

# Landslide Detection Based on Efficient Residual Channel Attention Mechanism Network and Faster R-CNN

Yabing Jin<sup>1</sup>, Ou Ou<sup>2,\*</sup>, Shanwen Wang<sup>2</sup>, Yijun Liu<sup>1</sup>, Haoqing Niu<sup>2</sup>, and Xiaopeng Leng<sup>2</sup>

<sup>1</sup> Geological Bureau of Shenzhen, Shenzhen 518028, China  
jinyabing25@sina.com

<sup>2</sup> College of Computer and Network Security, Chengdu University of Technology,  
Chengdu 610051, Sichuan, China  
ouou@cdut.edu.cn

**Abstract.** Accurate landslide detection plays an important role in land planning, disaster prediction and disaster relief. At present, field investigation and exploration based on professional personnel is the most widely used landslide mapping and detection technology, but this method consumes a lot of manpower and material resources and is inefficient. With the development of artificial intelligence, landslide identification and target detection based on deep learning have attracted more and more attention due to their remarkable advantages over traditional technologies. It is a technical problem to identify landslides from satellite remote sensing images. Although there are some methods at present, there is still room for improvement in the target detection algorithm of landslides against the background of the diversity and complexity of landslides. In this paper, target detection algorithm models such as Faster R-CNN apply to landslide recognition and detection tasks, and various commonly used recognition and detection algorithm network structures are used as the basic models for landslide recognition. Efficient residual channel soft thresholding attention mechanism algorithm (ERCA) is proposed, which intends to reduce the background noise of images in complex environments by means of deep learning adaptive soft thresholding to improve the feature learning capability of deep learning target detection algorithms. ERCA is added to the backbone network of the target detection algorithm for basic feature extraction to enhance the feature extraction and expression capability of the network. During the experiment ERCA combined with ResNet50, ResNet101 and other backbone networks, the objective indicators of detection results such as AP50 (Average Precision at IOU=0.50), AP75 (Average Precision at IOU=0.75) and AP (Average Precision) were improved, and the AP values were all improved to about 4%, and the final detection results using ResNet101 combined with ERCA as the backbone network reached 76.4% AP value. ERCA and other advanced channel attention networks such as ECA (Efficient Channel Attention for Deep Convolutional Neural Networks) and SENet (Squeeze-and-Excitation Networks) are fused into the backbone network of the target detection algorithm and experimented on the landslide identification detection task, and the detection results are that the objective detection indexes AP50, AP75, AP, etc. are higher for ERCA compared with other channel attention, and the subjective detection image detection effect and feature map visualization display are also better.<sup>3</sup>

\* Corresponding author

<sup>3</sup> We released our code at: <https://github.com/fluoritess/Efficient-residual-channel-attention-mechanism-network-and-Faster-R-CNN>.

**Keywords:** landslide detection, deep learning, Faster R-CNN, ERCA.

## 1. Introduction

Landslide is a common geological natural disaster, causing serious damage to the natural environment, personal safety and property of all countries. Landslides may be caused by many factors, including earthquake [1, 2], heavy rainfall [3,4], human factors[5], etc. Field investigation of potential landslide areas by professionals is a common and reliable method, but this is time-consuming, expensive and inefficient [6], especially for large-area landslide detection. Due to the above reasons, more and more scholars have started to explore semi-automated or automated landslide detection methods based on remote sensing images in the last decade or so [38].

Remotely sensed images are images acquired from ground observations by aerial aircraft or artificial satellites. Based on the acquisition method, remote sensing images can be classified into the categories of SAR images, infrared images, multispectral images, and visible images [30]. Due to synthetic aperture radar (SAR) images based on microwave coherence imaging have a single color and lack texture detail information; multispectral images have poor resolution and image information is difficult to understand; infrared images are more suitable for identifying heat-emitting targets, visible images become the most commonly used remote sensing image category in landslide detection, and visible images have intuitive content, high resolution, and contain a large amount of information, with rich spatial information, clear geometric structure and texture information, and can truly reflect the ground geographic conditions [31-34]. Therefore, the improved model as well as the chosen dataset in this paper are for remote sensing images in the visible light category. Because most of the landslides are small in scale, large in number and the surrounding environment of the landslide is complex, detecting landslides from remote sensing images is a very challenging problem [7].

