channel dimension. By integrating high-level semantic information with low-level spatial information, it provides a more comprehensive representation of the target features. This enhances the network's feature representation capability and improves the model's robustness and accuracy when handling complex scenes or multi-scale targets. In this adaptation, the convolutional structure used in the PAN-FPN up-sampling phase in YOLOv5 is removed, and the C3 module is replaced with the C2f module as the new neck structure. The PAN-FPN structure is designed to construct a hybrid network with both top-down and bottom-up information flow. This allows for the complementary integration of shallow positional information with deep semantic information through feature fusion, maintaining the original performance while achieving a lightweight model. It facilitates comprehensive feature fusion between multi-scale information, enhancing feature diversity and completeness.

YOLOv8 model head module

The YOLOv8 head adopts a decoupled structure, where the decoupled head architecture separates object classification from the bounding box prediction. It employs two independent branches for object classification

and bounding box regression, using distinct loss functions for these two tasks. The YOLOv8 introduces a departure from the traditional Anchor-Based approach, instead embracing the Anchor-Free concept, which offers advantages in detecting objects with irregular aspect ratios. As an anchor-free detection model, it simplifies the specification of positive and negative samples. Furthermore, it incorporates a task alignment allocator to dynamically assign samples, thereby enhancing the model's detection accuracy and robustness.

YOLOv8 model loss function

The loss function of the YOLOv8 consists of two main components: category loss and location loss. The category loss employs the Varifocal Loss (VFL), as represented by Eq. (1)⁶⁸:

$$\text{VFL}(p, q) = \begin{cases} -q(q\log(p) + (1-q)\log(1-q)) & q > 0 \\ -\alpha p^\gamma \log(1-q) & q = 0 \end{cases} \quad (1)$$

Where q is the intersection-union ratio (IoU) between the prediction box and the real box, and p is the probability. The bounding box regression loss takes the form of the Complete IoU (CIoU)⁶⁹ supplemented by the Distribution Focal Loss (DFL), with the DFL defined by Eq. (2):

$$\text{DFL}(S_i, S_{i+1}) = -((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1})) \quad (2)$$

Where S_i is the output of its sigmoid, y_i and y_{i+1} represent the indices of the coordinate axis intervals, and y denotes the value of the label. VFL is an improved loss function over the cross-entropy loss, which better handles class imbalances and enhances detection accuracy. The CIoU Loss builds upon the DIoU (Distance-IoU) by introducing an innovative scale loss term, which handles overlaps between the predicted and the true bounding boxes more effectively. DFL tackles the problem of class imbalance and background class issues by enabling the network to quickly reach a distribution of positions closer to the desired target location.

YOLOv8 model improvement

In this study, a new object detection head is designed for the YOLOv8n model to address the specifics of landslide detection in InSAR data. The novel C2f_Faster module is introduced, with the SIoU as the loss function, to enhance the landslide detection accuracy. We refer to the improved model as InSAR-YOLOv8, and its network architecture is illustrated in Fig. 3.

Small target detection module design

The original YOLOv8 network model's feature extraction network comprises three feature detection layers: a 20×20 large-object detection head, a 40×40 medium-object detection head, and a 80×80 small-object detection head. However, the model often has a large downsampling multiple, making it challenging to capture the features of small objects from deeper feature maps. This frequently results in overlooked detections or subpar detection performance for diminutive items in datasets that include them. In the process of detecting landslides, in addition to identifying large landslide areas, smaller-scale landslides are also frequently observed. When utilizing YOLOv8 for identification purposes, it can be challenging to accurately extract the features of certain minor landslides, which may also become indistinguishable from the surrounding background. For the detection of geological hazards, any overlook can have significant consequences.

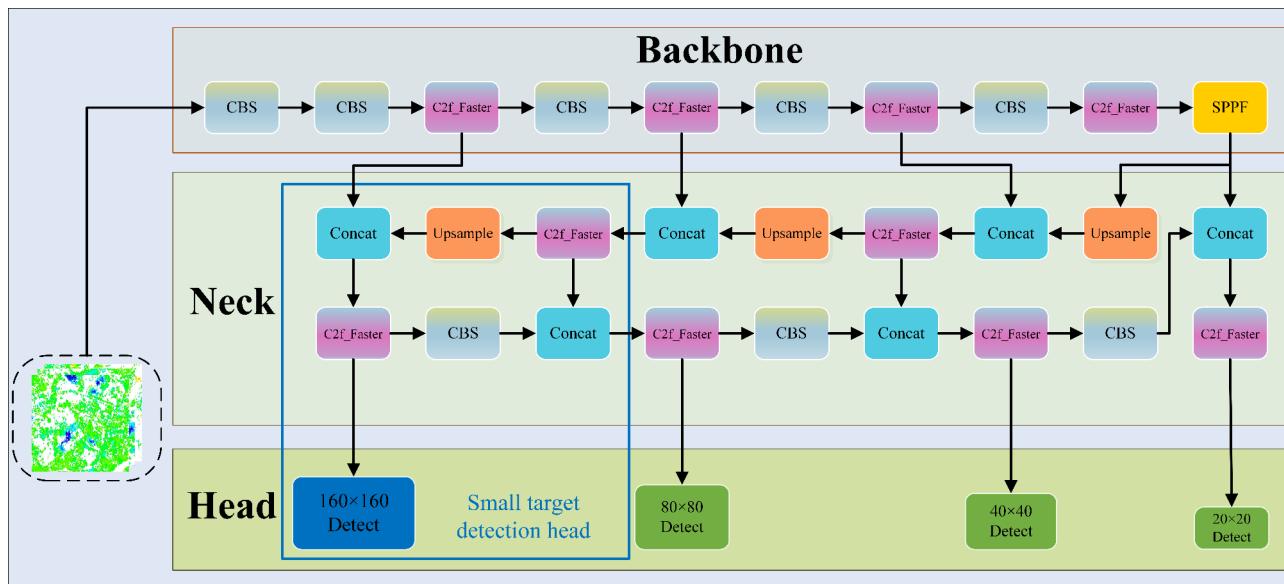


Fig. 3. InSAR-YOLOv8 model network structure.

To address this issue, we propose adding an additional small detection head tailored to small objects for further enhancing the detection accuracy. As shown in the content of the red box in Fig. 3, we introduced a 160×160 scale feature map detection head specifically designed to extract relevant features from smaller objects, thereby increasing the model's resolution and enhancing its detection capabilities. This adjustment enhances the detection capabilities of the network across a broader spectrum of applications and effectively improves recognition accuracy. When doing InSAR landslide detection, incorporating four detection heads allows the model to accurately recognize landslides of different sizes. This reduces the chances of missing any landslides and significantly enhances its detection capability.

C2f_Faster module design

Deep learning imposes specific requirements on various hardware devices, such as CPUs and GPUs. As a result, a neural network with low computational complexity and high operational speed becomes particularly crucial. Chen et al.⁷⁰ proposed a novel, fast, and efficient detection network called FasterNet along with the Partial Convolution (PCnov) module. The PCnov achieved an efficient feature extraction by reducing the redundant computations and excessive memory access, resulting in lower floating-point operations (FLOPs) compared to the conventional Conv module. Figure 4 illustrates the working modes of standard convolution and partial convolution. In Fig. 4(a), when the feature input is of $h \times w \times c$ in size and the regular convolution is of $k \times k$ in size, the FLOPs formula for the standard Convolution becomes:

$$\text{FLOPs}_{\text{Cnov}} = h \times w \times k^2 \times c^2 \quad (3)$$

As depicted in Fig. 4(b), the PCnov performs convolutions only on the first continuous feature with a feature channel count of c_p , and its FLOP calculation formula is as shown in Eq. (4):

$$\text{FLOPs}_{\text{PCnov}} = h \times w \times k^2 \times c_p^2 \quad (4)$$

When c_p is a quarter of c , the FLOPs of PCnov are only one-sixteenth of the normal convolution, which efficiently performs spatial information feature extraction while reducing the number of parameters and memory accesses.

FasterNet is based on the PCnov theory, featuring a relatively simple architecture, a fast operational speed, and an excellent performance. The C2f module in YOLOv8 utilizes the Bottleneck module to improve performance, but this also results in a substantial increase in the number of parameters and computations. This paper proposes the integration of the FasterNet Block module from the FasterNet network with the C2f module of the YOLOv8. The FasterNet Block module's structure is depicted in Fig. 5, characterized by its straightforward architecture and fewer parameters. It comprises a PCnov convolution and two 1×1 standard convolution modules.

Thus, this study integrates the C2f module and the FasterNet Block module to construct the C2f_Faster module. Compared to the original C2f module, this new module enhances the network's learning capability, reduces computational redundancy and memory access, and achieves higher computational efficiency. The network using the C2f_Faster module significantly reduces model size, parameters, and computational requirements, enabling the InSAR-YOLOv8 model to improve inference speed while achieving better detection accuracy. This helps lower the hardware requirements of deep learning models, allowing deployment on more general-purpose devices. In InSAR-based landslide detection tasks, the improved module can extract landslide deformation features with higher efficiency, and its lightweight nature further enhances the model's scalability.

SIoU loss function

The YOLOv8 bounding box regression loss takes the form of the CIoU Loss combined with the DFL Loss. The CIoU primarily considers three geometric parameters: overlap area, center point distance, and aspect ratio.

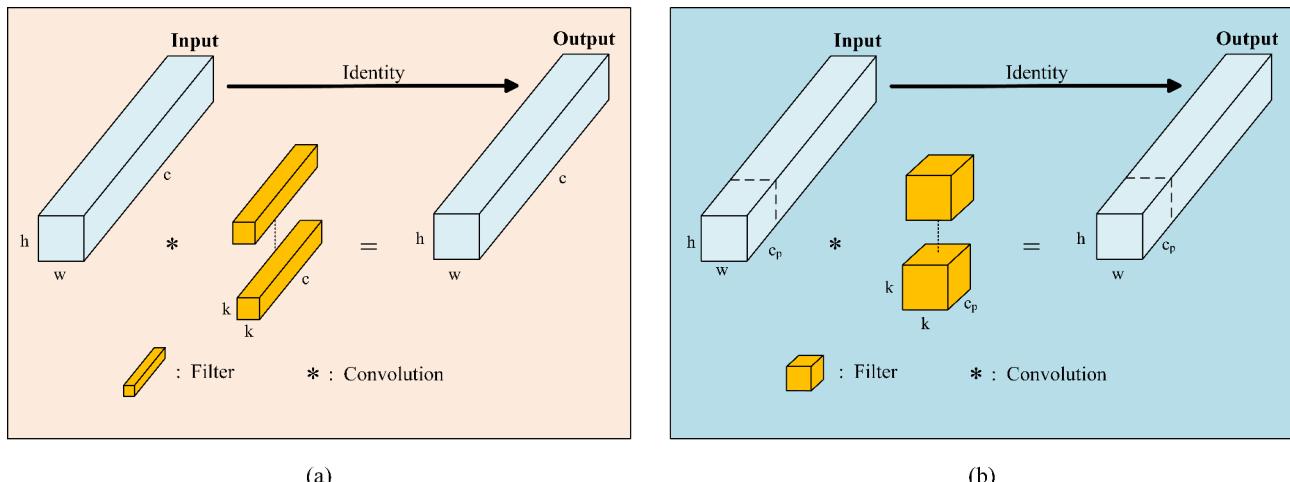
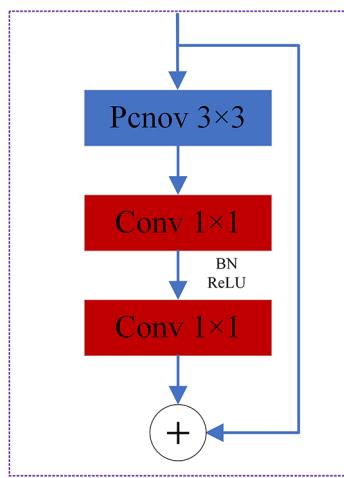


Fig. 4. Standard Convolution (a) and Partial Convolution (b).

