A novel Dynahead-Yolo neural network for the detection of landslides with variable proportions using remote sensing images

Zheng Han[1,2], Zhenxiong Fang[1], Yange Li[1]* and Bangjie Fu[1]

[1]School of Civil Engineering, Central South University, Changsha, China, [2]Hunan Provincial Key Laboratory for Disaster Prevention and Mitigation of Rail Transit Engineering Structures, Changsha, China

Efficient and automatic landslide detection solutions are beneficial for regional hazard mitigation. At present, scholars have carried out landslide detection based on deep learning. However, continuous improvement regarding the accuracy of landslide detection with better feature extraction of landslides remain an essential issue, especially small-proportion landslides in the remote sensing images are difficult to identify up to date. To address this issue, we propose a detection model, the so-called Dynahead-Yolo which is designed by combining unifying scale-aware, space-aware, and task-aware attention mechanisms into the YOLOv3 framework. The proposed method focuses on the detailed features of landslide images with variable proportions, improving the ability to decode landslides in complex background environments. We determine the most efficient cascade order of the three modules and compare previous detection networks based on randomly generated prediction sets from the three study areas. Compared with the traditional YOLOv3, the detection rate of Dynahead-Yolo in small-proportion landslides and complex background landslides is increased by 13.67% and 14.12%, respectively.

1 Introduction

Landslide is a common and extremely hazardous natural phenomenon, that poses serious threats to human lives and cause huge economic losses (Clague et al., 2012; Alam, 2020; Valdés Carrera et al., 2021). After a landslide, it is essential to determine quickly and accurately the magnitude and distribution area of the landslides for subsequent rescue. It is also helpful to update the existing landslide database and provide data support for landslide research (Hungr et al., 2014; Ghorbanzadeh et al., 2019; Guzzetti et al., 2012).

Traditional in-site investigation is a common method of landslide investigation, requiring surveyors spend plenty of time and efforts during in-site surveys. With the

booming and recent development of remote sensing technology, the resolution of remote sensing images has significantly enhanced, which provides potential to obtain large-proportion feature information and has been widely used in geological hazard interpretation (Meena et al., 2022; Liu and Wu, 2016). Currently, there are four major types of methods for landslide identification using remote sensing images, i.e., visual interpretation, pixel-based, object-oriented, and artificial intelligence methods (Ju et al., 2020; Ju et al., 2022). Visual interpretation is a conventional method of using remote sensing images to obtain landslide information (Xu et al., 2014; Petschko et al., 2016; Fiorucci et al., 2019). It requires experts to comprehensively utilize image features such as object shape, texture, and spectrum, and then combine some non-remote sensing data for analysis and reasoning. This method consumes a lot of time and energy, and has limitations such as large errors and low efficiency (Hölbling et al., 2014; Moosavi et al., 2014; Wang et al., 2017; Zhao et al., 2017; Singh et al., 2021). Pixel-based methods usually uses binarization algorithm to determine whether a pixel of the image belong to the landslide (Li et al., 2014; Han et al., 2019). Recently development of this kind of methods refers to Han et al. (2022), in which we proposed a pixel-based landslide interpretation method and designed a multi-strategy feature fusion strategy, which combined the terrain slope of the detection object, the main axis features and the Normalized Difference Vegetation Index (NDVI) for screening, to reduce the false detection rate of landslides. However, it is difficult to distinguish them correctly when there are objects in the image with spectral characteristics similar to those of landslides. The object-oriented recognition method segments remote sensing images by setting certain thresholds based on spectrum, shape, and texture information (Sandric et al., 2010; Eeckhaut et al., 2011; Lu et al., 2011; Stumpf and Kerle. 2011). However, the method is less applicable since the pre-defined feature thresholds are varying case by case, and therefore need empirical adjusting.

The above remarkable methods attempt detect landslides in remote sensing images. These methods have promoted the progress of landslide detection research to varying degrees. However, inherent limitation should be noticed, such as a long detection time and low accuracy in the case of the complex background in remote sensing images. With the continuous development of artificial intelligence, big data and other technologies, recent attempts are trying to apply artificial intelligence (AI)-based methods for landslide detection. AI-based methods can be generally divided into machine learning and deep learning methods. Many studies have been performed on the development of machine learning landslide detection algorithms such as primary logistic regression, support vector machines, Bayesian methods, and decision trees (Parker, 2013; Korup and Stolle, 2014; Hu et al., 2019; Piralilou et al., 2019). This type of methods generally requires manual construction and selection of features, followed by classification with a classifier,

which complicates the algorithm and limits real-time applications. The main algorithms of deep learning currently include Convolutional Neural Networks (CNNs) (Chumerin, 2017), Recurrent Neural Networks (RNNs) (Zaremba et al., 2014), and Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). CNN-based methods have excellent non-linear mapping capabilities and can automatically learn features of landslide data (Hao et al., 2016; Vargas et al., 2017), which can quickly and accurately identify landslides. Ghorbanzadeh et al. (2022a) established a benchmark dataset by manually annotating landslide images and evaluated the performance of 11 deep learning models for landslide boundary detection. Furthermore, they applied the U-Net and ResU-Net models to landslide detection from free satellite data for the first time. They train the model using three case study regions and evaluate the transferability of the model through different training-test scenarios (Ghorbanzadeh et al., 2021). A recent development of the U-Net model for landslide detection could be referred to Fu et al. (2022). Cai et al. (2021) introduced dense convolutional networks into landslide detection, and significantly improved the detection ability of the model through measures such as feature reuse and feature enhancement. It is challenging to detect landslides in complex backgrounds based on CNN, but there are some excellent studies that have achieved such goals. Ju et al. (2020) combined deep learning and Google Earth data to detect historical landslides in typical loess regions in China. They established a database of historical loess landslides and used mask region-based convolutional networks to automatically identify loess landslides. Yu et al. (2022) realized the detection of loess landslides in complex backgrounds by improving the You Only Look Once X (YOLO X) model. Ghorbanzadeh et al. (2020) proposed the Dempster–Shafer model based on convolutional neural networks and combined it with the analysis of terrain factors to reduce the false detection rate of landslides with complex backgrounds. According to the different forms of landslide detection results, these CNN-based detection models are mainly divided into two categories. The first type is the segmentation model represented by U-Net and ResNet. These kinds of models classify the foreground and background based on pixels and predicts the boundaries of landslides (Jiang et al., 2021; Liu et al., 2022). The second type is the bounding box detection model represented by YOLO and Faster R-CNN. The method first divides the input image into patches of different sizes, and then classifies each sub-image to distinguish whether the image patch is a landslide (Hou et al., 2022; Liu et al., 2022). The final predicted results are bounding boxes of landslides.

Attention mechanism originate from the study of the human visual system, which can automatically locate useful information and suppress useless information (Mnih et al., 2014). Dai et al. (2021) compared the effects of introducing multiple attention channels, a single attention channel or no attention channel on

the model detection ability through experiments. The results show that models that introduce multiple attention channels perform best for detection. To improve the model detection ability of complex background landslides, some studies have introduced an attention mechanism into the model. Ji et al. (2020) designed a novel 3D attention module to emphasize the unique features of landslide images. Cheng et al. (2021) introduced an attention module designed based on the visual system and incorporated it into the yolov4 model for training, which improved the attention to landslide features and reduced background noise. Amankwah et al. (2022) introduced an attention module into the deep network structure to improve the ability to suppress background noise. The research results show that the attention module can significantly improve the landslide detection performance. The above research shows that introducing attention mechanism into the model is an effective way to improve the detection ability.