At present, there are two main methods for landslide detection in remote sensing images: one is the traditional machine learning-based landslide detection method for remote sensing images [35], which firstly uses two methods, pixel-based method or object-based method [8], to obtain the suspected landslide area in remote sensing images. In the pixel-based landslide detection method, a single pixel in the remote sensing image is the most basic processing unit [10], which determines whether a certain area in the image is a landslide. The object-based landslide detection method calculates the texture and spectral similarity between the pixels in the remote sensing image, clusters a single pixel into multiple candidate objects, and then sets a threshold to classify each candidate object for landslide classification. Then the acquired suspected landslide areas were classified, and the early rule-based classification systems were established mainly relying on the professional judgment of relevant experts on data features [39]. With the rapid development of technology, machine learning has been widely applied to landslide and other geological hazards research, and many machine learning-based landslide classification algorithms have been proposed one after another, such as Stumpf and Kerle [11] implemented object-based landslide detection with random forest (RF), Van Den Eeckhaut et al.[12], which used an object-based method and support vector machine (SVM) to identify landslides in forested areas with LiDAR and its derivatives, etc. The traditional machine learning-based landslide detection method for remote sensing images can explore large landslide

areas in complex contexts and has the advantages of lower cost and faster than field survey methods, but the accuracy of this method relies heavily on the selection of parameters for the classification of candidate landslide images, i.e., the background knowledge of the landslide domain [36].

Another one is a remote sensing image landslide detection method based on convolutional neural network [9]. With the rapid development of deep learning, convolutional neural networks (CNN) can effectively extract key features from image training samples by the two advantages of local perception and parameter sharing, which has become one of the most important feature extraction methods [13,14,15] for image processing tasks such as image classification, target recognition, etc. A convolutional neural network-based landslide detection method for remote sensing images can automatically extract important features of landslide remote sensing images through a multilayer convolution operation [37], thus avoiding the manual feature design and related parameter setting process that requires landslide expertise to perform, and making the landslide detection task more straightforward and simple. Wang [40] used an integrated geographic database to compare the recognition accuracy of five machine learning methods, namely, convolutional neural network, random forest, logistic regression, reinforcement learning, and support vector machine, in identifying landslides in natural terrain, and among the five methods, convolutional neural network had the highest recognition accuracy, while pointing out that recognition techniques based on machine learning and deep learning have excellent It is also pointed out that the recognition techniques based on machine learning and deep learning have excellent robustness and great potential for problem solving in landslide recognition research. Recently, many scholars have proposed several remote sensing image landslide detection methods based on deep learning for convolutional neural networks. Ding [16] proposed to use the traditional convolutional neural network to extract image features to find suspicious areas where landslides occurred, and then confirm these suspicious areas through change detection methods based on image texture features. Because the traditional convolutional neural network has poor characterization ability for the detection object with multiple scales [10], and landslides usually appear at different scales, with the landslide length from several meters to several kilometers [17]. Therefore, Lei [18] et al. proposed a fully convolutional neural network based on pyramid pooling, which can extract feature semantics in remote sensing images more efficiently, and performs better in multi-scale landslide detection.

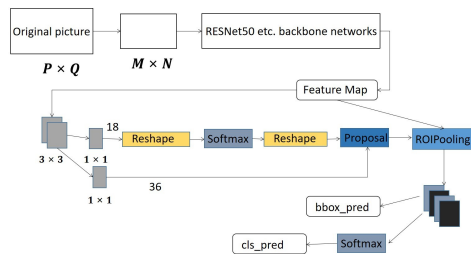
This paper proposes an Efficient Residual Channel Attention Mechanism Network (ERCA). ERCA intends to improve the feature learning ability of the deep learning target detection algorithm by reducing the background noise of images in complex environments through a deep learning adaptive soft thresholding approach to improve the accuracy of landslide identification detection algorithms in scenarios with complex land cover and uncertainty of light and dark intensity of remote sensing images. ERCA is highly portable and can be easily added to mainstream networks such as ResNet, VGG. The ERCA is integrated into the Faster R-CNN model to improve the model's ability to extract landslide features in remote sensing images. Compared with other current algorithms, the improved algorithm has higher AP values. The algorithm in this paper applies to landslide images taken by remote sensing satellites of common resolution, and the landslide images used in the experiments in this paper are taken by TripleSat satellites.

## 2. Landslide detection and identification method

In view of the complex surrounding environment and the many types of landslides, this paper presents an Efficient Residual Channel Attention Mechanism Network (ERCA) in order to improve the effect of landslide target detection, and integrates ERCA into The Faster R-CNN model to propose a more significant landslide characterization from the background to improve the detection effect.