**Fig. 5.** FasterNet Block Module.

Nevertheless, the description of its aspect ratio is subjective, resulting in a certain level of inherent uncertainty and disregarding the discrepancy in orientation between the actual box and the predicted box. Gevorgyan⁷¹ proposed a novel loss function, SIoU, which introduces the vector angle between the real box and the predicted box. This redefined penalty metric enhances the training and convergence speed. Now, assuming w and w^{gt} represent the width of the predicted box and the real box, respectively, with h and h^{gt} representing their heights, $(b_{cx}^{gt}, b_{cy}^{gt})$ denotes the center coordinates of the real box, while (b_{cx}, b_{cy}) represents those of the predicted box, while b and b^{gt} denote the boundary center point of the predicted and real boxes. Furthermore, w_c and h_c stands for the width and height of the minimum enclosing rectangle of the real box and the predicted box. The signs σ and γ represent the distance and height difference between the center points of the real box and the predicted box. The IoU signifies the intersection over the union ratio.

The computational formula for the CIoU can be expressed as Eq. (5):

$$\text{Loss}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\sigma^2 (b + b^{gt})}{(w_c)^2 + (h_c)^2} + \alpha v \quad (5)$$

Where:

$$\alpha = \frac{v}{(1 - \text{IoU}) + v}, v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \text{IoU} = \frac{\text{Intersection Area}}{\text{Union Area}} \quad (6)$$

The SIoU has redefined the related loss functions by introducing the concept of angle cost. It divides the computation into four distinct components: Angle cost, Distance cost, Shape cost, and IoU cost. The formula for calculating the Angle cost is presented in Eq. (7):

$$\theta = \cos \left(2 \times \left(\arcsin \left(\frac{\gamma}{\sigma} \right) - \frac{\pi}{4} \right) \right) \quad (7)$$

The distance cost calculation is shown in Eq. (8):

$$\Delta = \sum_{t=x,y} \left(1 - e^{-(2-\theta)pt} \right) \quad (8)$$

Among them:

$$p_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{w_c} \right)^2, p_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{h_c} \right)^2 \quad (9)$$

The shape cost calculation is shown in Eq. (10):

$$\Omega = \left(1 - e^{-\omega_w} \right)^\theta + \left(1 - e^{-\omega_h} \right)^\theta \quad (10)$$

Among them:

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (11)$$

The final SIoU calculation is presented in Eq. (12):

$$\text{Loss}_{\text{SIoU}} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \quad (12)$$

Overall, the SIoU function builds upon the excellent properties of CIoU by further enhancing the regression capability of bounding boxes, enabling the model to achieve more precise localization. In InSAR measurements, landslide areas often exhibit various shapes and scales, with some areas located in close proximity to each other, posing significant challenges for traditional loss functions. SIoU introduces angular and shape penalties to optimize the alignment between predicted and ground truth bounding boxes, enhancing the model's alignment capability. Therefore, this study replaces the CIoU function with SIoU to enable more precise localization when handling targets of different scales and shapes, effectively reducing errors between predicted and ground truth boxes. For landslide-prone areas with irregular shapes and uneven distributions, SIoU better adapts to these complex target characteristics, improving detection accuracy and model performance. SIoU demonstrates superior performance in InSAR-based landslide detection tasks, providing effective support for enhancing the accuracy and stability of landslide target detection.

Results

Experimental design

Datasets

The dataset utilized in the study is the InSAR landslide dataset employed in reference⁷², with example images depicted in Fig. 6. The dataset is based on SBAS-InSAR technology, collecting and processing 473 ALOS-2 PALSAR-2 images and 1168 Sentinel-1 images acquired in Guizhou Province, China, from 2017 to 2021. The processing methods include joint registration, combination of interferometric pairs, differential interferometry, and spatiotemporal joint filtering. Then, through further data processing and analysis, the deformation velocity and time series for the region were ultimately obtained. Following the processing, the results underwent color mapping for data visualization, facilitating a more intuitive understanding of the data. The InSAR measurement results were cross-referenced with the optical imagery from Google Earth and high-resolution remote sensing imagery obtained from the Unmanned Aerial Vehicles (UAVs), resulting in the manual identification and labeling of 363 active landslides displaying displacement signals. Ultimately, in order to fulfill the need for a substantial quantity of samples for training a deep learning model, the dataset underwent data augmentation procedures. The data augmentation methods included rotating, flipping, and scale transition of displacement

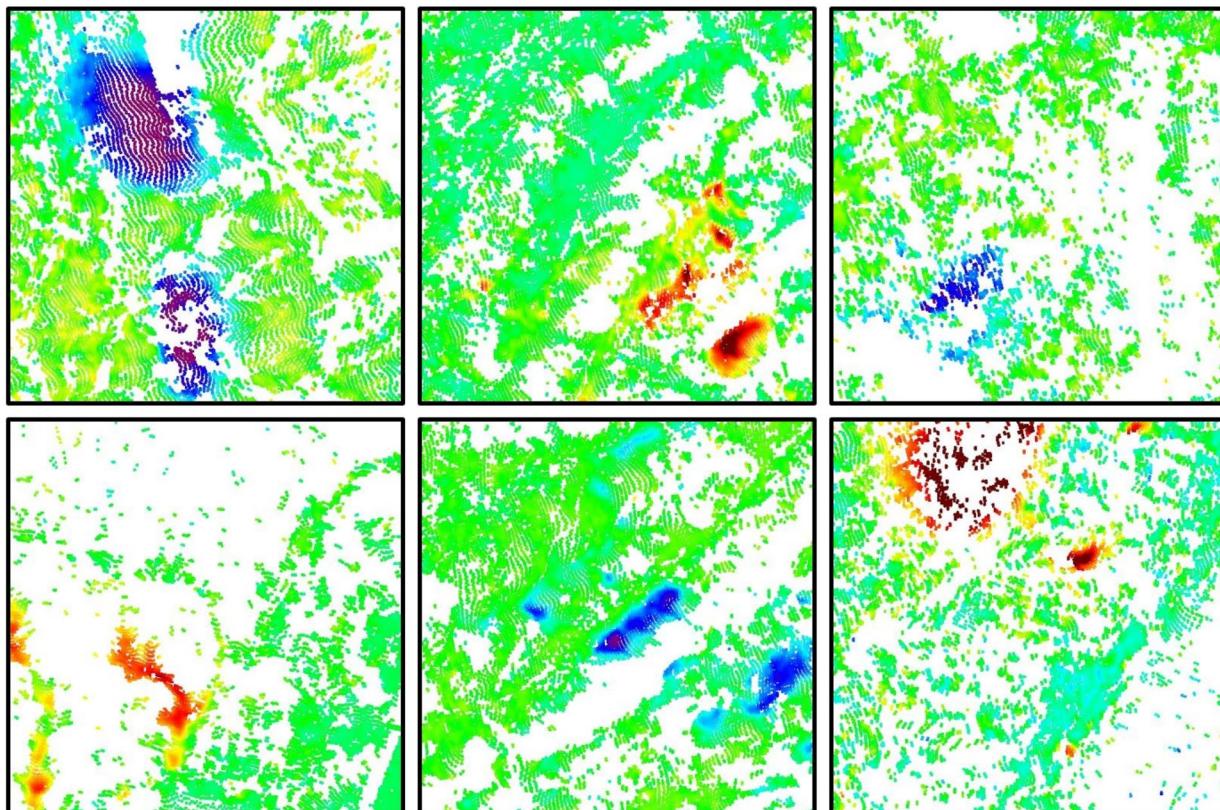


Fig. 6. Example images from the dataset.

rates. The final dataset contained 4984 sample images, each 512×512 pixels in size. In this experiment, the dataset is divided into a training set and a validation set, with the training set containing 3980 images and the validation set containing 1004 images, maintaining an approximate ratio of 8:2.

Experimental environment

The hardware setup for this study comprises an Intel(R) Core(TM) i9-9900 K CPU @ 3.60 GHz, 64GB of RAM, and an NVIDIA GeForce RTX 2080 Ti graphics card with 12GB of video memory. The software environment includes Windows 10 as the operating system, Python as the development language, and the PyTorch open-source deep learning framework as the network. The specific information is shown in Table 1.

Design of training strategies

The hyperparameters chosen for training are as follows: an input image size of 512×512 pixels, SGD as the optimizer, a batch size of 32, a model training period (in epochs) of 200, an initial learning rate ($Lr0$) of 0.01, a final learning rate (Lrf) of 0.001, and a momentum of 0.937. All hyperparameter settings are exhaustively tested to ensure that the best overall combination is used. Additionally, Mosaic augmentation is employed to prevent overfitting.

Evaluation indicators

This study follows the evaluation approach for target detection algorithms and emphasizes using the mean average precision (mAP) as the main metric for comprehensively evaluating the performance of target detection models. Specifically, it focuses on two IoU thresholds: 0.5 and $0.5 \sim 0.95$, referred to as mAP50 and mAP50:95, as the key indicators for evaluating model accuracy. The mAP metric offers a comprehensive evaluation by considering both the model's detection Precision and the Recall rate. The formula is presented in Eq. (13):

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (13)$$

Where AP (average precision) is an important index in the target detection task and is calculated as:

$$AP = \int_0^1 \text{Precision}(\text{Recall}) d(\text{Recall}) \quad (14)$$

Where Recall and Precision represent the ratio of true positive samples correctly predicted by the model to the actual positive samples present and the ratio of true positive samples correctly predicted by the model to all samples predicted as positive, respectively. Their formulas are shown in Eq. (15):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

Here, TP (True Positives) indicates the instances where the model correctly identifies the positive category of a target, while FP (False Positives) corresponds to the cases where the model incorrectly predicts the negative category of a target as positive. Likewise, FN (False Negatives) refers to the cases where the model inaccurately predicts the positive category of a target as negative.

The F1 score is derived from two metrics, namely Recall and Precision, which collectively reflect the model's correctness and accuracy. The formula is expressed in Eq. (16):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

Analysis of the experimental results

To validate the performance of the model proposed in this paper, a comparative analysis is conducted by comparing InSAR-YOLOv8 separately with YOLOv8n and other advanced models (e.g., DH-Faster-RCNN, RTMDet, and YOLOvX).

Comparison of experimental results with the YOLOv8n model

For the same dataset, 200 rounds of training are performed using both YOLOv8n model and the InSAR-YOLOv8 model without loading pre-training weights. The comparison of mAP50 and mAP50:95 curves is illustrated in Fig. 7. It is evident that, as the number of iterations increases, the accuracy of the InSAR-YOLOv8 model gradually surpasses that of the YOLOv8n model and becomes more stable.