Although the above studies have employed deep learning algorithms for landslide detection and achieved satisfactory results, there are still some limitations. First, the majority of these models are focusing on the impact of a single attention mechanism on the detection performance, ignoring the composite effect of multiple attention mechanisms. Dai et al. (2021) compared the detection effect of single attention mechanism and compound attention mechanism, and the results proved that the compound attention mechanism achieves better detection performance. Besides, some landslide detection models mainly focus on large-proportion landslide images. For landslide images with complex backgrounds and different proportions in their size, especially small-proportion landslide images, the resulting false detection rate is high. The problem with respect to the low detection rate of small-proportion landslide images can be explained by two main reasons as below. First, small-proportion landslides occupy fewer pixels in the image, and the features of landslides are more likely to be lost during the encoding process, especially after the pooling process (Wang et al., 2018; Ghorbanzadeh et al., 2022b). For example, an image with a size of 256 × 256 is downsampled to 128 × 128 after a pooling layer, and some pixel information will be lost. Second, with the deepening of network layers and increasing receptive field, the image features of small-proportion landslides are more difficult to retain than large-proportion images (Luo et al., 2016; Krishna and Jawahar, 2018; Ajaz et al., 2022). This means that the model is more sensitive to large-proportion landslides, and it is possible to miss small-proportion landslides. For landslide detection with complex backgrounds, false detection often occurs in regions with spectral characteristics similar to those of landslide regions (Han et al., 2019; Han et al., 2022).

To address these problems, we propose the Dynahead-Yolo object detection model based on the attention mechanism. We choose YOLOv3 as the basic detection framework, which deals with the landslide detection as a mathematically regression problem and directly predicts the bounding box coordinates of the landslide area (Ju et al., 2022; Pang et al., 2022). The model has three detection branches with different scales, which is more effective for small proportion landslide detection. Based on the YOLOv3 detection framework, we redesign the detection head module. We employ scale-aware, space-aware and task-aware attention modules in the detection head, so that it can learn rich detailed features and achieve high-precision detection of landslides with variable proportions and complex background landslides. We conduct multiple comparison experiments based on the dataset consisting of three study areas. The experimental results demonstrate the feasibility and effectiveness of Dynahead-Yolo in landslide detection.

2 Materials and methods

2.1 Study areas

We select three study areas in this paper, including Ludian County in Yunnan Province, Bijie City in Guizhou Province and Beichuan County in Sichuan Province, where earthquakes and co-seismic landslides are often reported (Chang and Zhang. 2017; Ji et al., 2020). The three chosen areas have common characteristics. They are all located in SouthWestern China, with a large mountainous terrain, high mountains and sharped valleys, and staggered rivers and ditches. Landslides occur basically every year in these areas, causing serious damage to facilities such as human settlements, tunnels, roads, bridges, farmland and reservoirs, and greatly disrupting human life (Chen et al., 2016). The occurrence of landslides is also exacerbated by human production activities, such as logging, mining and agricultural production. At present, there are two main methods to obtain the location and boundaries of landslides in the study area. The first method is that experts interpret the landslide based on the images obtained by the drone, and then conduct on-site surveys to determine the landslide area and boundary. The second method is that local residents report to the government, and then the government assigns personnel to conduct landslide surveys.

2.2 Landslide datasets

In this study, we obtain remote sensing images of landslides from 91 wemap software and Bijie open-source database (Ji et al., 2020). These images are combined and pre-processed to generate the landslide datasets. We manually select a total of 950 valid and clear landslide images from the three study areas, and the selected images are additionally confirmed by experts. Part of the data consists of the Bijie open-source dataset (Ji et al., 2020), which was created for landslide segmentation. Due to the
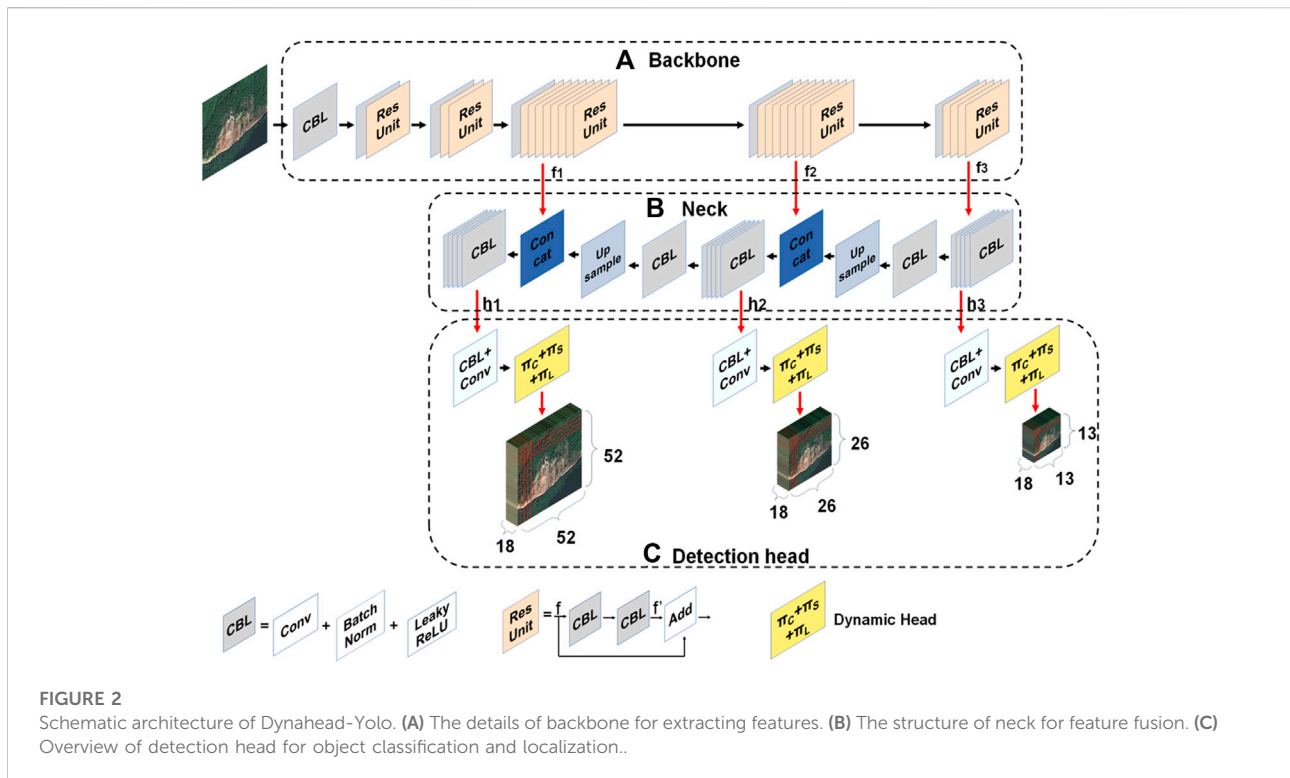
FIGURE 1
Examples of labelled images.

low resolution of images, we select approximately 200 images with a higher resolution. The rest consists of the Ludian and Beichuan areas obtained by 91 wemap software. All images are three-channel (RGB) data. We randomly select 200 images as the test set. Then, we perform data augmentation to improve the robustness of the model. We randomly select 250 images of landslides from the remaining images, and perform data augmentation by geometric transformation and noise processing. The geometric transformation mainly includes operations such as rotation, cropping, vertical flipping and horizontal flipping of landslide pictures. The noise processing includes adding salt and pepper noise, Gaussian noise and random noise to pictures. Therefore, the dataset includes a total of 1,200 images, of which 900 are used for training, 100 for validation, and 200 for testing. The proposed landslide detection model is a supervised model, so we annotate the dataset based on the Labelimg platform with PASCAL VOC2007 format. The label files mainly record the category of the object and the coordinates of the upper left and lower right corners of the label frame. Figure 1 shows a partially labelled image.

2.3 Model architecture

The schematic architecture of Dynahead-Yolo is shown in Figure 2. It consists of three parts: a backbone for extracting features, a neck for feature fusion, and a detection head for object classification and localization. The three critical components are described in detail below.