### 2.1. Faster R-CNN network structure

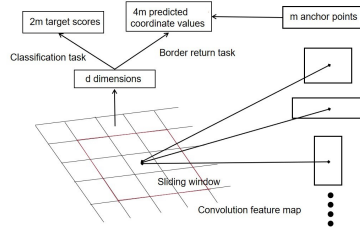
The detection process of the Faster R-CNN algorithm is shown in Figure 1. The process is as follows: 1) perform the feature extraction on the original image through the basic convolutional backbone network (ResNet50[21], VGG16[22], etc.); 2) use the Feature Map extracted in step 1 to generate multiple candidate regions through the RPN network; 3) output a fixed-size feature map through the ROI pooling layer based on the Feature Map extracted in step 1 and the candidate region generated in step 2; 4) classify the categories based on the feature map output in step 3, and perform the frame regression to obtain the precise position of the detection frame. The RPN network is one of the biggest



**Fig. 1.** Faster R-CNN Structure

innovations of the Faster R-CNN algorithm. The previous candidate region extraction methods are usually very time-consuming, such as the SS (Selective Search) algorithm adopted by R-CNN and Fast R-CNN [23] and the Sliding Window algorithm used in traditional target detection. RPN is implemented by a fully convolutional network, which is essentially a classless object detector based on a sliding window. Since RPN can share the convolutional features of the entire image with the detection network, it can output a series of candidate region suggestion frames for input images of any scale at almost no cost.

The structure of the RPN network is shown in Figure 2. First, use the sliding window to generate  $m$  anchor points for the center position of each window on the shared feature map as the initial detection frame, and then perform object classification and border regression on the generated anchor points. Since object classification is a two-classification problem, that is, to determine whether the anchor point is the detection target or the background, the object classification obtains  $2m$  target scores. Similarly, because the bounding



**Fig. 2.** RPN structure

box regression needs to modify the four coordinate values (x, y, w, h), the bounding box regression obtains 4m predicted coordinate values.

It can be seen from the above that the RPN network is a multi-task network, and its overall loss function is composed of two parts. The equation is as follows:

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

The left side of the plus sign (+) in Eq. (1) is the loss value of the classification task, where  $p_i$  represents the probability that the current anchor is the target;  $P_i^*$  represents the target label value, as in Eq. (2), that is, if the current anchor is a positive sample, its value is 1, otherwise it is 0.

$$P_i^*(x) = \begin{cases} 0 & x \in \text{Negativesamples} \\ 1 & x \in \text{Positivesamples} \end{cases} \quad (2)$$

The loss value function used in the classification task is cross entropy, as shown in Eq. (3).

$$L_{cls}(p_i, p_i^*) = -\log [p_i^* p_i + (1 - p_i^*) (1 - p_i)] \quad (3)$$

The right side of the plus sign (+) in Eq. (1) is the loss value of the bounding box regression task, where  $t_i = \{t_x, t_y, t_w, t_h\}$  represents the four predicted coordinate values of the rectangular bounding box;  $t_i^*$  represents the four marker coordinate values in the positive sample; the loss function of the bounding box regression task is only considered when  $p_i^*$  is 1, if  $p_i^*$  is 0, the bounding box regression loss value is also 0, as shown in Eq. (4), in which R is the Smooth L1 function as in Eq. (5).

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \quad (4)$$

$$R(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (5)$$

**2.2. Efficient residual channel attention mechanism network (ERCA)**

This paper is inspired by the literature [24,25,26] to propose the efficient residual channel attention mechanism network (ERCA) which structure is shown in Figure 3. The ERCA

is implemented through the three structures of 1D convolution, soft threshold and residual network.

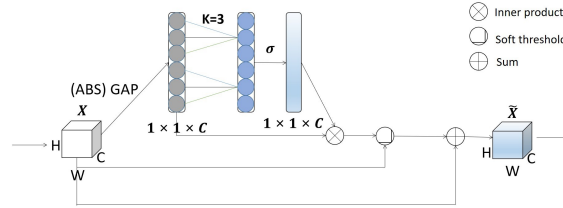


Fig. 3. ERCA Structure

**1D convolution** Given the normalized feature map of the data, global average pooling (GAP) is performed before 1D convolution [27]. Global average pooling adds up all the pixel values of each channel in the feature map to obtain a value, which is used to represent the feature map of this channel. The feature map for n channels is pooled by global averaging to obtain n values then  $X \in R^{H \times W \times C}$  becomes C values. ERCA captures cross-channel attention interaction by considering each channel and its k neighbors. It is realised as a one-dimensional convolution with a k-size convolution kernel captures the attention of neighbors to participate in a channel, in which k represents the coverage of local cross-channel interaction. In order to avoid manual parameter adjustment, the method of reference [24] in this article realizes the automatic learning of k.