Table 2 Presents a quantitative comparison between performance metrics before and after improvement. According to our results, the proposed InSAR-YOLOv8 model outperforms the original model across all performance metrics. Specifically, compared to the original YOLOv8n model, the InSAR-YOLOv8 model

System environment	Python version	CUDA version	Deep learning frameworks	Base network
Windows 10	3.8	11.8	pytorch:2.0.0	YOLOv8n

Table 1. Operation setup.

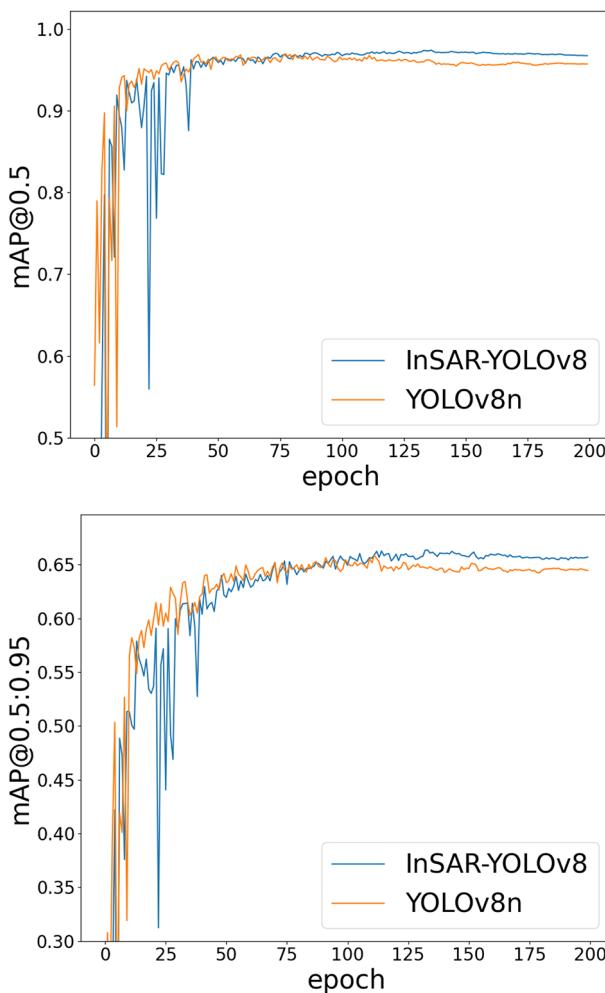


Fig. 7. Comparison of Accuracy Between InSAR-YOLOv8 Model and YOLOv8n Model.

Model	Model size	mAP50	mAP50:95	precision	recall	F1	Parameters
YOLOv8n	5.93MB	96.76%	65.95%	93.02%	90.78%	91.88%	3.01M
InSAR-YOLOv8	4.63MB	97.41%	66.47%	93.26%	90.88%	92.06%	2.26M

Table 2. Comparison of Performance Metrics between YOLOv8 and InSAR-YOLOv8

increases mAP50 from 96.76% to 97.41%, representing a 0.65% improvement. The mAP50:95 also shows an improvement from 65.95% to 66.47%, with a 0.52% improvement. The F1 score rose from 91.88% to 92.06%, marking a 0.18% increase. Additionally, the InSAR-YOLOv8 model reduced the parameter count from 3M to 2.26M, a reduction of approximately 25%. The model size decreased from 5.93MB to 4.63MB by 22%. This demonstrates that the InSAR-YOLOv8 model achieves a better detection accuracy while becoming more lightweight. This enhancement not only improves the model's performance but also optimizes its applicability, making it more suitable for InSAR landslide detection tasks.

To obtain more robust conclusions, we incorporated different random seeds in multiple experiments and calculated the mean and confidence intervals for various performance metrics. The experimental results are presented in Table 3.

Based on Table 3, the InSAR-YOLOv8 model demonstrates clear advantages over the YOLOv8n model in terms of average results across multiple experiments. The mAP50 increased from 96.66% to 97.29%, reflecting an improvement of 0.63%. Similarly, the mAP50:95 improved from 65.66% to 66.55%, showing an increase of 0.89%, and the F1 score increased from 91.35% to 91.80%, achieving a gain of 0.45%. Moreover, in terms of the confidence intervals (CI), the mAP50 CI of InSAR-YOLOv8 is [97.20%, 97.38%], while that of YOLOv8n is [96.49%, 96.83%], indicating that the improved model offers greater stability and less performance fluctuation. Both the mAP50:95 CI and F1 CI also demonstrate better performance, further validating the robustness and consistency of the model across multiple experiments. A comprehensive analysis of the experimental results

Model	mAP50 Mean	mAP50: 95 Mean	F1 Mean	mAP50 CI	mAP50: 95 CI	F1 CI
YOLOv8	96.66%	65.66%	91.35%	[96.49%, 96.83%]	[65.48%, 65.84%]	[91.17%, 91.53%]
InSAR-YOLOv8	97.29%	66.55%	91.80%	[97.20%, 97.38%]	[66.41%, 66.69%]	[91.56%, 92.06%]

Table 3. Comparison of Accuracy Stability between YOLOv8 and InSAR-YOLOv8

reveals that the InSAR-YOLOv8 model not only improves detection accuracy and stability but also maintains strong performance across various random seeds. InSAR-YOLOv8 has demonstrated promising performance in landslide detection tasks, suggesting enhanced reliability and practicality for real-world applications.

Figure 8 clearly presents a comparison between InSAR-YOLOv8 and YOLOv8 in the task of InSAR landslide detection. Panels (1), (2), (3), and (4) display detection results for four different sets of images. Observing the results in panels (1) and (4), it is evident that the InSAR-YOLOv8 model exhibits higher confidence levels, indicating that the model is more confident and accurate in identifying landslide targets, with superior data processing and feature extraction performance. In panel (2), the InSAR-YOLOv8 model, benefiting from the enhanced small target detection head, effectively identifies small landslide areas, whereas the YOLOv8 model fails to do so, resulting in missed detections. In panel (3), the YOLOv8n model exhibits misjudgment issues, while the InSAR-YOLOv8 model shows a greater stability, yielding detection results that closely match the labels. Overall, the InSAR-YOLOv8 model not only achieves a higher detection accuracy but also exhibits a more stable recognition performance, significantly reducing false positives and missed detections. This indicates that the model maintains a high performance across different scenarios, particularly excelling in detecting smaller landslides. Therefore, the model is more reliable and effective in practical applications.

Comparison of detection accuracy across different landslide scales

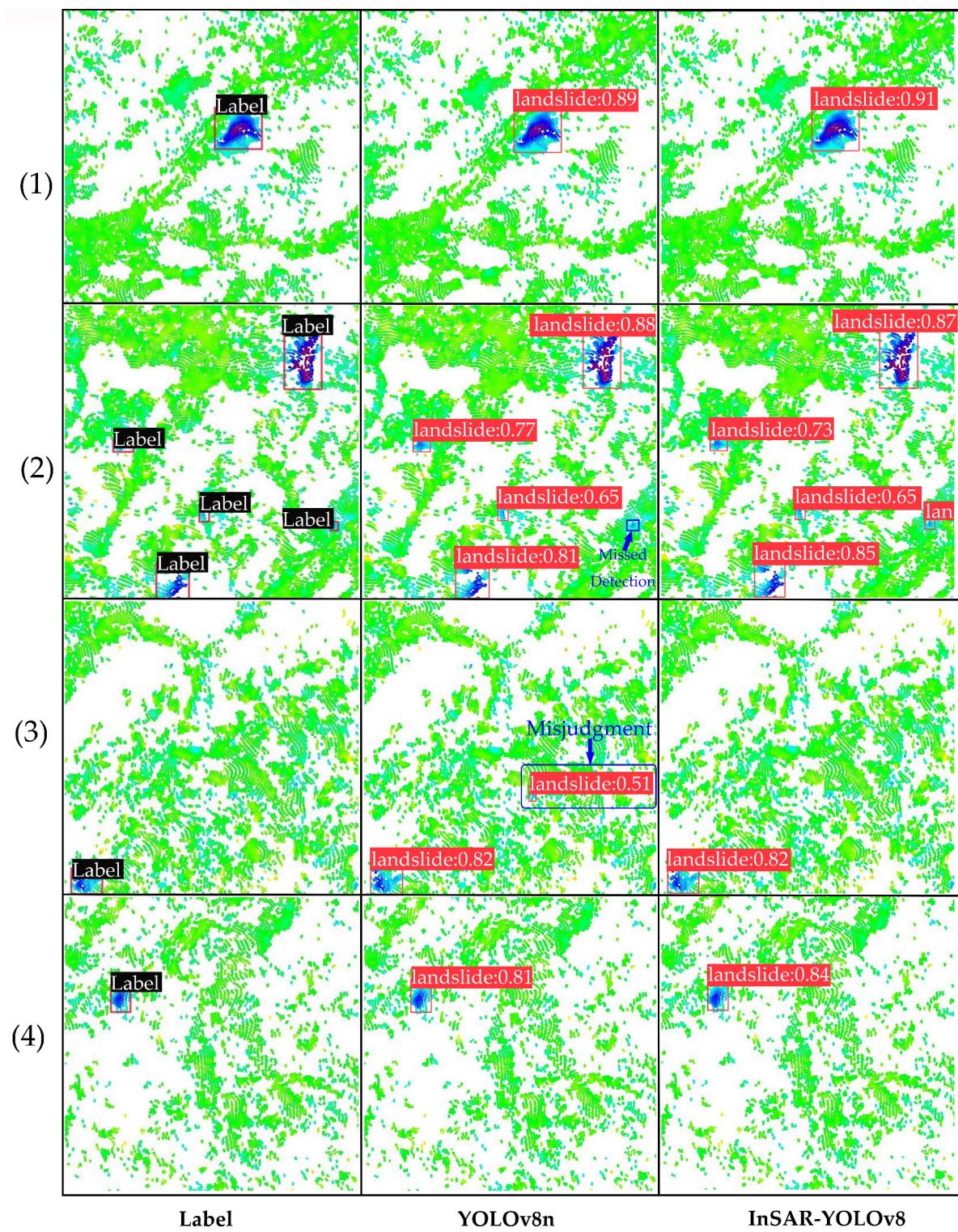
Based on the visualization results and detection accuracy, the InSAR-YOLOv8 model has demonstrated superior performance. To further validate the performance of InSAR-YOLOv8 in detecting landslides of different scales, this study defines small, medium, and large landslides following the commonly used COCO metrics in object detection algorithms. Specifically, landslides with an area smaller than 32 square pixels are classified as small landslides, those with an area between 32 and 96 square pixels are considered medium landslides, and those larger than 96 square pixels are classified as large landslides. In 512×512 resolution images, landslides covering less than 0.39% of the total image area are categorized as small landslides, those between 0.39% and 3.52% as medium landslides, and those exceeding 3.52% as large landslides. The detection accuracy results for small, medium, and large landslides are shown in Table 4.