2.3.1 Backbone

Our proposed model is an improved landslide detection network based on YOLOv3. Its backbone is Darknet-53 which is similar to that of YOLOv3, and the difference is mainly on the detection head of the model. The structure of Darknet-53 is pictured in Figure 2A. It consists of one convolutional layer and five residual structures, where each residual structure contains a different number of basic residual blocks. The first and second residual structures include 1 and 2 basic residual blocks, respectively. Both the third and fourth residual structures contain 8 basic residual blocks, and the last residual structure contains 4 basic residual blocks. When a feature map tensor f is sent into the basic residual block, 1×1 and 3×3 convolutional

FIGURE 2
Schematic architecture of Dynahead-Yolo. **(A)** The details of backbone for extracting features. **(B)** The structure of neck for feature fusion. **(C)** Overview of detection head for object classification and localization..

layers are utilized for feature extraction to obtain a feature f'. Then f and f' are summed by a shortcut to acquire the output of the residual block. In addition, each residual structure also includes a 3×3 convolutional layer to compress the height and width of the input features. The convolutional layers mentioned above are followed by a batch normalization (BN) layer and a leaky rectified linear unit (Leaky ReLU). In the model structure diagram, we use a CBL module to represent the combination of these three layers, as shown in the CBL block of the legend in Figure 2.

Given an image with height H and width W, through the backbone network, we can acquire three different sizes of feature maps at the last three residual structures, denoted as $f_1 \in R^{C_1 \times \frac{H}{8} \times \frac{W}{8}}$ ($C_1 = 256$), $f_2 \in R^{C_2 \times \frac{H}{16} \times \frac{W}{16}}$ ($C_2 = 512$), and $f_3 \in R^{C_3 \times \frac{H}{32} \times \frac{W}{32}}$ ($C_3 = 1024$).

2.3.2 Neck

The main function of the neck network is to fuse the feature maps obtained from the backbone for feature enhancement. The neck network of Dynahead-Yolo includes the three branches depicted in Figure 2B. The first branch is to acquire a fusion feature map h_3, which is obtained from f_3 through five convolutional layers. The second branch merges feature maps of different sizes. The h_3 obtained from the first branch is first processed by a convolution and an up-sampling layers, and then concatenate with f_2 to obtain h_2 through five convolutional layers. The processing process of the third branch is similar to that of the second branch. The fusion feature map h_2 is

processed by a convolution and an up-sampling layers, and then concatenate with f_1 and through five convolutional layers to obtain h_1. The fusion feature maps obtained by the three branches are sent to the head module for object classification and localization. By fusing features of different scales, the neck module can make full use of the extracted information and improve the performance of the detection network.

2.3.3 Detection head

The detection head module is employed to predict the class and location of objects. In this research, we utilize the dynamic head module to improve the original head of YOLOv3, as shown in Figure 2C. The detection head module consists of three detection branches, each of which includes a convolutional layer and a dynamic head module. The dynamic head module (Dai et al., 2021) combines three attention mechanisms: spatial-aware, scale-aware and task-aware. In our Dynahead-Yolo model, we explore the effect of the connection order of three perception modules on the model performance and design the cascade order that is most useful for landslide image detection, as depicted in Figure 3. Details of the specific cascade sequence are provided in the discussion section.

The task-aware block can adapt to detection tasks by activating the channels of feature maps and improve the detection performance. The specific process is given an input feature map x, and it is first passed through an average pooling layer to reduce the feature dimension. Then, two fully connected layers

FIGURE 3
The overall structure of Dynamic Head.
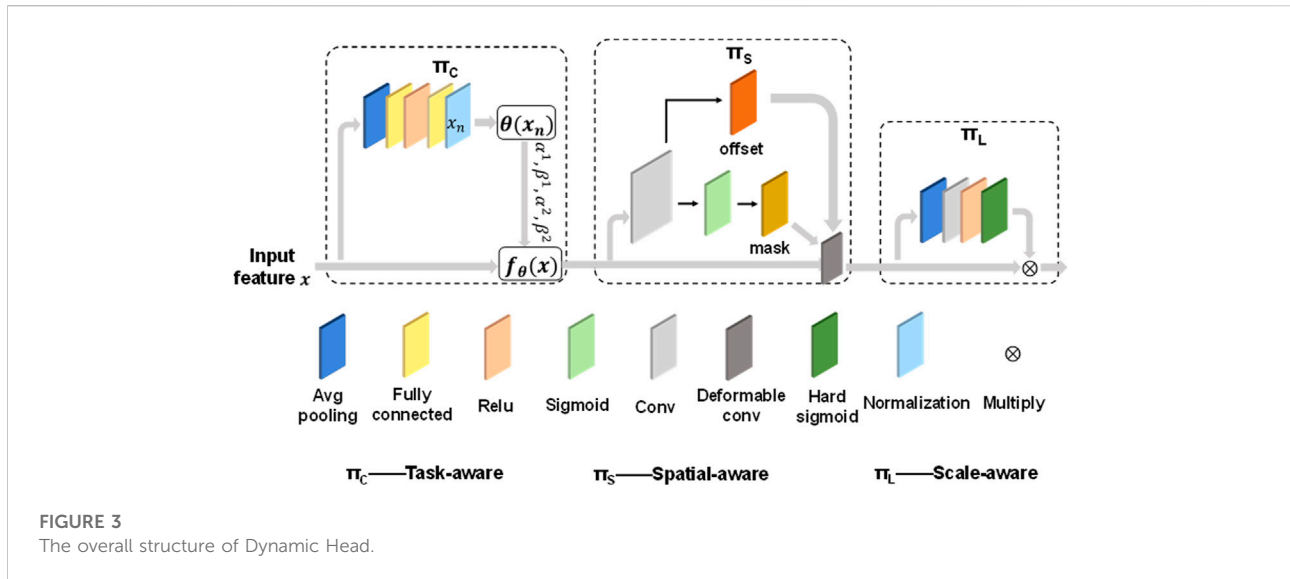
and a normalization layer are employed to map the feature to the range of −1 to 1. The normalization layer is obtained by scaling and shifting the sigmoid function and the normalized result x_n is sent into a hyperfunction $\theta(x_n)$ to generate four learnable parameters α^1, β^1, α^2 and β^2 for subsequent computations. Finally, an activation function $f_\theta(x)$ is used to dynamically activate different channels of the input feature x to obtain the final output of the task perception block. A more detailed introduction to the hyperfunction $\theta(x_n)$ the and activation function $f_\theta(x)$ can be found in the literature (Chen et al., 2020). The spatial-aware attention block is mainly deployed in the space dimension using deformable convolution to adapt to the shape and scale variations of objects. A convolutional layer is first utilized to obtain the offset and mask, which are to be applied for each position in the convolution kernel. Then, the offset, mask and input features are sent into the deformable convolution (Dai et al., 2017) layer to obtain the geometric transformation of the landslide. The scale-aware bock fuses features of different scales based on the weights of the input features. The input feature first goes through an average pooling layer to remove redundant information and reduce the number of parameters. Then it is sent into a convolutional layer with a kernel size of 1 and a ReLU activation layer. The last hard sigmoid activation function layer is utilized to speed up training. The hard ac activation function can be regarded as a classifier that approximates the sigmoid with a linear piecewise function, which is calculated by:

$$\sigma(z) = max\left(0, min\left(1, \frac{z+1}{2}\right)\right) \qquad (1)$$

The output of the activation function is multiplied by the corresponding elements of the input map to obtain the final result of the scale-aware block.

When the fusion features of three sizes obtained by the neck module are sent to the detection head module, they will go through two convolution layers and the dynamic head module for classification and positioning. The first convolutional layer is followed by an BN layer and a Leaky ReLU layer. Three detection result maps of different sizes are generated in the detection head branches, with sizes 52, 26 and 13 respectively. The number of channels for the three result plots is 18. Each grid region of three different result maps predicts 3 bounding boxes to generate a total of $(52 \times 52 + 26 \times 26 + 13 \times 13) \times 3 = 10\,647$ bounding boxes, thus the predictions for every six channels make up a vector P of each predicted bounding box, and the composition of the vector P is as follows:

$$P = \left(t_x + t_y + t_w + t_h\right) + P_0 + P_1 \qquad (2)$$

The first to fourth elements represents the coordinate information of the prediction box. P_0 means the confidence that there is an object in the predicted box, and P_1 represents the probability that the object belongs to landslides. Finally, non-maximum suppression is performed on the generated prediction frame to obtain the final prediction result.