$$k = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \tag{6}$$

In Eq. (6), C is the number of channels,  $\gamma, b$  the constants which are set  $\gamma = 1, b = 2$ , and  $\lfloor \cdot \rfloor_{odd}$  the nearest odd. After the global average pooling and 1D convolution, the activation function  $\sigma$  (sigmoid) is used to activate the final output  $V = [v_1, \dots, v_i, \dots, v_c]$ , in which  $v_i$  is a constant.

**Soft threshold** A soft threshold is inserted as a non-linear transformation layer in the deep learning network to eliminate unimportant features:

$$\eta_i(x_i, \lambda_i) = \text{sgn}(x_i) (|x_i| - \lambda_i)_+ \tag{7}$$

In Eq. (7),  $\lambda_i$  represents a non-negative threshold.  $(|x_i| - \lambda_i)_+$  equals  $|x_i| - \lambda_i$  if  $(|x_i| - \lambda_i) > 0$ , while it equals 0 if  $(|x_i| - \lambda_i) < 0$ . The soft threshold  $\lambda_i$  in the paper adopts the similar method in reference[25], which presents  $\lambda_i = v_i \cdot \text{average}_{i,j,c} |x_{i,j,c}|$ .  $\lambda_i$  is the threshold of the feature map of c, where  $i, j, c$  are the width, height and current channel, respectively.

**Residual Network** The final output  $\tilde{X} \in R^{H \times W \times C}$  is obtained after the re-assigned feature map is added to the original one through the residual network, as shown in Eq. (8).

$$\tilde{X} = \eta(X, \lambda) + X = [\eta_1 + x_1, \dots, \eta_i + x_i, \dots, \eta_c + x_c] \quad (8)$$

**ERCA Module for Deep CNN Networks** Figure 3 shows the basic structure of the efficient residual channel attention mechanism network (ERCA). Without dimensionality reduction, the convolutional features are aggregated through the global average pooling operation, and then the cross-channel attention is captured through 1D convolution. The sigmoid function is used to activate the learning channel attention, a soft threshold is inserted as a nonlinear transformation layer to eliminate unimportant features, and finally the residual structure is used for summation. The research adopts the embedding method, replacing the SENet structure with ERCA as embedding ERCA in the CNN network is similar to SENet.

**Parameter Analysis** The process of the SENet model can be simply described as follows: given a feature map  $X \in R^{H \times W \times C}$ , the first step is global average pooling (GAP), the weights in SENet are defined as  $W_1$  and  $W_2$  as in Eq. (9), and the model is adaptively adjusted by a fully connected neural network, and the final output of the model is  $U \in R^{H \times W \times C}$ . The process is as in Eqs. (10-12), where the GAP operation is defined as  $F_{GAP}()$  and  $\sigma$  is the activation function.

$$W_1 = \begin{bmatrix} w_{1,1} & \cdots & w_{1,c} \\ \vdots & \ddots & \vdots \\ w_{c,1} & \cdots & w_{c,c} \end{bmatrix}, W_2 = \begin{bmatrix} w_{1,1} & \cdots & w_{1,c} \\ \vdots & \ddots & \vdots \\ w_{c,1} & \cdots & w_{c,c} \end{bmatrix} \quad (9)$$

$$Avg = F_{GAP}(X) \quad (10)$$

$$T = ReLU(Avg * W_1) \quad (11)$$

$$U = X * \sigma(T * W_2) \quad (12)$$

Inspired by the SENet model, the ECA model works with a similar network structure. However, unlike SENet, the ECA model uses 1D convolution to train and acquire the channel-related features, which greatly reduces the parameters of the model. We define the weights of the ECA model as  $W_3$ , and the final output is  $U \in R^{H \times W \times C}$  after adaptive adjustment by 1D convolution, as in Eq. (14).