Based on Table 4, it can be observed that InSAR-YOLOv8 outperforms YOLOv8 in terms of precision across different scales of landslide detection under the IoU threshold of 0.50:0.95. Specifically, the precision for InSAR-YOLOv8 in detecting small, medium, and large landslides are 57.2%, 73.1%, and 64.1%, respectively, all higher than those of YOLOv8. This indicates that InSAR-YOLOv8 has stronger multi-scale adaptability, especially in landslide detection tasks, demonstrating higher detection accuracy and more stable performance. It implies that the model can more accurately capture features of landslides of various sizes while effectively reducing false positives. In terms of recall, InSAR-YOLOv8 achieves 64.8%, 78.2%, and 73.3% for small, medium, and large landslides, respectively. Although its recall rate for large landslides is slightly lower than that of YOLOv8, InSAR-YOLOv8 still exhibits superior recall capability across different scales of landslides, demonstrating its advantage in reducing missed detections and improving the comprehensiveness of detection.

It is worth noting that the F1 Score comprehensively reflects the model's actual detection capability and stability, making it a key metric for evaluating detection accuracy. The F1 Scores of InSAR-YOLOv8 for small, medium, and large landslides are 60.9%, 75.6%, and 69.1%, respectively, showing improvements of 0.7%, 1.0%, and 1.9% compared to YOLOv8. These improvements indicate that InSAR-YOLOv8 achieves superior detection results across different scales of landslides, further enhancing the comprehensiveness and stability of detection. It also reflects the model's improved robustness in capturing complex target features and adapting to diverse scenarios. Compared to YOLOv8, InSAR-YOLOv8 demonstrates improved detection of small landslides occupying less than 0.39% of the image area and achieves higher accuracy in detail extraction and localization for medium and large landslides. Overall, InSAR-YOLOv8 maintains a lightweight structure while exhibiting more stable performance across different scales of landslide detection. Its superior comprehensive performance offers more reliable support for landslide disaster monitoring and early warning.

To more intuitively demonstrate the improvements in feature extraction achieved by the InSAR-YOLOv8 model, we utilize the Grad-CAM-Plus-Plus method to visualize the heatmaps of detection layers from different models, as illustrated in Fig. 9.

Both image sets contain multiple targets of varying scales, accompanied by complex boundary information and significant noise interference. Specifically, panel (1) provides a further analysis of panel (2) in Fig. 8. Upon examination, it can be observed that while the YOLOv8 model is capable of capturing certain prominent target regions, its attention distribution is relatively scattered, with weak responses in some areas. Furthermore, the generated visual boundaries exhibit substantial deviations from the actual landslide boundaries. Some regions of deep focus in the model do not align with the actual deformation areas, and for extremely small landslides that are missed, no effective heatmap responses were generated. This indicates that the YOLOv8 model has limitations in terms of the precision of its feature extraction and the focus of its attention mechanism, making it difficult to fully and accurately capture the complete morphology and salient features of landslides. However, with the introduction of a small target detection head in the InSAR-YOLOv8 model, its feature extraction

**Fig. 8.** Comparison between Recognition Results of InSAR-YOLOv8 and YOLOv8n.

Model	Precision (Small)	Precision (Medium)	Precision (Large)	Recall (Small)	Recall (Medium)	Recall (Large)	F1 (Small)	F1 (Medium)	F1 (Large)
YOLOv8n	56.80%	72.10%	60.60%	63.90%	77.10%	74.20%	60.20%	74.60%	67.20%
InSAR-YOLOv8	57.20%	73.10%	65.10%	64.80%	78.20%	73.30%	60.90%	75.60%	69.10%

Table 4. Comparison of Detection Results for Different Landslide Scales

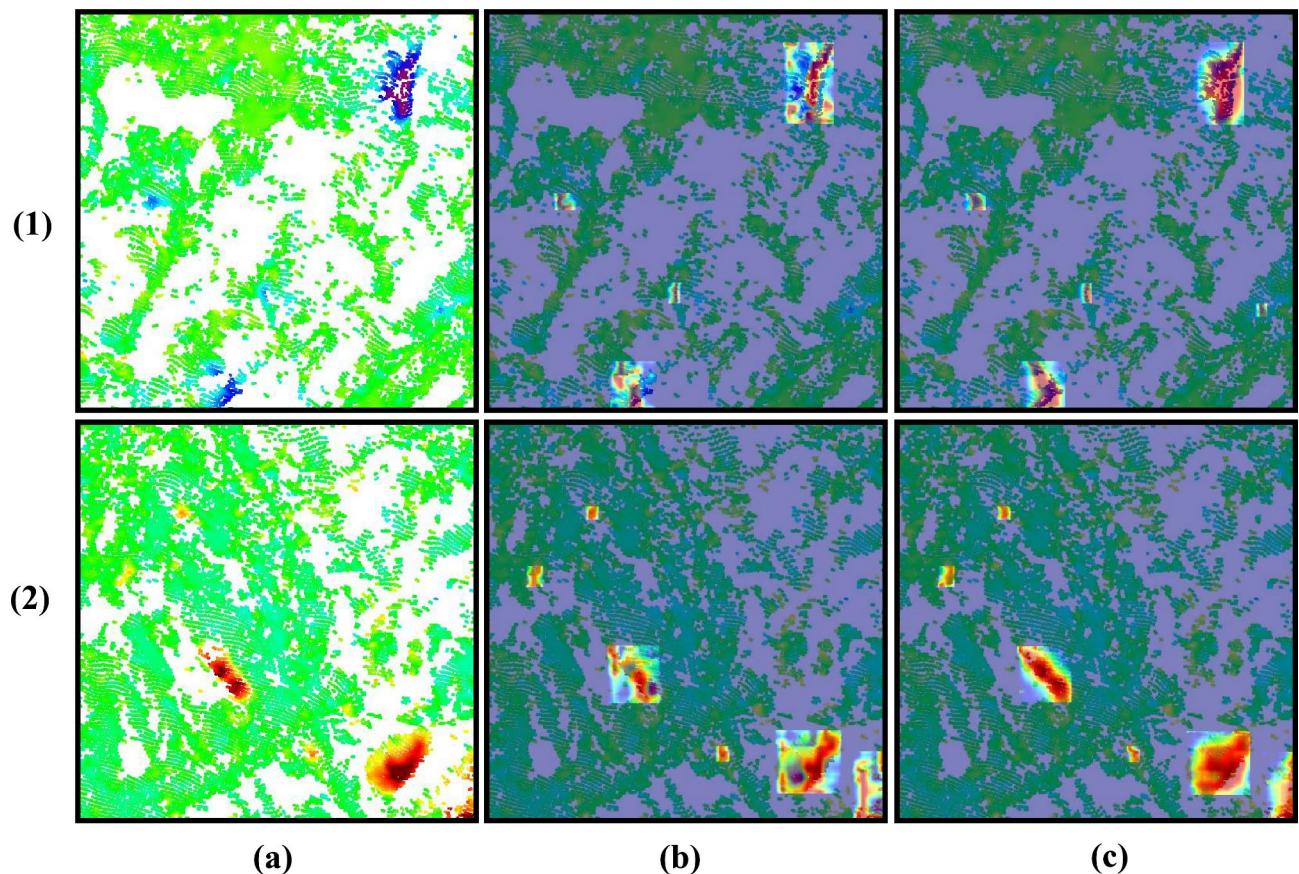


Fig. 9. Visualized heatmap of different networks, where (a) is the original image; (b) is the heatmap of the YOLOv8 model; (c) is the heatmap of the InSAR-YOLOv8 model.

capability has been significantly enhanced. The small target detection head enables the model to focus on finer-grained feature representations, resulting in clearer feature maps with more distinct and easily detectable target boundaries and morphological information. Additionally, the generated heatmaps exhibit more concentrated and smooth characteristics, with visual contours that align more closely with the true landslide boundaries. This indicates that InSAR-YOLOv8 demonstrates improved feature extraction capability and stability in multi-scale target scenarios, offering greater practical value.

Comparison of experimental results with other advanced models

To further validate the performance of the InSAR-YOLOv8 model, this study trained the Faster-R-CNN⁵⁰, RTMDet⁷³, Double-head-faster-RCNN⁷⁴, YOLOv3, Nas-fcos⁷⁵, and YOLOvX⁷⁶ models under the identical conditions. Among them, Faster-R-CNN and YOLOv3 represent classic two-stage and one-stage detection models, respectively, and have been widely applied across various detection tasks. Double-head-Faster-RCNN and YOLOvX are advanced iterations of these models, achieving significant improvements in both detection performance and efficiency. RTMDet and Nas-FCOS are outstanding models that have emerged in recent years in the field of object detection, characterized by excellent performance and strong representativeness. We employed mAP50, mAP75, mAP50:95 and model parameters as evaluation metrics. The precision of each model is shown in Table 5.

The experimental results clearly demonstrate that the InSAR-YOLOv8 model surpasses various benchmark models across multiple performance metrics. Specifically, compared to the Faster-R-CNN, RTMDet, Double-Head-Faster-RCNN, YOLOv3, Nas-fcos, and YOLOvX models, the InSAR-YOLOv8 model achieves higher mAP50 scores by 8.31%, 2.5%, 5.61%, 9.11%, 6.11%, and 3.71%, respectively. For mAP75, the InSAR-YOLOv8 model scores are higher by 23%, 8.06%, 19.76%, 26.66%, 22.56%, and 18.26%, respectively. In terms of mAP50:95, the InSAR-YOLOv8 model scores are higher by 15.27%, 5.97%, 12.37%, 16.77%, 14.57%, and 11.77%, respectively. These results demonstrate the superior detection accuracy and precision of the InSAR-YOLOv8 model, particularly at higher IoU thresholds. It validates that the InSAR-YOLOv8 model maintains a stable performance and high reliability across different application scenarios. It is worth noting that the InSAR-YOLOv8 model not only excels in detection accuracy but also has significantly fewer parameters compared to other models, thus avoiding limitations imposed by low hardware resources. The InSAR-YOLOv8 showcases extensive versatility and exceptional efficiency in real-world scenarios, offering reliable assistance for the automated identification and surveillance of landslide hazards.

Model	mAP50	mAP75	mAP50:95	Parameters
Faster-R-CNN	89.10%	54.60%	51.20%	99.65M
RTMDet	94.90%	70.00%	60.50%	8.89M
DH-faster-RCNN	91.80%	58.30%	54.10%	47.34M
Yolov3	88.30%	51.40%	49.70%	61.95M
Nas-fcos	91.30%	55.50%	51.90%	39.11M
YolovX	93.70%	59.80%	54.70%	5.06M
InSAR-YOLOv8	97.41%	78.06%	66.47%	2.26M

Table 5. Accuracy of each model.

Figure 10 illustrates the detection results of different models, where (a)-(h) represent the Label, Faster-R-CNN, RTMDet, DH-faster-RCNN, YOLOv3, Nas-fcos, YOLOvX, and InSAR-YOLOv8, respectively. Panels (1)-(3) show the detection results of the three sets of images using different models. Panels (1) and (2) contain multi-scale targets, with some targets being closely spaced. Other models tend to merge multiple landslides into a single target or detect the same target multiple times. In contrast, our model offers more accurate recognition results and effectively avoids the issues of false positives and duplicate detections seen in other models. The detection boxes are also closer to the actual boundaries.