2.4 Model evaluation metrics

To evaluate the performance of the model in terms of the detection accuracy, the precision, recall, F1 score and average precision (AP) evaluation metrics are employed in our experiments. The precision represents the size of correct predictions in the samples predicted to be landslides. The recall stands for the size of all samples that can be predicted to be landslides. The precision and recall are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

$$Recall = \frac{TP}{TP + FN} \qquad (4)$$

where TP is a true positive, FP is a false positive, and FN is a false positive. The F1 score weights precision and recall, and is a common target detection index. The F1 score is defined as follows:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (5)$$

In addition, we also utilize the AP to evaluate the performance of model. First, we need to sort the detection results in descending order according to the confidence of the prediction, and then calculate the precision and recall of the cumulative for each sample, and finally draw the precision-recall curve. The AP value is calculated through the area under the curve. The calculation formula is as follows:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) \times P_{inter} \qquad (6)$$

$r_1, r_2, \ldots, r_{i+1}$ are the values of recall in ascending order, $r_{i+1} - r_i$ means the deviation between adjacent recall values; P_{inter} is the maximum value of the corresponding precision when the recall rates are r_{i+1} and r_i. In addition, we adopt intersection over union (IoU) to measure the overlapping area between the detection and label boxes. The larger the area is, the more credible the detection box. IoU is defined as follows:

$$IoU = \frac{P \cap G}{P \cup G} \qquad (7)$$

where p and G represent the prediction box and label box, respectively. $\cap$ and $\cup$ indicate the intersection and union between two bounding boxes, respectively.

3 Model training and results

3.1 Model training

To evaluate the effectiveness of the proposed model, we conduct a series of comparative experiments and ablation experiments. To ensure the same hardware conditions for each experiment, all object detection networks are trained based on the same dataset, and all experiments are performed on a server with an NVIDIA RTX2080Ti, 12 GB of GPU memory, and a 6× Xeon E5-2678 v3 CPU. The weights of the model are initialized according to the pretrained model on the VOC2007 dataset. During training, the max epoch is 200 epochs for each experiment, including 100 freeze training epochs and 100 non-freeze training epochs. Freeze training means that the parameters of the backbone feature extraction network will not be updated during the training process. Non-freezing training

will update all parameters of the model during training. This training method can improve the training speed and make the model converge quickly. The Adam optimizer with an initial learning rate of 3×10^{-4} and decayed weights of 5×10^{-4} is utilized for model training. The learning rate decays by a fixed step size of 1, and the decay coefficient is .94.

We plot the loss variation curves for the training set and validation set, as shown in Figure 4A. The model loss gradually decreases during the training process, and finally oscillates and stabilizes, indicating that the model has been fully trained. In addition, we calculate the mean average Precision (mAP) of the training weights on the training and validation datasets and plot the curves, as shown in Figure 4B The mAP of the training and validation datasets are stable at approximately 97.8% and 83.2%, respectively, suggesting that the proposed Dynahead-Yolo has an excellent effect on landslide detection. We select the model weight of the 185th epoch with the smallest validation loss for testing, and compare it with other models to verify the feasibility of the study.

3.2 Results

To illustrate the recognition ability of Dynahead-Yolo for landslides in complex backgrounds and the detection ability with variable proportions landslide images, especially for small-proportion landslides. We randomly generated 200 test images and count the number of different proportion landslides. The object detection model extracts features from pixels in the local area through the sliding window. We could classify landslides into large, medium and small proportion landslides according to the proportion of pixels in the image. The proportion of pixels also represents the percentage of the landslide area to the image area, when the landslide area accounts for more than 50% of the total image area, it can be considered as a large-proportion landslide. If the area ratio is between 10% and 50%, it can be regarded as a medium-proportion landslide, and when the area ratio is less than 10%, it can be treated as a small-proportion landslide. As follows:

$$\alpha = \frac{S_L}{S_I} \qquad (8)$$

$$\begin{cases} \alpha \leq 10\% & small - proportion\ landslides \\ 10\% < \alpha \leq 50\% & medium - proportion\ landslides \\ \alpha > 50\% & large - proportion\ landslides \end{cases} \qquad (9)$$

where S_L is the area of the landslides; S_I represents the area of the entire image; and α represents the area ratio, which is the proportion of landslides in the image.

The complexity of the background mainly includes four aspects: 1) interference from clouds and fog when imaging, 2) interference from houses near the landslide, 3) interference from bare sand with similar characteristics to the landslide image, and 4) interference from terraces in mountainous areas. Based on the
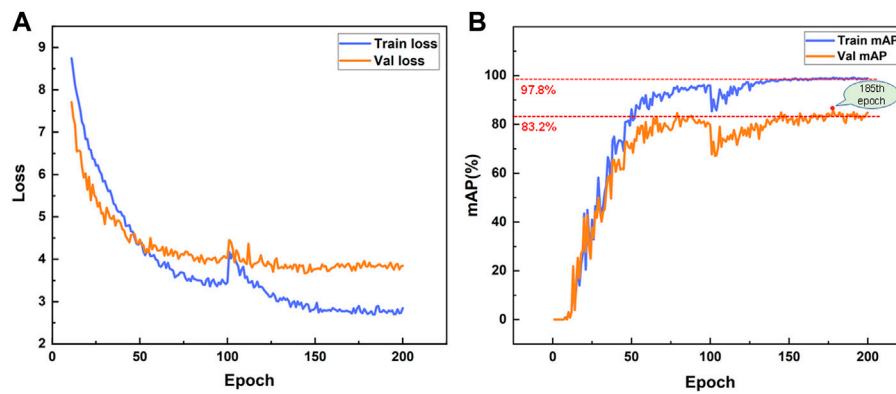
FIGURE 4
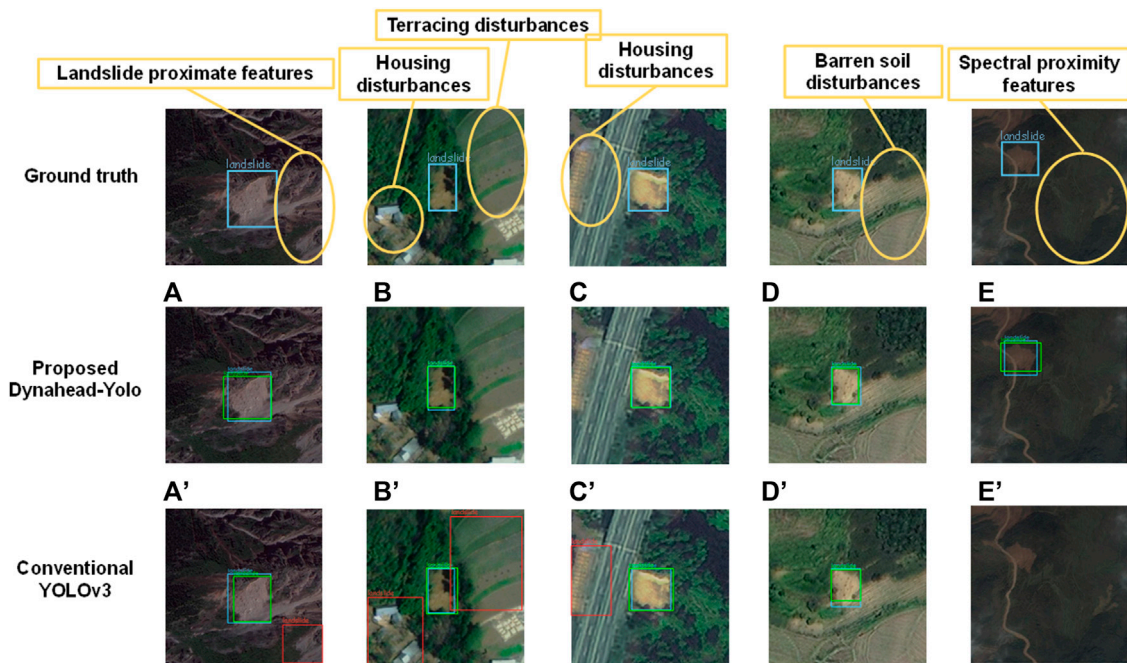**(A)** Loss curve during training **(B)** Model mAP curve.