Using the images from Group (2) as an example, we can see that there are four distinct landslide zones present in this area; two of these zones have quite isolated targets, while the other two are relatively close together. By analyzing the detection results of different models, it can be observed that the Faster-R-CNN and DH-faster-RCNN belong to the same object detection category with large number of parameters and a complex structure, giving them high precision. These models can effectively detect isolated landslide targets but perform poorly when detecting closely spaced targets, often merging two targets into one and resulting in a significant error. The RTMDet exhibits both erroneous detections and duplicate marking of the same target, which is definitely unacceptable. The YOLOv3 only detects two targets, exhibiting the worst performance. Such missed detections can lead to significant and unforeseeable losses in practice. The Nas-fcos model detected all four targets but with a low confidence level in the detection boxes. The YOLOvX, with the smallest number of parameters among the models, except for the InSAR-YOLOv8, showed a similar performance to the Faster-R-CNN, suffering from the same issues. The InSAR-YOLOv8 successfully detected all four targets with a high confidence level and effectively avoided the issues observed in other models. It demonstrated the best performance in InSAR landslide detection.

Panels (3) in Fig. 10 comprise images with targets that display vague characteristics and are of diminutive dimensions. When detecting landslides with indistinct features and small areas, all other models skipped some landslides. In contrast, the proposed improved model showed superior capability for detecting small targets and feature extraction, effectively mitigating this problem.

Overall, the InSAR-YOLOv8 model has achieved an exceptional detection accuracy while maintaining an extremely low number of parameters. The model detects all targets in the recognition images, reducing misjudgment and false detection, displaying higher performance. It is capable of meeting the demands of rapid identification in large-scale InSAR landslide detection tasks and showcasing robust practicality and reliability in real-world applications.

The effectiveness of different modules-ablation experiment

To assess the effectiveness of the various modules in our improved model (InSAR-YOLOv8), we selected the YOLOv8n configuration model as the baseline. We use the evaluation metrics such as mAP50, mAP50:95, F1, Parameters, and Model size. Ablation experiments are conducted by combining different modules, and the results are presented in Table 6.

From Table 6, it can be observed that the introduction of the small target detection head increased mAP50, mAP50:95, and F1 by 0.49%, 0.47%, and 0.15%, respectively. This is due to the introduction of the small target detection head, which enhances the model's ability to detect small-scale targets by strengthening feature extraction and multi-scale information fusion, effectively reducing missed detections of smaller objects. Additionally, it improves the transmission and integration of deep and shallow features, resulting in more stable performance across targets of various scales. This enhancement not only boosts the model's sensitivity and robustness but also improves overall detection performance, enabling the model to accurately capture a diverse range of targets in complex scenarios while reducing false positives and missed detections. However, this newly introduced structure increases the number of layers in the network architecture, resulting in a slight increase in model size. On the other hand, replacing the original C2f module with the C2f_Faster module not only reduced redundant information but also decreased the model size and the number of parameters by approximately 22% and 25%, respectively. This reduction results in a more compact and lightweight model, offering relative advantages for real-time applications and deployments on resource-constrained devices. The improvements in precision indicate that the C2f_Faster module effectively balances detection accuracy and computational efficiency, making it more scalable for large-scale datasets and practical applications. Additionally, the enhanced convolutional layer configuration in the C2f_Faster module reduces redundancies and promotes more efficient feature extraction, ensuring that the model maintains high detection accuracy across diverse scenarios. Although using the SIoU loss function⁷¹ alone led to a decrease in the model's performance metrics,

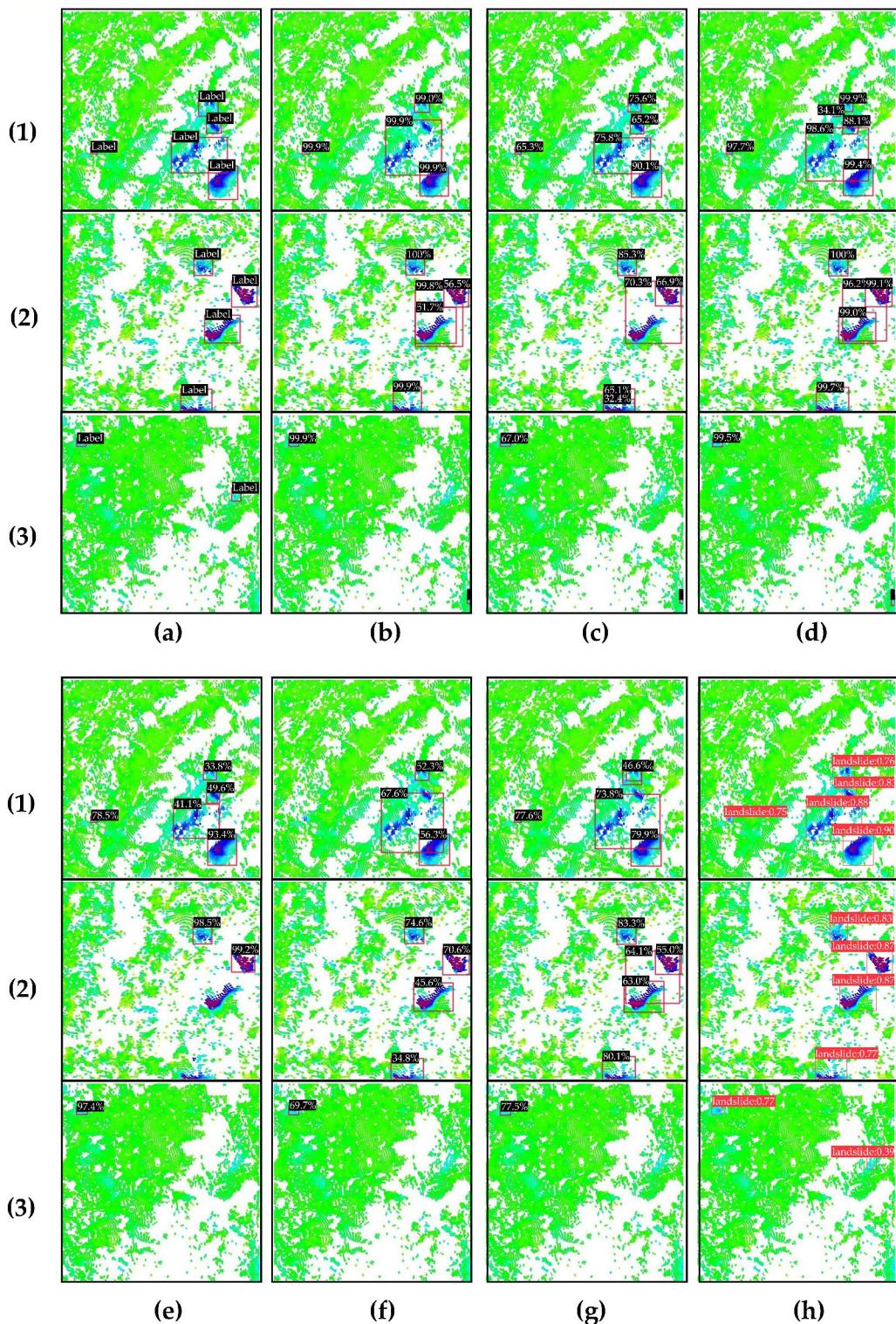
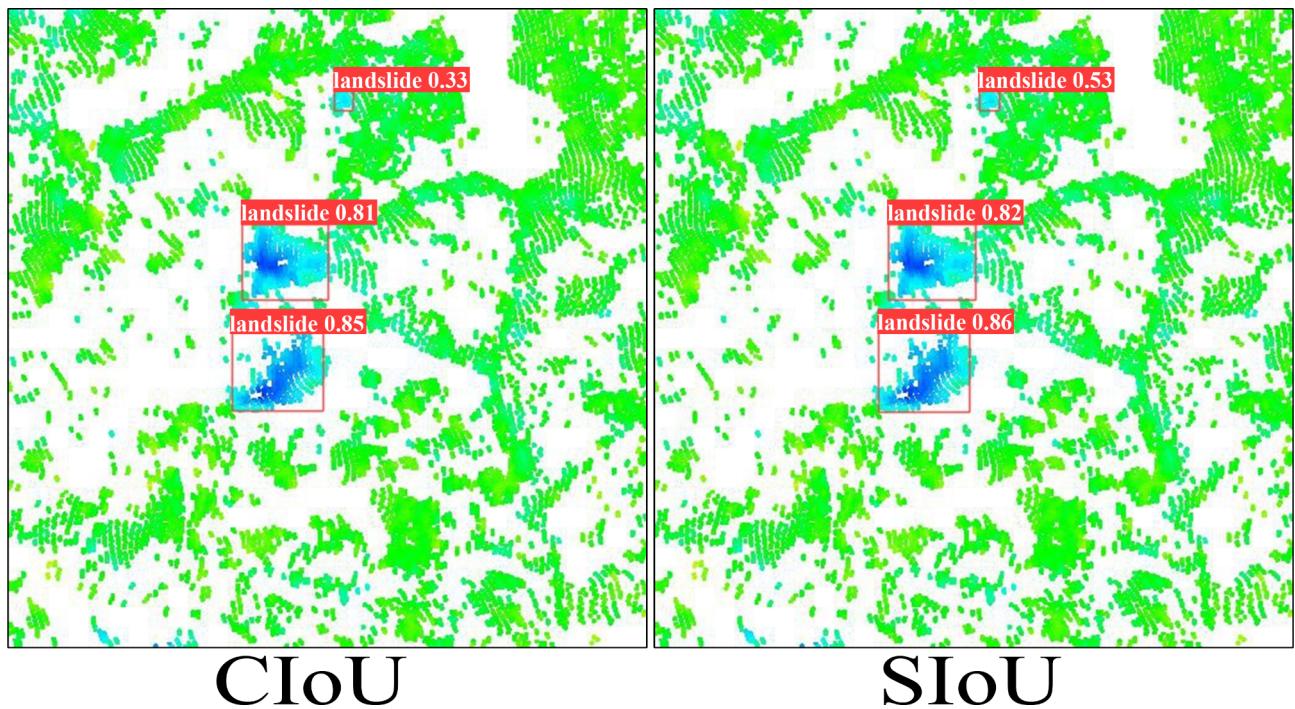


Fig. 10. (a) Label (b) Faster-R-CNN (c) RTMDet (d) DH-faster-RCNN (e) Yolov3 (f) Nas-fcos (g) YolovX (h) InSAR-YOLOv8.

however, combining SIoU with other modules effectively enhanced accuracy. For instance, applying the SIoU loss function in conjunction with the small target detection head further increased mAP_{50:95} and reached a peak of 66.78%. This improvement is attributed to the enhanced regression capability of the bounding boxes due to the introduction of SIoU which resulted in a more accurate target localization and prediction boxes that more closely align with actual conditions. Consequently, the model demonstrates a superior performance across different IoU thresholds and maintains efficient and accurate target detection capabilities under various complex

Method				mAP50	mAP50:95	F1	Model size	Parameters
Yolov8n	Small target detection head	C2f_Faster	SIoU					
✓				96.76%	65.95%	91.88%	5.93MB	3.01M
✓	✓			97.25%	66.42%	92.03%	6.02MB	2.98M
✓		✓		97.21%	66.43%	91.75%	4.58MB	2.30M
✓			✓	96.63%	65.60%	91.08%	5.93MB	3.01M
✓	✓		✓	97.05%	66.78%	91.72%	6.02MB	2.98M
✓		✓	✓	97.32%	66.31%	91.89%	4.58MB	2.30M
✓	✓	✓		97.16%	66.04%	91.59%	4.63MB	2.26M
✓	✓	✓	✓	97.41%	66.47%	92.06%	4.63MB	2.26M

Table 6. Results of the ablation experiments**Fig. 11.** Comparison between Recognition Results of CIoU and SIoU.

circumstance. Using the C2f_Faster module and SIoU loss function in combination further improved the mAP50 and F1 metrics and enhanced the localization capability of the bounding boxes under certain conditions, with all metrics surpassing the original model. When the small target detection head and C2f_Faster module are combined, changes in the convolutional layer structure and network hierarchy alter the model's parameters and computational complexity. There may be conflicts in feature extraction between the two which leads to a less-than-expected feature extraction performance. Despite a slight decrease in some metrics, the performance still exceeds that of the original model. However, when these modules are paired with the SIoU loss function, it enhances the ability to accurately locate targets of varying scales, resulting in overall performance improvement and resolution of potential problems in feature extraction. As illustrated in Fig. 11, it presents the detection results of the improved model using CIoU and the improved model employing SIoU for the same region.

The image contains three landslides of varying scales and shapes. For the two larger landslides, the confidence scores of the detection boxes using SIoU are slightly higher than those using CIoU. However, when detecting the smaller landslide, the confidence score with CIoU is only 0.33, while it increases to 0.53 with SIoU, reflecting an improvement of 0.2. With a confidence threshold of 0.5, the smaller landslide would be overlooked when using CIoU, resulting in a missed detection. Higher confidence scores across different targets indicate that the InSAR-YOLOv8 model utilizing SIoU achieves a closer fit to the ground truth bounding boxes, demonstrating greater reliability in handling multi-scale targets. Moreover, the SIoU loss function exhibits better compatibility with the small target detection head and the C2f_Faster module compared to CIoU, further enhancing the model's performance. The integration of these three modules enables them to synergistically collaborate, capitalizing on their own strengths to optimize the overall performance of the model. Ultimately, the improved InSAR-YOLOv8

increased mAP50, mAP50:95, and F1 by 0.65%, 0.52%, and 0.18%, respectively, and reduced the model size and parameters by 1.3MB and 0.75M, and achieved an optimal overall performance.

In summary, the proposed improvements significantly reduced the model's parameters while enhancing its detection accuracy. This allows the model to perform exceptionally well in target detection tasks across different scales and complex scenarios, providing robust technical support for landslide detection in large-scale InSAR images.

Discussion

Discuss the performance under different hyperparameters

In order to determine the most optimal hyperparameters, we conducted experiments with a range of different hyperparameters. We also evaluated the InSAR-YOLOv8 model with different batch sizes (2, 4, 8, 16, and 32), different optimizers (Adam, Adamax, AdamW, Nadam, and SGD)^{77–79}, and different initial learning rates (0.1, 0.05, 0.02, 0.01, and 0.005). The experimental results are presented in Table 7, with the hyperparameters that are ultimately chosen indicated by superscripts.

As shown in Table 7, when the batch size is set to 2, the smaller batch size offers finer-grained gradient estimates and achieves the best performance on specific metrics, such as mAP75. Nevertheless, the precision of other measurements is not as elevated as when the batch size is 32, and smaller batch sizes also result in longer training durations. The batch size of 32 provided the optimal overall performance. This indicates that larger batch sizes in deep learning training provide more stable gradient estimates and a higher computational efficiency, which help achieve better performance metrics. Therefore, we chose a batch size of 32 for our hyperparameter settings. Regarding the optimizer, the SGD optimizer performed the best, achieving the highest precision across all metrics. Compared to other adaptive optimization algorithms (such as Adam and AdamW), SGD demonstrates greater adaptability in handling noise and complex data distributions. Although SGD is a more classical optimization method, its strong generalization capability and stability in complex scenarios enhance the model's robustness and reliability. During training, SGD updates gradients uniformly, making it less prone to getting trapped in local optima, thus delivering superior performance. For the initial learning rate, a learning rate of 0.01 provided the best performance. The reasonable update step size ensures that the gradient updates are neither too large, which could result in skipping over the ideal solution, nor too small, which would slow down convergence. This allows the model to fully learn the data features during training and find the optimal solution stably and quickly. Overall, choosing an initial learning rate of 0.01 provides a superior performance.

Through these experiments, we validated the impact of batch size, optimizer, and initial learning rate on the model's performance. These experimental results confirm that carefully selecting and optimizing hyperparameters can enhance the performance of the InSAR-YOLOv8 model. This optimization approach improves the performance of the InSAR-YOLOv8 model, making it more reliable and applicable for landslide detection tasks based on InSAR measurements.

Discuss the performance under different parameter configurations

The improvement method proposed in this paper is conducted under the smallest parameter configuration version of the YOLOv8, namely YOLOv8n. To fully validate the advantages of the proposed improvement strategy, the same enhancement approach is applied to the YOLOv8s model, which has a larger set of parameters. Experiments under identical conditions are performed and the obtained results are shown in Table 8.

The data shown in Table 8 clearly demonstrates that implementing the improvement technique outlined in this study on configurations with different parameters still yields significant advantages. The YOLOv8s model resulted in an increase in mAP50 to 97.27%, mAP75 to 78.59%, mAP50:95 to 67.29%, and the F1 score to 92.5%. Compared to the original model, these represent increases of 0.47%, 0.61%, 0.88%, and 0.44% respectively,

Hyperparameters	mAP 50	mAP 75	mAP 50:95	F1
batch size	2	96.96%	78.51%	66.33%
	4	97.20%	77.46%	65.90%
	8	96.91%	76.96%	65.91%
	16	97.01%	77.97%	66.13%
	32*	97.41%	78.06%	66.47%
optimizer	Adam	97.01%	75.72%	64.94%
	Adamax	97.03%	77.53%	66.16%
	AdamW	97.08%	77.44%	65.83%
	Nadam	97.37%	76.50%	65.70%
	SGD*	97.41%	78.06%	66.47%
lr0	0.1	97.02%	76.63%	65.83%
	0.05	97.07%	76.98%	65.88%
	0.02	97.21%	78.08%	66.34%
	0.01*	97.41%	78.06%	66.47%
	0.005	97.14%	76.83%	65.95%

Table 7. Experimental Results under Different Hyperparameters

Model	mAP 50	mAP 75	mAP 50:95	F1	Model size	Parameters
YOLOv8n	96.76%	77.72%	65.95%	91.88%	5.93MB	3.01M
InSAR-YOLOv8n	97.41%	78.06%	66.47%	92.06%	4.63MB	2.26M
YOLOv8s	96.80%	77.98%	66.41%	92.06%	21.45MB	11.14M
InSAR-YOLOv8s	97.27%	78.59%	67.29%	92.50%	15.55MB	7.96M

Table 8. Experimental results of different configuration models

with a 27.5% reduction in the model volume and a 28.5% in the number of parameters. These experimental results further demonstrate the strong applicability of the improvement method proposed in this paper. When applied to the YOLOv8 models with larger parameter configurations, it can further enhance accuracy and reduce parameters, proving the advantages of our improvement strategy.

Limitations and future research directions

Overall, our experiments have achieved some promising results, but there are still potential limitations and areas for improvement.

The generalization ability of a model is an important indicator of its excellence. Training on diverse datasets ensures that the model maintains a stable performance in various practical applications. In this study, the model was primarily trained and evaluated on a dataset collected from Guizhou Province, China, which comprehensively reflects the region's terrain and climatic conditions. However, landslide occurrences are influenced by a complex interplay of factors, such as geological structures, topographic features, climate and precipitation patterns, vegetation cover, and human activities. In different regions, these factors may interact differently, resulting in significant variations in landslide characteristics. Although the model has shown promising detection results in this region, its ability to generalize to areas with different geological and environmental characteristics remains to be validated. To address this limitation and enhance the model's adaptability, it is necessary to further evaluate its performance on datasets collected from other geographical and environmental conditions. Such cross-regional validation will help ensure the model's robustness across diverse terrain types, vegetation coverage, and climatic scenarios, thereby strengthening its capability to detect landslides in a variety of environmental contexts. Due to the relative scarcity of publicly available datasets in this field, we plan to expand the dataset coverage in the future to include a broader range of landslide events, geographical regions, and environmental conditions. Additionally, we plan to utilize data augmentation techniques such as image blending and data synthesis to simulate more diverse topographical and environmental conditions. These techniques will help the model learn more generalized features, thereby enhancing its robustness in unknown scenarios. At the same time, transfer learning will be leveraged to help the model adapt to variations in landslide characteristics under different geological conditions, thereby improving its generalization ability across regions and environmental contexts. This will help ensure that the InSAR-YOLOv8 model maintains high accuracy and reliability across various landslide scenarios, thus validating its effectiveness and robustness for large-scale landslide detection using InSAR measurements.

We recognize that our model may encounter challenges in landslide detection tasks, such as noise interference and the spatial scale of landslides, which can impact detection accuracy and effectiveness. Specifically, during the detection process, the model may be influenced by noise, misidentifying it as a landslide, leading to false positives. Furthermore, for very small landslides, the resolution of the data and the presence of noise may cause their features to be smoothed or confounded with background noise, resulting in missed detections. To address these issues, future research will focus on two main areas. On the one hand, we will optimize the model's architecture to enhance its detection capability, enabling it to more effectively overcome noise interference, accurately distinguish between noise and deformation features, and improve detection precision for small landslide targets. On the other hand, we will focus on improving the quality of InSAR measurements by employing more efficient noise reduction and filtering algorithms to minimize the impact of noise. Additionally, using higher-resolution SAR data will allow for clearer capture of subtle surface deformation features, thereby better preserving the characteristics of small landslides. These approaches are expected to facilitate the acquisition of more accurate landslide information, further enhancing the model's stability and accuracy, and providing more comprehensive and precise technical support for practical applications.

In practical applications, we first obtain InSAR measurement data for the target area and use the model to quickly predict the location and scale of potential landslides. Subsequently, optical remote sensing imagery is employed for auxiliary validation to confirm the specific details of the landslide event. This coordinated use of multi-sensor data facilitates the verification of actual landslide occurrences and effectively mitigates the risk of false alarms caused by non-landslide-induced surface deformations, thereby minimizing unnecessary risks. In future research, we will further optimize the model to reduce its computational resource requirements. A lower computational resource demand will help the model maintain stable performance and high accuracy when deployed on various devices. We will also optimize the model's feature extraction network to better capture the fine details of landslide deformation areas, thereby achieving higher detection accuracies in InSAR landslide detection. Our goal is to develop a more reliable and efficient landslide detection model for practical applications, providing strong technical support for landslide disaster warning and monitoring. It is worth noting that we believe this method is not only suitable for detecting landslide disasters but can also be widely applied in monitoring other types of geological hazards. For example, geological hazards such as earthquakes and volcanic activity exhibit different effects on the InSAR measurement maps. For earthquakes, the ground

deformation typically propagates outward from the epicenter, potentially causing extensive surface fractures or subsidence. The model needs to enhance its ability to detect large-scale targets, accurately delineate boundary regions, and learn to recognize displacement differences along fault lines. In contrast, volcanic deformation zones often exhibit elliptical or concentric patterns due to the pressure exerted by subsurface magma chambers. However, volcanic deformation can also display irregular or complex patterns, influenced by factors such as topography, fault lines, or uneven magma distribution. To address these challenges, the model needs to improve its ability to detect irregular shapes and complex targets, ensuring that it can identify diverse deformation modes that may emerge during different stages of volcanic activity. By instructing deep learning models to acquire knowledge of the InSAR imaging properties of diverse geological disasters, the models may identify and categorize distinct geological hazards, hence improving their value and precision in monitoring geological disasters. This study combines image detection technology with InSAR geological disaster detection, utilizing data from multiple sensors for a comprehensive analysis. In the future, the integration of multi-source data is expected to further enhance the accuracy and reliability of geohazard detection models. This approach will provide more comprehensive and precise information to support practical applications and further advance the intelligence of geohazard detection.