FIGURE 5
Complex background landslide detection. **A–E** show the landslide detection results of Dynahead-Yolo with different complex backgrounds.
A′–E′ represent the detection results of YOLO v3 in the corresponding complex background.

above criteria, we counted the number of landslide images with complex backgrounds in the test set, for a total of 80 images. We employed the developed Dynahead-Yolo to detect complex background landslides images, and compared them with the prediction results of YOLOv3. Figure 5 shows the ground truth, detected results of Dynahead-Yolo and YOLOv3. The blue bounding box is manually annotated label. The green bounding box is the correctly predicted detection box, which

should ideally overlap with the label, and the red box is the result of the wrong prediction. YOLOv3 mistakenly detected landslide adjacent areas, terraced fields, and houses as landslides, as listed in Figures 5A′–C′. Besides, in Figures 5E′, YOLOv3 cannot detect landslides when the spectral radiance of the target region is similar to that of the background region. Overall, the proposed model can well identify landslide images in such complex backgrounds and achieve higher detection accuracy.
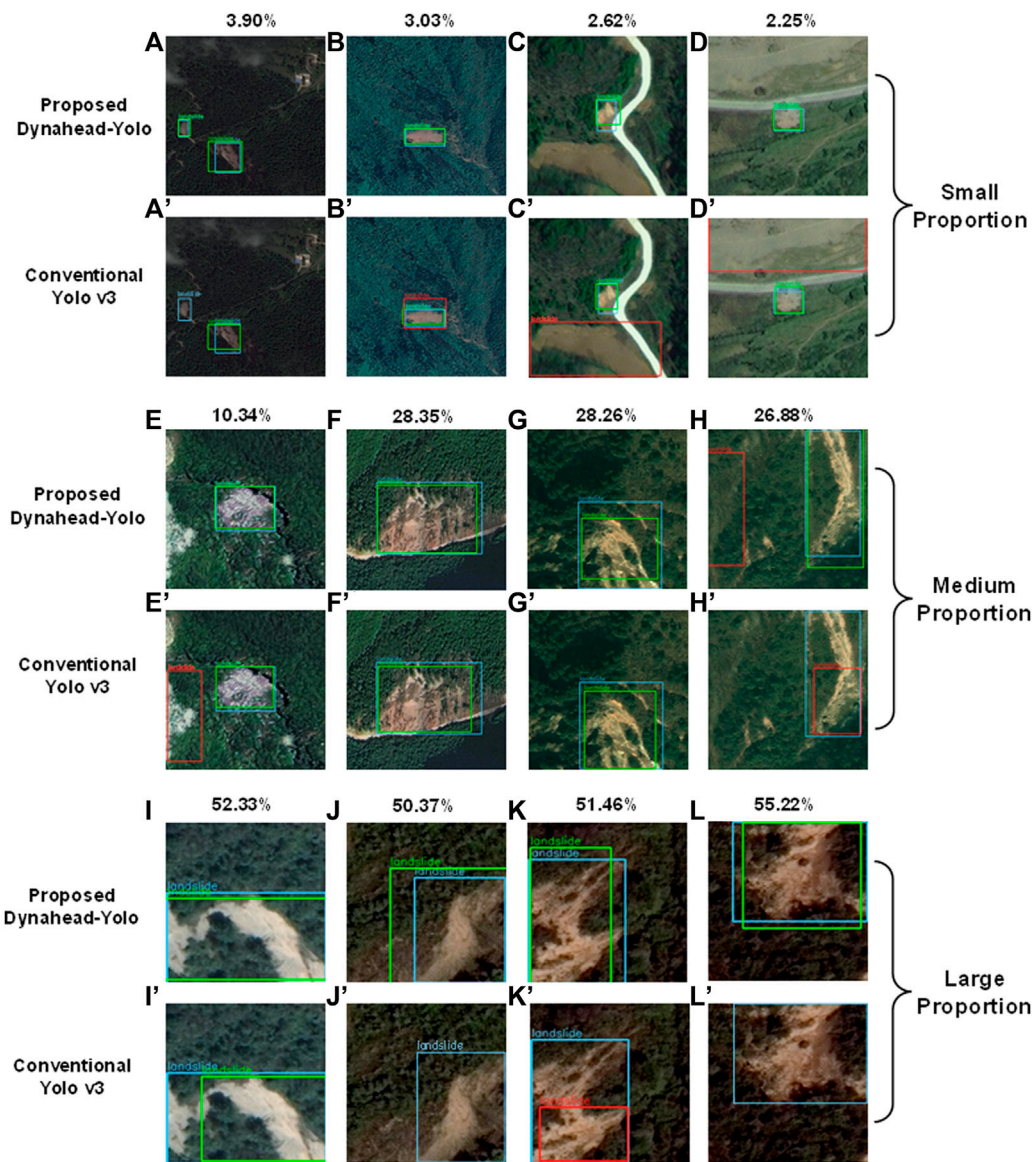
FIGURE 6
Landslides detection with variable sizes. **A−D** and **A'−D'** show the detection results of Dynahead-Yolo and YOLO v3 for small-proportion landslides, respectively. **E−H** and **E'−H'** compare the effect of Dynahead-Yolo and YOLO v3 on the medium-proportion landslides. **I−L** and **I'−L'** represent the detection results of Dynahead-Yolo and YOLO v3 for large-proportion landslides, separately.

The detection results of landslides with variable proportions are shown in Figure 6. We compared the results of two different models in small, medium, and large proportion landslides. The number above the image indicates the percentage of landslides in the image, which is the proportion of the area of a detected landslide compared to the image proportion. In Figures 6A', J', L', we found that no matter whether the proportion of landslides is too small or too large, the Dynahead-Yolo can locate the landslide and achieve accurate bounding box prediction. As shown in Figures 6C', D', the YOLOv3 model may mistake

TABLE 1 Percentage of correct landslides detection with variable proportions and complex backgrounds.

	Dynahead-Yolo	YOLOv3
Complex background landslides (%)	**82.33**	72.10
Small-proportion landslides (%)	**86.63**	76.21
Medium-proportion landslides (%)	**83.33**	79.22
Large-proportion landslides (%)	**90.00**	83.51
Average accuracy (%)	**85.57**	77.76

TABLE 2 Performance of the model with different concatenation sequences.

	Precision (%)	Recall (%)	F1	AP (%)
$\pi_L \to \pi_S \to \pi_C$	86.34	89.30	0.87	83.91
$\pi_L \to \pi_C \to \pi_S$	86.34	78.67	0.82	75.25
$\pi_S \to \pi_L \to \pi_C$	85.65	84.89	0.85	81.35
$\pi_S \to \pi_C \to \pi_S$	87.20	81.78	0.84	79.21
$\pi_C \to \pi_L \to \pi_S$	89.40	86.22	0.87	83.92
$\pi_C \to \pi_S \to \pi_L$	**87.17**	**87.56**	**0.87**	**85.53**

exposed soil for landslide. In addition, it is obvious that the Dynahead-Yolo can achieve higher IoU between the label and detected box in Figures 6I, I′. A more detailed comparison of the distribution of the IoU is provided in the Discussion section.

To further verify the ability to detect landslide images with complex environments and variable proportions, we counted the percentage of the correct prediction results of two different models in different scenarios, as summarized in Table 1. According to the comparison results, the correct detection rate of our method for complex background landslides reaches 82.3%, which is 14.1% higher than that of YOLOv3. Moreover, our method also improved the detection rate of small-proportion landslides and large-proportion landslides by 13.6% and 7.8%, respectively. The results show the advantages of the Dynahead-Yolo model for landslides with complex backgrounds and variable proportions of landslide detection.

4 Discussion

4.1 Model results for different concatenation sequences

We first investigate the effect of the cascade order for three attention modules on the model performance. As shown in Table 2, we evaluate the performance of six concatenation sequences for landslide detection from fort aspects: precision, recall, F1 value, and AP value. According to the comparison results, we find that different concatenation sequences have a significant impact on the performance of the model. The detection performance in the order of task-aware, space-aware and scale-aware is the best, reaching an AP value of 85.53%. The cascade sequence first activates different channels according to the detection task, then enhances the spatial location features of foreground objects through a spatial-aware attention module, and finally improves the detection ability of landslide areas with

different proportions through a scale-aware attention module.
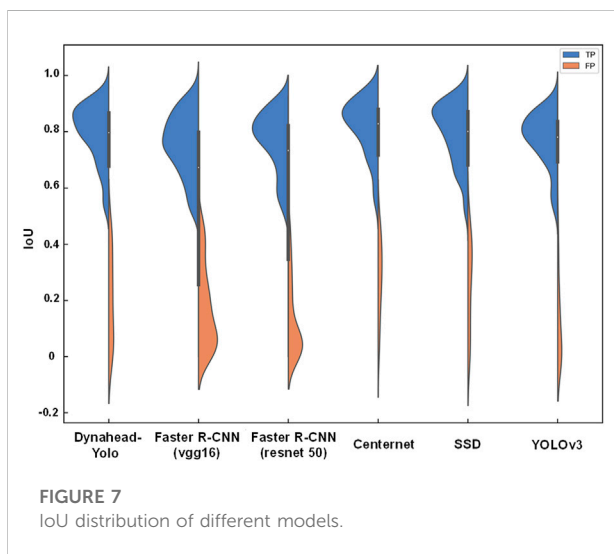
4.2 Comparison of different detection models

To further demonstrate the performance of Dynahead-Yolo, we compare the detection results of previous object models such as Faster R-CNN, Faster R-CNN, YOLOv3, SSD, and Centernet. All models were trained based on the RTX2080TI GPU with 200 epochs each and tested with the same dataset. Table 3 summarizes the comparison results for the four evaluation indexes of different models. The faster R-CNN with resnet50 backbone network has the highest recall value, but its precision is only 45.56%, which shows that the model has high false positives in landslide detection. Faster R-CNN is a two-stage detection model. The first stage is to generate many proposals, and the second stage adjusts the coordinates of proposals. In the first step, all areas suspected of landslides will be detected as proposals, which leads to a high false detection rate. Dynahead-Yolo directly extracts features from the network and predicts the location of landslides without generating proposals. The addition of the compound attention module enhances the ability of the model to acquire features of different scales and spatial features, enabling to obtain better detection results. The precision and recall can evaluate the performance from different perspectives. We usually employ the F1 score and mAP combining the two indexes to comprehensively evaluate the effect of the model. The F1 score and mAP of Dynahead-Yolo are 0.87 and 85.53%, respectively, which are 0.02 and 6.95% higher than those of YOLOv3. The results show that the proposed method is more suitable for the automatic detection of landslide remote sensing images.

We calculate the IoUs of the detection results for each model and plotted a violin plot, where the horizontal coordinates are the individual model and the vertical coordinates are the IoU values, as pictured in Figure 7. The blue and red areas represent the number of true positive and false positive samples, respectively. From Figure 7, it is clear that the IoUs of the correct prediction results for Dynahead-Yolo are mainly concentrated in the range of 0.8–0.87, which is obviously higher than that of other models. The results indicate that the coordinates of detection and prediction bounding boxes are close to each other.

To illustrate the performance of Dynahead-Yolo for landslide with variable proportions, we compare it with the results of other studies. We select three published papers (Ye et al., 2019; Ju et al., 2022; Li et al., 2022) based on deep learning methods for landslide detection, and calculate the percentage of landslides in the detected images shown in each paper. We draw a box-plot and represent the results of the comparison, as shown in Figure 8. It can be considered that the paper shows

TABLE 3 Models performance.

Description	Precision (%)	Recall (%)	F1	mAP (%)
Dynahead-Yolo	**87.17**	**87.56**	**0.870**	**85.53**
Faster R-CNN (resnet50)	45.56	91.11	0.610	83.06
Faster R-CNN (vgg16)	45.09	85.78	0.590	74.17
YOLOv3	89.60	80.44	0.850	78.58
SSD	81.28	79.11	0.800	71.00
Centernet	88.89	56.89	0.70	54.82

FIGURE 7
IoU distribution of different models.

FIGURE 8
Comparison of landslides detection results with variable sizes.

representative landslide images reflecting the detection effect of the model. Therefore, the drawn box-plot shows that the Dynahead-YOLO model is more suitable for detecting a certain proportion of landslides. According to Figure 8, it can be observed that the models (Ye et al., 2019; Ju et al., 2022) are more useful for medium-proportion and large-proportion landslides. The proportion of landslides in Paper (Li et al., 2022) is mainly distributed in areas greater than 50%, and the model of the article is more effective for large-proportion landslides. Dynahead-Yolo has good detection performance for different proportions of landslides.

4.3 Limitations and future research

In this study, we propose a named Dynahead-Yolo object detection model for high accuracy detection in complex environments and landslides with variable proportions. We discuss the effects of various concatenation sequences for three attention modules on the detection capability of the model, and compare them with classical object detection networks. Nevertheless, there are some problems in this study due to the limitation of data resources and experimental equipment. First of all, there are very few public datasets related to landslides, and it is difficult to obtain effective and clear landslide data. Therefore, the study only utilizes 1,200 landslide images, and the amount of data is small. Besides, the labels for training are manually labeled. Although the labels have been confirmed and evaluated by experts, there will still be errors, which will affect the accuracy and precision of the model. In the future research, we will acquire more landslide data containing various types of remote sensing images to improve the generalization and detection capability of the model.

In subsequent studies, we believe that it remains an important means to improve the performance of the model by an attention mechanism. We will also try to combine the Dynamic head with other object detection models to achieve the purpose of improving detection capabilities.
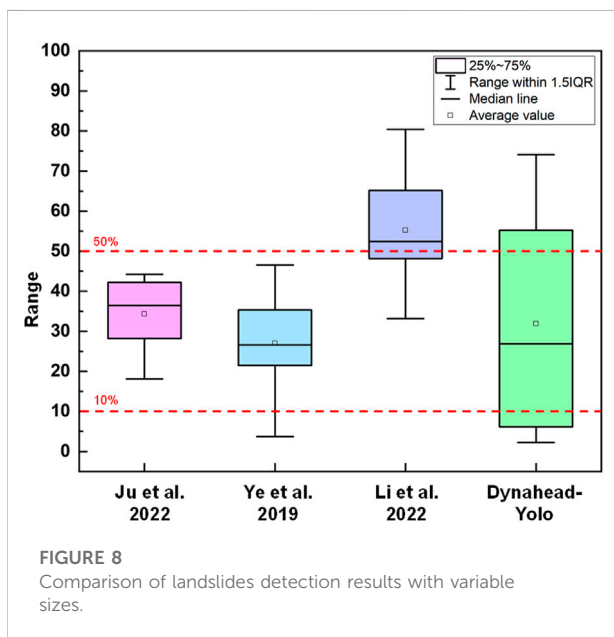
5 Conclusion

In this paper, we proposed a novel Dynahead-Yolo neural network for the detection of landslides using remote sensing images. This neural network was specifically designed to address the poor detection of the small proportion landslides in the remote sensing images. The combination of the three attention modules including scale-aware, spatial-aware, and task-aware enhanced the feature extraction ability and improved the adaptability to landslide images with different proportions and complex backgrounds. Landslide images from Bijie city, Ludian County and Beichuan County were collected to generate datasets and randomly separated as prediction sets to verify the performance of the model. Results show that, compared with the conventional YOLOv3, the accuracy of the proposed Dynahead-Yolo for complex backgrounds and small-proportion landslides were improved by 14.19% and 13.67%, respectively, while the F1 score and AP of the model were 0.87 and 85.53%, respectively. The results indicate an outperforming ability of the proposed Dynahead-Yolo for detecting small proportion landslides comparing to the conventional YOLOv3. However, the performance of the model may be limited by the number of datasets and the reliability of the labels. Therefore, increasing the dataset or designing a suitable model to adapt to small sample landslide detection is the main work in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/Abbott-max/dataset; doi:10.4121/20401740.