Conclusions

Long-term monitoring and rapid identification of landslides are crucial for addressing landslide disasters. In the task of landslide detection based on InSAR measurements, we encounter numerous challenges, including the identification of small targets, multi-scale processing, and limited hardware resources. Most of existing models struggle to balance detection accuracy with resource efficiency, failing to meet the demands for quick and precise landslide identification. To address these issues, we propose an improved InSAR landslide detection model called InSAR-YOLOv8 based on the YOLOv8n. In the proposed method, a small target detection head is initially added to bolster the model's capability for feature extraction from small-scale targets, thus enhancing the model's ability to recognize small landslides and effectively addressing the significant scale differences of landslides observed in InSAR measurements. Then, a more efficient and lighter C2f_Faster module replaces the original C2f module, improving the model's information processing, eliminating redundancies, decreasing model parameters, and increasing accuracy. By using the SIoU loss function, the regression ability of the bounding box is improved, resulting in more accurate target localization and ultimately enhancing the model's detection performance. Experimental results demonstrate that the InSAR-YOLOv8 model in this research can effectively identify landslides. The InSAR-YOLOv8 model obtains a mAP50 of 97.41%, a mAP50:95 of 66.47%, and an F1 score of 92.06% by reducing parameters by around 25%. Compared with other mainstream models (such as Faster-R-CNN, RTMDet, Double-head-faster-RCNN, YOLOv3, Nas-fcos, and YOLOvX), the InSAR-YOLOv8 model exhibits improvements in mAP50:95 by 15.27%, 5.97%, 12.37%, 16.77%, 14.57%, and 11.77%, respectively, reflecting better stability and accuracy.

In summary, we have applied the deep learning object detection algorithms to the field of wide-area InSAR landslide detection and proposed a model with higher accuracy and fewer parameters. This model can quickly identify landslide hazard areas from large-scale InSAR images and effectively overcome the inefficiencies of the traditional manual identification methods. When dealing with complex scenarios involving landslides of varying scales, this model effectively avoids false positives and missed detections and maintains a stable performance and a high recognition accuracy. In practical applications, the InSAR-YOLOv8 can quickly and accurately locate landslide hazards in large-scale InSAR measurements, providing a new technical means for landslide disaster early warning and monitoring. Additionally, this model enhances the application of deep learning in the field of InSAR disaster monitoring and provides new insights into monitoring other surface deformations and natural disaster warnings.

Data availability

The Landslide_InSAR_image_datasets are available online at https://github.com/whuinsar/Automatic_Landslides_identification_with_InSAR.git.

Received: 29 July 2024; Accepted: 25 December 2024

Published online: 10 January 2025

References

1. Liu, W., Yan, S. & He, S. Landslide damage incurred to buildings: a case study of Shenzhen landslide. *Eng. Geol.* **247**, 69–83. <https://doi.org/10.1016/j.enggeo.2018.10.025> (2018).
2. Wang, Y., Zhao, B. & Li, J. Mechanism of the catastrophic June 2017 landslide at Xinmo Village, Songping River, Sichuan Province, China. *Landslides* **15**, 333–345. <https://doi.org/10.1007/s10346-017-0927-3> (2018).
3. Shan, Y. et al. Landslide Hazard Assessment combined with InSAR Deformation: a Case Study in the Zagunao River Basin, Sichuan Province, Southwestern China. *Remote Sens.* **16**, 99. <https://doi.org/10.3390/rs16010099> (2024).
4. Scaioni, M., Longoni, L., Melillo, V. & Papini, M. Remote sensing for landslide investigations: an overview of recent achievements and perspectives. *Remote Sens.* **6**, 9600–9652. <https://doi.org/10.3390/rs6109600> (2014).
5. Liu, W. et al. Forecast volume of potential landslides in alpine-canyon terrain using time-series InSAR technology: a case study in the Bailong River basin, China. (2023). <https://doi.org/10.1007/s10346-023-02135-2>
6. Cotecchia, V., Grassi, D. & Merenda, L. Fragilità dell'area Urbana Occidentale Di Ancona dovuta a movimenti di massa profondi e superficiali ripetutisi nel 1982. *Geologia Appl. e Idrogeologia*. **30**, 633–657 (1995).
7. Fu, J. & Zhang, G. Application and research of GPS Digital Image Measurement Technology in Landslide Disaster Monitoring and Prediction. *IOP Conf. Ser. Earth Environ. Sci.* **558**, 042050. <https://doi.org/10.1088/1755-1315/558/4/042050> (2020).
8. Li, X., Zhou, L., Su & Wu, W. Research progress on the application of InSAR technology in landslide disaster. *J. Remote Sens.* **25**, 614–629 (2021).
9. Zhao, C. & Lu, Z. Remote sensing of Landslides—A review. *Remote Sens.* **10**, 279. <https://doi.org/10.3390/rs10020279> (2018).

10. Strozzi, T., Ambrosi, C. & Raetzo, H. Interpretation of aerial photographs and Satellite SAR Interferometry for the Inventory of landslides. *Remote Sens.* **5**, 2554–2570. <https://doi.org/10.3390/rs5052554> (2013).
11. Turner, D., Lucieer, A. & De Jong, S. M. Time Series Analysis of Landslide Dynamics using an unmanned aerial vehicle (UAV). *Remote Sens.* **7**, 1736–1757. <https://doi.org/10.3390/rs70201736> (2015).
12. Burrows, K., Walters, R. J., Milledge, D., Spaans, K. & Densmore, A. L. A New Method for large-scale landslide classification from Satellite Radar. *Remote Sens.* **11**, 237. <https://doi.org/10.3390/rs11030237> (2019).
13. Lu, H. et al. Early identification of landslide hazards in the Upper and Lower reaches of the Jinsha River Baige Landslide combining Optical Remote sensing with InSAR. *J. Wuhan Univ. (Information Sci. Edition)* **44**, 1342–1354. <https://doi.org/10.13203/j.whugis20190086> (2019).
14. Dong, J. et al. Detection and displacement characterization of landslides using multi-temporal satellite SAR interferometry: a case study of Danba County in the Dadu River Basin. *Eng. Geol.* **240**, 95–109. <https://doi.org/10.1016/j.engeo.2018.04.015> (2018).
15. Zhao, C., Liu, X., Zhang, Q., PENG, J. & XU, Q. Research on loess landslide identification, monitoring and failure mode with InSAR technique in Heifangtai. *Gansu Geomatics Inform. Sci. Wuhan Univ.* **44**, 996–1007. <https://doi.org/10.13203/j.whugis20190072> (2019).
16. Wei, Z. et al. The Identification and Influence Factor Analysis of landslides using SBAS-InSAR technique: a case study of Hongya Village, China. *Appl. Sci.* **14**, 8413. <https://doi.org/10.3390/app14188413> (2024).
17. Liu, G. Introduction to Advanced Synthetic aperture radar (SAR) Remote Sensing Technology: interferometric synthetic aperture radar (InSAR). *Sichuan Surveying Mapp.*, 92–95 (2004).
18. Wasowski, J. & Bovenga, F. Investigating landslides and unstable slopes with satellite Multi temporal interferometry: current issues and future perspectives. *Eng. Geol.* **174**, 103–138. <https://doi.org/10.1016/j.engeo.2014.03.003> (2014).
19. Zhang, Y. et al. Investigating slow-moving landslides in the Zhouqu region of China using InSAR time series. *Landslides* **15**, 1299–1315. <https://doi.org/10.1007/s10346-018-0954-8> (2018).
20. Li, Z. et al. Application of Satellite Radar Remote sensing in Landslide Disaster Detection and Monitoring: challenges and countermeasures. *Eomatics Inform. Sci. Wuhan Univ.* **44**, 967–979. <https://doi.org/10.13203/j.whugis20190098> (2019).
21. Bekaert, D. P., Handwerger, A. L., Agram, P. & Kirschbaum, D. B. InSAR-based detection method for mapping and monitoring slow-moving landslides in remote regions with steep and mountainous terrain: an application to Nepal. *Remote Sens. Environ.* **249**, 111983. <https://doi.org/10.1016/j.rse.2020.111983> (2020).
22. Bouali, E. H., Oomen, T. & Escobar-Wolf, R. Mapping of slow landslides on the Palos Verdes Peninsula using the California landslide inventory and persistent scatterer interferometry. *Landslides* **15**, 439–452. <https://doi.org/10.1007/s10346-017-0882-z> (2018).
23. Pedretti, L. et al. InterpolatiON of InSAR Time series for the dEtection of ground deforMatiOn eVEnts (ONtheMOVE): application to slow-moving landslides. *Landslides* **20**. <https://doi.org/10.1007/s10346-023-02073-z> (2023).
24. Bianchini, S., Cigna, F., Righini, G., Proietti, C. & Casagli, N. Landslide HotSpot Mapping by means of Persistent Scatterer Interferometry. *Environ. Earth Sci.* **67**, 1155–1172. <https://doi.org/10.1007/s12665-012-1559-5> (2012).
25. Cigna, F., Bianchini, S. & Casagli, N. How to assess landslide activity and intensity with persistent Scatterer Interferometry (PSI): the PSI-based matrix approach. *Landslides* **10**, 267–283. <https://doi.org/10.1007/s10346-012-0335-7> (2013).
26. Tao, C., Zhang, Y. & Ren, S. Research Progress on the application of InSAR Technology in Landslide Detection and Monitoring. *Geol. China*, 1–16 (2024).
27. Zhu, Y., Yao, X., Yao, L. & Yao, C. Detection and characterization of active landslides with multisource SAR data and remote sensing in western Guizhou. *China Nat. Hazards*, 1–22. <https://doi.org/10.1007/s11069-021-05087-9> (2022).
28. Xie, M., Huang, J., Wang, L., Huang, J. & Wang, Z. Early landslide detection based on D-InSAR technique at the Wudongde hydropower reservoir. *Environ. Earth Sci.* **75**, 717. <https://doi.org/10.1007/s12665-016-5446-3> (2016).
29. Hussain, M. A., Chen, Z., Wang, R. & Shoib, M. PS-InSAR-Based validated Landslide susceptibility mapping along Karakorum Highway, Pakistan. *Remote Sens.* **13**, 4129. <https://doi.org/10.3390/rs13204129> (2021).
30. Michoud, C. et al. Large slope deformations detection and monitoring along shores of the Potrerillos dam reservoir, Argentina, based on a small-baseline InSAR approach. *Landslides* **13**, 451–465. <https://doi.org/10.1007/s10346-015-0583-4> (2016).
31. Ferretti, A., Prati, C. M. & Rocca, F. Permanent scatterers in SAR interferometry. IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293) 3, 1528–1530 vol.1523, (1999). <https://doi.org/10.1109/36.898661>
32. Berardino, P., Fornaro, G., Lanari, R. & Sansosti, E. A New Algorithm for Surface Deformation Monitoring based on small baseline Differential SAR interferograms. *IEEE Trans. Geoscience Remote Sens.* **40**, 2375–2383. <https://doi.org/10.1109/TGRS.2002.803792> (2002).
33. Dun, J., Feng, W., Yi, X., Zhang, G. & Wu, M. Detection and mapping of active landslides before Impoundment in the Baihetan Reservoir Area (China) based on the Time-Series InSAR Method. *Remote Sens.* **13**, 3213. <https://doi.org/10.3390/rs13163213> (2021).
34. Li, Y., Feng, X., Li, Y., Jiang, W. & Yu, W. Detection and analysis of potential landslides based on SBAS-InSAR technology in alpine canyon region. *Environ. Sci. Pollut. Res.* **31**, 6492–6510. <https://doi.org/10.1007/s11356-023-31473-w> (2024).
35. Dong, J., Niu, R., Li, B., Xu, H. & Wang, S. Potential landslides identification based on temporal and spatial filtering of SBAS-InSAR results. *Geomatics Nat. Hazards Risk* **14**, 52–75. <https://doi.org/10.1080/19475705.2022.2154574> (2023).
36. Li, J., Xing, X. & Ou, J. Locating and characterizing potential rainfall-induced landslides on a regional scale based on SBAS-InSAR technique. *Bull. Eng. Geol. Environ.* **82**, 329. <https://doi.org/10.1007/s10064-023-03356-4> (2023).
37. Luo, S. et al. An Improved Method for Automatic Identification and Assessment of potential geohazards based on MT-InSAR measurements. *Remote Sens.* **13**, 3490. <https://doi.org/10.3390/rs13173490> (2021).
38. Rawat, W. & Wang, Z. Deep convolutional neural networks for image classification: a Comprehensive Review. *Neural Comput.* **29**, 2352–2449. https://doi.org/10.1162/neco_a_00990 (2017).
39. Anantrasirichai, N., Biggs, J., Albino, F., Hill, P. & Bull, D. Application of machine learning to classification of volcanic deformation in routinely generated InSAR Data. *J. Geophys. Research: Solid Earth* **123**, 6592–6606. <https://doi.org/10.1029/2018JB015911> (2018).
40. Sun, J. et al. Automatic detection of volcanic surface deformation using deep learning. *J. Geophys. Research: Solid Earth* **125**. <https://doi.org/10.1029/2020JB019840> (2020). e2020JB019840, doi.
41. , J., B. C. M., D. &, B. W. Identification of Surface Deformation in InSAR using machine learning. *Geochem. Geophys. Geosyst.* **22**. <https://doi.org/10.1029/2020GC009204> (2021).
42. Wu, Z., Wang, T., Wang, Y., Wang, R. & Ge, D. Deep learning for the detection and phase unwrapping of mining-induced deformation in large-scale interferograms. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–18. <https://doi.org/10.1109/TGRS.2021.3121907> (2021).
43. Wu, Z. et al. Automatic detection and classification of land subsidence in deltaic metropolitan areas using distributed scatterer InSAR and oriented R-CNN. *Remote Sens. Environ.* **290**, 113545. <https://doi.org/10.1016/j.rse.2023.113545> (2023).
44. Zhao, Y. et al. A new algorithm for intelligent detection of geohazards incorporating attention mechanism. *Int. J. Appl. Earth Obs. Geoinf.* **113**, 102988. <https://doi.org/10.1016/j.jag.2022.102988> (2022).
45. Lin, T., Fan, H., Sun, Y. & Li, X. & Zhongwenqian Zhuang, H. Research on Automatic Identification Method for wide-area InSAR Mining Subsidence zones based on U2-Net. *Metal Mine* 1–11 (2023).
46. McColl, S. T. & Cook, S. J. A universal size classification system for landslides. *Landslides* **21**, 111–120. <https://doi.org/10.1007/s10346-023-02131-6> (2023).