Author contributions

ZH contributed to the conception of the study; ZF performed the experiment; BF contributed significantly to analysis and manuscript preparation; ZF performed the data analyses and wrote the manuscript; YL helped perform the analysis with constructive.

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ajaz, A., Salar, A., Jamal, T., and Khan, A. U. (2022). Small object detection using deep learning[J]. arXiv e-prints. doi:10.48550/arXiv.2201.03243

Alam, E. (2020). Landslide hazard knowledge, risk perception and preparedness in Southeast Bangladesh. *Sustainability* 12 (16), 6305. doi:10.3390/su12166305

Amankwah, S. O. Y., Wang, G., Gnyawali, K., Hagan, D. F. T., Sarfo, I., Zhen, D., et al. (2022). Landslide detection from bitemporal satellite imagery using attention-based deep neural networks. *Landslides* 19 (10), 2459–2471. doi:10.1007/s10346-022-01915-6

Cai, H., Chen, T., Niu, R., and Plaza, A. (2021). Landslide detection using densely connected convolutional networks and environmental conditions. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 14, 5235–5247. doi:10.1109/JSTARS.2021.3079196

Chang, Hao, and Zhang, Lv (2017). Analysis of landslide susceptibility in the earthquake zone of Ludian Ms6.5 earthquake. *Chin. J. Geol. Hazards Prev.* 28 (02), 38–48. doi:10.16031/j.cnki.issn.1003-8035.2017.02.06

Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., and Liu, Z. (2020). "Dynamic relu," in *European conference on computer vision* (Cham: Springer), 351–367. doi:10.1007/978-3-030-58529-7_21

Chen, Z., Meng, X., Yin, Y., Dijkstra, T., Winter, M., and Wasowski, J. (2016). Landslide research in China. *Q. J. Eng. Geol. Hydrogeology* 49 (4), 279–285. doi:10.1144/qjegh2016-100

Cheng, L., Li, J., Duan, P., and Wang, M. (2021). A small attentional YOLO model for landslide detection from satellite remote sensing images. *Landslides* 18 (8), 2751–2765. doi:10.1007/s10346-021-01694-6

Chumerin, N. (2017). Convolutional neural network. Apress. doi:10.1007/978-1-4842-2845-6_6

Clague, J. J., Roberts, N. J., and Stead, D. (2012). "Landslide hazard and risk," in *Landslides: Types, mechanisms and modeling*. Editors J. J. Clague and D. Stead (Cambridge, UK: Cambridge University Press), 1–9. doi:10.1017/CBO9780511740367.002

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., et al. (2017). "Deformable convolutional networks," in Proceedings of the IEEE international conference on computer vision, Venice, Italy, Oct. 22 2017 to Oct. 29 2017, 764–773. doi:10.1109/ICCV.2017.89

Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., et al. (2021). "Dynamic head: Unifying object detection heads with attentions," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20-25 June 2021, 7373–7382. doi:10.48550/arXiv.2106.08322

Eeckhaut, M. V. D., Kerle, N., and Hervás, J. (2011). "Mapping vegetated landslides using LiDAR derivatives and object-oriented analysis," in Ninth International Workshop on Remote Sensing for Disaster Response, Stanford, USA, 15-16 September 2011.

Fiorucci, F., Ardizzone, F., Mondini, A. C., Viero, A., and Guzzetti, F. (2019). Visual interpretation of stereoscopic NDVI satellite images to map rainfall-induced landslides. *Landslides* 16, 165–174. doi:10.1007/s10346-018-1069-y

Fu, B., Li, Y., Han, Z., Fang, Z., Ceng, N., et al. (2022). RIPF-unet for regional landslides detection: A novel deep learning model boosted by reversed image pyramid features. doi:10.21203/rs.3.rs-1886017/v1

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S., Tiede, D., and Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* 11 (2), 196. doi:10.3390/rs11020196

Ghorbanzadeh, O., Crivellari, A., Ghamisi, P., Shahabi, H., and Blaschke, T. (2021). A comprehensive transferability evaluation of U-Net and ResU-Net for landslide detection from Sentinel-2 data (case study areas from Taiwan, China, and Japan). *Sci. Rep.* 11 (1), 14629–14720. doi:10.1038/s41598-021-94190-9

Ghorbanzadeh, O., Meena, S. R., Abadi, H. S. S., Piralilou, S. T., Zhiyong, L., and Blaschke, T. (2020). Landslide mapping using two main deep-learning convolution neural network streams combined by the dempster–shafer model. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 14, 452–463. doi:10.1109/JSTARS.2020.3043836

Ghorbanzadeh, O., Xu, Y., Zhao, H., Wang, J., Zhong, Y., Zhao, D., et al. (2022a). The outcome of the 2022 Landslide4Sense competition: Advanced landslide detection from multi-source satellite imagery. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 9927–9942. doi:10.1109/JSTARS.2022.3220845

Ghorbanzadeh, O., Xu, Y., Ghamis, P., Kopp, M., and Kreil, D. (2022b). Landslide4Sense: Reference benchmark data and deep learning models for landslide detection. arXiv preprint arXiv:2206.00515. doi:10.1109/TGRS.2022.3215209

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *Adv. neural Inf. Process. Syst.* 63, (11) 139–144. doi:10.1007/978-3-030-50017-7_16

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K. T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Rev.* 112 (1–2), 42–66. doi:10.1016/j.earscirev.2012.02.001

Han, Z., Fang, Z., Fu, B., Wu, B., Li, Y., Li, C., et al. (2022). A multi-feature fusion interpretation method for optical remote sensing images of isoseismic landslides. *Chin. J. Geol. Hazards Prev.* 33 (6), 104–114. doi:10.16031/j.cnki.issn.1003-8035.202111008

Han, Z., Su, B., Li, Y. G., Ma, Y. F., Wang, W. D., and Chen, G. Q. (2019). An enhanced image binarization method incorporating with Monte-Carlo simulation. *J. Central South Univ.* 26 (6), 1661–1671. doi:10.1007/s11771-019-4120-9

Hao, X., Zhang, G., and Ma, S. (2016). Deep learning. *Int. J. Semantic Comput.* 10 (03), 417–439. doi:10.1142/S1793351X16500045

Hölbling, D., Tsai, T. T., Eisank, C., Friedl, B., Shieh, C. L., and Blaschke, T. (2014). "Pixel-based and object-based landslide mapping: A methodological comparison," in Geological Society of America-Abstracts with Programs, Vancouver, Canada, 19-22 October 2014.

Hou, H., Chen, M., Tie, Y., and Li, W. (2022). A universal landslide detection method in optical remote sensing images based on improved YOLOX. *Remote Sens.* 14 (19), 4939. doi:10.3390/RS14194939

Hu, Q., Zhou, Y., Wang, S., Wang, F., and Wang, H. (2019). Improving the accuracy of landslide detection in "off-site" area by machine learning model portability comparison: A case study of jiuzhaigou earthquake, China. *Remote Sens.* 11 (21), 2530. doi:10.3390/rs11212530

Hungr, O., Leroueil, S., and Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides* 11 (2), 167–194. doi:10.1007/s10346-013-0436-y

Ji, S., Yu, D., Shen, C., Li, W., and Xu, Q. (2020). Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* 17 (6), 1337–1352. doi:10.1007/s10346-020-01353-2

Jiang, W., Xi, J., Li, Z., Ding, M., Yang, L., and Xie, D. (2021). Mask R-CNN landslide segmentation and recognition of simulated difficult samples. *J. Wuhan Univ. Inf. Sci. Ed.* doi:10.13203/j.whugis20200692

Ju, Y., Xu, Q., Jin, S., Li, W., Dong, X., and Guo, Q. (2020). Using deep learning method to realize automatic identification of loess landslides. *J. Wuhan Univ. Inf. Sci. Ed.* 45 (11), 1747–1755. doi:10.13203/j.whugis20200132

Ju, Y., Xu, Q., Jin, S., Li, W., Su, Y., Dong, X., et al. (2022). Loess landslide detection using object detection algorithms in northwest China. *Remote Sens.* 14 (5), 1182. doi:10.3390/rs14051182

Korup, O., and Stolle, A. (2014). Landslide prediction from machine learning. *Geol. today* 30 (1), 26–33. doi:10.1111/gto.12034

Krishna, H., and Jawahar, C. V. (2018). "Improving small object detection[C]," in 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), Nanjing, China, Nov. 26 2017 to Nov. 29 2017. doi:10.1109/ACPR.2017.149

Li, H., He, Y., Xu, Q., Deng, J., Li, W., and Wei, Y. (2022). Detection and segmentation of loess landslides via satellite images: A two-phase framework. *Landslides* 19 (3), 673–686. doi:10.1007/s10346-021-01789-0

Liu, J., Wu, Y., Gao, X., and Si, W. (2022). Coseismic landslide identification method based on gee and u-net models. *Chin. J. Earth Inf. Sci.* 24 (7), 11. doi:10.12082/dqxxkx.2022.210704

Li, Y., Chen, G., Han, Z., Zheng, L., and Zhang, F. (2014). A hybrid automatic thresholding approach using panchromatic imagery for rapid mapping of landslides. *GIScience Remote Sens.* 51 (6), 710–730. doi:10.1080/15481603.2014.972867

Liu, X., Ou, O., Zhang, W., and Du, X. (2022). Landslide image detection fusion attention mechanism and atrous convolution. *Comput. Mod.* (04), 45–51. doi:10.3969/j.issn.1006-2475.2022.04.009

Liu, Y., and Wu, L. (2016). Geological disaster recognition on optical remote sensing images using deep learning. *Procedia Comput. Sci.* 91, 566–575. doi:10.1016/j.procs.2016.07.144

Lu, P., Stumpf, A., Kerle, N., and Casagli, N. (2011). Object-oriented change detection for landslide rapid mapping. *IEEE Geoscience remote Sens. Lett.* 8 (4), 701–705. doi:10.1109/LGRS.2010.2101045

Luo, W., Li, Y., Urtasun, R., and Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. *Adv. neural Inf. Process. Syst.* 29. doi:10.48550/arXiv.1701.04128

Meena, S. R., Nava, L., Bhuyan, K., Puliero, S., Soares, L. P., Dias, H. C., et al. (2022). HR-GLDD: A globally distributed dataset using generalized dl for rapid landslide mapping on hr satellite imagery. *Earth Syst. Sci. Data Discuss.*, 1–21. doi:10.5194/essd-2022-350

Mnih, V., Heess, N., Graves, A., and Kavukcuoglu, K. (2014). Recurrent models of visual attention. *Adv. Neural Inf. Process. Syst.* 2, 2204–2212. doi:10.48550/arXiv.1406.6247

Moosavi, V., Talebi, A., and Shirmohammadi, B. (2014). Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. *Geomorphology* 204, 646–656. doi:10.1016/j.geomorph.2013.09.012

Pang, D., Liu, G., He, J., Li, W., and Fu, R. (2022). Automatic remote sensing identification of Co-seismic landslides using deep learning methods. *Forests* 13 (8), 1213. doi:10.3390/f13081213

Parker, O. P. (2013). Object-based segmentation and machine learning classification for landslide detection from multi-temporal WorldView-2 imagery California: California State University. Masters Thesis.

Petschko, H., Bell, R., and Glade, T. (2016). Effectiveness of visually analyzing LiDAR DTM derivatives for Earth and debris slide inventory mapping for statistical susceptibility modeling. *Landslides* 13, 857–872. doi:10.1007/s10346-015-0622-1

Piralilou, S. T., Blaschke, T., and Ghorbanzadeh, O. (2019). "An integrated approach of machine-learning models and Dempster-Shafer theory for landslide detection," in Living Planet Symposium, Milan, Italy, 13–17 May 2019. doi:10.13140/RG.2.2.33724.18567

Sandric, I., Mihai, B., Chitu, Z., Gutu, A., and Savulescu, I. (2010). "Object-oriented methods for landslides detection using high resolution imagery, morphometric properties and meteorological data," in ISPRS TC VII Symposium, Vienna, Austria, July 5–7, 2010.

Singh, P., Maurya, V., and Dwivedi, R. (2021). "Pixel based landslide identification using landsat 8 and GEE," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11-16 July 2021 (IEEE), 8444–8447. doi:10.1109/IGARSS47720.2021.9553358

Stumpf, A., and Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote Sens. Environ.* 115 (10), 2564–2577. doi:10.1016/j.rse.2011.05.013

Valdés Carrera, A. C., Mendoza, M. E., Allende, T. C., and Macías, J. L. (2021). A review of recent studies on landslide hazard in Latin America. *Phys. Geogr.*, 1–44. doi:10.1080/02723646.2021.1978372

Vargas, R., Mosavi, A., and Ruiz, L. (2017). Deep learning A review. In Advances in Intelligent Systems and Computing 5 (2) (Accessed June 2017).

Wang, G., Xiong, Z., Dong, L., and Luo, C. (2018). "Cascade mask generation framework for fast small object detection[C]," in 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23-27 July 2018 (IEEE Computer Society). doi:10.1109/ICME.2018.8486561

Wang, Y., Ren, G., Wang, J., Wang, M., and Yu, T. (2017). A review of remote sensing interpretation of landslides. *Northwest Hydropower* 1, 17–21. doi:10.3969/j.issn.1006-2610.2017.01.005

Xu, C., Shyu, J. B. H., and Xu, X. (2014). Landslides triggered by the 12 january 2010 port-au-prince, Haiti, <i&gt;M&lt;/i&gt;<sub&gt;w&lt;/sub&gt; = 7.0 earthquake: Visual interpretation, inventory compiling, and spatial distribution statistical analysis. *Nat. Hazards Earth Syst. Sci.* 14, 1789–1818. doi:10.5194/nhess-14-1789-2014

Ye, C., Li, Y., Cui, P., Liang, L., Pirasteh, S., Marcato, J., et al. (2019). Landslide detection of hyperspectral remote sensing data based on deep learning with constrains. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 12 (12), 5047–5060. doi:10.1109/JSTARS.2019.2951725

Yu, Z., Chang, R., and Chen, Z. (2022). Automatic detection method for loess landslides based on GEE and an improved YOLOX algorithm. *Remote Sens.* 14 (18), 4599. doi:10.3390/rs14184599

Zaremba, W., Sutskever, I., and Vinyals, O. (2014). Recurrent neural network regularization. Eprint Arxiv. doi:10.48550/arXiv.1409.2329

Zhao, W., Li, A., Nan, X., Zhang, Z., and Lei, G. (2017). Postearthquake landslides mapping from Landsat-8 data for the 2015 Nepal earthquake using a pixel-based change detection method. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 10 (5), 1758–1768. doi:10.1109/JSTARS.2017.2661802