47. Li, K., Wan, G., Cheng, G., Meng, L. & Han, J. Object detection in optical remote sensing images: a survey and a new benchmark. *ISPRS J. Photogrammetry Remote Sens.* **159**, 296–307. <https://doi.org/10.1016/j.isprsjprs.2019.11.023> (2020).
48. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 142–158. <https://doi.org/10.1109/TPAMI.2015.2437384> (2015).
49. Girshick, R. Fast r-cnn. arXiv preprint arXiv:1504.08083 (2015).
50. Ren, S., He, K., Girshick, R., Sun, J. & Faster, R-C-N-N. Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031> (2017).
51. Li, Q., Xu, X., Guan, J. & Yang, H. The improvement of Faster-RCNN Crack Recognition Model and parameters based on attention mechanism. *Symmetry* **16**, 1027. <https://doi.org/10.3390/sym16081027> (2024).
52. Varalakshmi, P. Extraction of building footprint using MASK-RCNN for high resolution aerial imagery. *Environ. Res. Commun.* **6**, 075015. <https://doi.org/10.1088/2515-7620/ad5b3d> (2024).
53. Wang, J. et al. Loess landslides detection via a partially supervised learning and improved Mask-RCNN with multi-source remote sensing data. *Catena* **231**, 107371. <https://doi.org/10.1016/j.catena.2023.107371> (2023).
54. Redmon, J., Divvala, S. K., Girshick, R. B. & Farhadi, A. You only look once: Unified, Real-Time object detection. *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. 779–788 <https://doi.org/10.1109/CVPR.2016.91> (2016).
55. Bi, F. & Yang, J. in 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC). 10–14.
56. Redmon, J. & Farhadi, A. YOLOv3: an incremental improvement. *ArXiv Abs.* <https://doi.org/10.48550/arXiv.1804.02767> (2018). /1804.02767, doi.
57. Jocher, G. et al. ultralytics/yolov5: v6. 0-YOLOv5n'Nano'models, roboflow integration, TensorFlow export, OpenCV DNN support. *Zenodo* <https://doi.org/10.5281/ZENODO.5563715> (2021).
58. Terven, J. R., Esparza, D. M. C. & Romero-González, J. A. A. Comprehensive Review of YOLO Architectures in Computer Vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **5**, 1680–1716. <https://doi.org/10.3390/make5040083> (2023).
59. Wang, C. Y., Bochkovskiy, A. & Liao, H. Y. M. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 7464–7475.
60. Song, Y., Guo, J., Wu, G., Ma, F. & Li, F. Automatic recognition of landslides based on YOLOv7 and attention mechanism. *J. Mt. Sci.* **21** <https://doi.org/10.1007/s11629-024-8669-x> (2024).
61. Arai, K., Nakaoka, Y. & Okumura, H. Method for disaster area detection with Just one SAR Data Acquired on the Day after Earthquake based on YOLOv8. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **15**. <https://doi.org/10.14569/IJACSA.2024.0150344> (2024).
62. He, C. et al. Analysis of the impact of different improvement methods based on YOLOv8 for weed detection. *Agriculture* **14**, 674. <https://doi.org/10.3390/agriculture14050674> (2024).
63. Ding, B., Zhang, Y. & Ma, S. A. Lightweight real-time infrared object detection model based on YOLOv8 for Unmanned Aerial Vehicles. *Drones* **8**, 479. <https://doi.org/10.3390/drones8090479> (2024).
64. Rasheed, A. F. & Zarkoosh, M. Optimized YOLOv8 for multi-scale object detection. *J. Real-Time Image Proc.* <https://doi.org/10.1007/s11554-024-01582-x> (2024).
65. Al Mudawi, N. et al. Vehicle detection and classification via YOLOv8 and Deep Belief Network over aerial image sequences. *Sustainability* **15**, 14597. <https://doi.org/10.3390/su151914597> (2023).
66. Oh, G. & Lim, S. One-stage Brake Light Status Detection based on YOLOv8. *Sensors* **23**, 7436. <https://doi.org/10.3390/s23177436> (2023).
67. Sampurno, R. M., Liu, Z., Abeyrathna, R. M. R. D. & Ahamed, T. Intrarow Uncut Weed Detection using you-only-look-once Instance Segmentation for Orchard plantations. *Sensors* **24**, 893. <https://doi.org/10.3390/s24030893> (2024).
68. Zhang, H., Wang, Y., Dayoub, F. & Sunderhauf, N. VarifocalNet: an IoU-aware dense object detector. *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. 8510–8519 <https://doi.org/10.48550/arXiv.2008.13367> (2020).
69. Zheng, Z. et al. Enhancing geometric factors in Model Learning and inference for object detection and Instance Segmentation. *IEEE Trans. Cybernetics* **52**, 8574–8586. <https://doi.org/10.48550/arXiv.2005.03572> (2020).
70. Chen, J. et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 12021–12031, (2023). <https://doi.org/10.1109/CVPR52729.2023.01157>
71. Gevorgyan, Z. SIoU loss: more powerful learning for bounding box regression. *ArXiv* <https://doi.org/10.48550/arXiv.2205.12740> (2022).
72. Cai, J. et al. Automatic identification of active landslides over wide areas from time-series InSAR measurements using faster RCNN. *Int. J. Appl. Earth Obs. Geoinf.* **124**, 103516. <https://doi.org/10.1016/j.jag.2023.103516> (2023).
73. Lyu, C. et al. RTMDet: an empirical study of Designing Real-Time object detectors. *ArXiv Abs.* <https://doi.org/10.48550/arXiv.2212.07784> (2022). /2212.07784, doi.
74. Wu, Y. et al. Rethinking classification and localization for object detection. 10183–10192, (2019). <https://doi.org/10.48550/arXiv.1904.06493>
75. Wang, N. et al. NAS-FCOS: efficient search for object detection architectures. *Int. J. Comput. Vis.* **129**, 3299–3312. <https://doi.org/10.1007/s11263-021-01523-2> (2021).
76. Ge, Z. et al. Exceeding YOLO Series in 2021. (2021). <https://doi.org/10.48550/arXiv.2107.08430>
77. Kingma, D. P., Ba, J. & Adam A method for stochastic optimization. *CoRR. abs.* <https://doi.org/10.48550/arXiv.1412.6980> (2014). /1412.6980,
78. Ruder, S. An overview of gradient descent optimization algorithms. *ArXiv Abs.* <https://doi.org/10.48550/arXiv.1609.04747> (2016). /1609.04747, doi.
79. Loshchilov, I. & Hutter, F. in International Conference on Learning Representations.

Author contributions

R.M. designed the experiments and wrote the main manuscript. H.Y. revised and suggested the paper and helped with the formatting review and editing of the paper. R.M. and H.Y. developed the experimental plan and methodology; R.M. conducted the experiments; X.L. and X.Y. performed the experimental validation; X.L., X.Y., and T.G. analyzed the data and generated the graphs; P.L., T.G., and X.L. conducted the literature survey and collected the data; All authors reviewed and wrote the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Consent to participate

All authors are contributing and accepting to submit the current work.

Additional information

Correspondence and requests for materials should be addressed to H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025