$$W_3 = \begin{bmatrix} w_{1,1} & \cdots & w_{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w_{2,2} & \cdots & w_{2,k+1} & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w_{c,c-k+1} & \cdots & w_{c,c} \end{bmatrix} \quad (13)$$

$$U = X * \sigma(Avg * W_3) \quad (14)$$

Define the weight of the ERCA model in this paper as  $W_4$ , and ERCA first performs the GAP operation consistent with SENet and ECA models, and the output after 1D convolutional adaptive adjustment is used as part of the soft threshold, and the final output is

$U \in R^{H \times W \times C}$ , as in Eqs. (16-17).

$$W_4 = \begin{bmatrix} w_{1,1} & \cdots & w_{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w_{2,2} & \cdots & w_{2,k+1} & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w_{c,c-k+1} & \cdots & w_{c,c} \end{bmatrix} \quad (15)$$

$$\lambda = \sigma (Avg * W_4) * F_{GAP}(|X|) \quad (16)$$

$$U = X + \eta(X, \lambda) \quad (17)$$

It is not difficult to find that the model weights  $W_4$  of our algorithm and the weights  $W_3$  of the ECA model are sparser than the weights  $W_1$  and  $W_2$  of the SENet model. We define the number of channels as  $C$ , then the parameters of SENet model weights are  $C^2$ , in contrast, our ERCA model is consistent with ECA model with only  $k$  parameters.

**ERCA Utilizes Both Maximum Pooling Outputs and Average Pooling** In this section, the single average pooling operation in ERCA in the previous section is replaced using the maximum pooling output and average pooling output of the shared network. The ERCA above is defined as the standard type and the ERCA using both maximum pooling output and average pooling is defined as ERCAMA. Its specific structure is shown in Figure 4. It is worth noting that although ERCAMA uses maximum pooling output and average

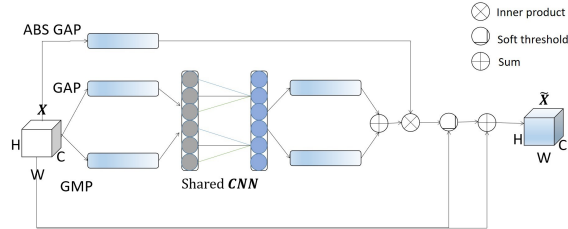


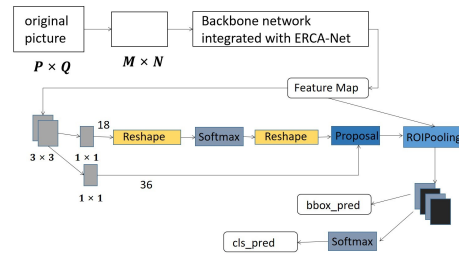
Fig. 4. ERCA Structure

pooling parallel output, the pooling results are convolved using the same set of parameters, so the parametric quantities of ERCAMA and ERCA are the same.

**2.3. The Final Structure of the Model**

Due to the vast and complex self-made landforms, regional differences, and diverse topography and climate in China, the method of investigating potential landslide areas through professionals is time-consuming, expensive, and inefficient. Meanwhile, with the landslides diverse and complex, it is very important to use artificial intelligence to quickly and accurately extract landslide information from satellite image data. The final detection network model of this research is shown in Figure 5, which mainly includes three parts:





**Fig. 5.** final model

(1) Feature extraction: Feature Map is obtained by extracting image features through the backbone network integrated into ERCA. Among them, the commonly used backbone networks are ResNet, VGG, etc. ResNet34, ResNet50 and ResNet101 are adopted as the backbone network in the experiments of this research. (2) RPN detection: The recommended target candidate area is obtained through the RPN detection network processing. (3) Object detection and classification: The target classification result is obtained by extracting and processing the Feature Map of the candidate area. Compared with the original Faster R-CNN, the improved algorithm in this paper realizes the channel attention mechanism by adding few parameters, and improves the target detection effect in complex environments.

### 3. Experimental Process and Analysis

#### 3.1. Experimental Environment and Data Enhancement

This paper adopts python3.7 as the development language, pytorch as the deep learning framework, and Pycharm as the development tool. Graphics card GeForce RTX 2080 Ti is employed with 11G video memory. In order to verify the effectiveness of the algorithm in this paper, we use landslide image in Bijie City[6]. In this experiment, more than 200 high-quality landslide images containing large, medium and small landslides were selected from the landslide images and labeled using the Labeling tool to form a training dataset in PASCAL VOC format<sup>4</sup>. Considering the phenomenon of model overfitting due to deeper network layers and smaller data volume in deep learning and in order to improve the accuracy of the landslide identification detection algorithm under the scenario of complexity of land cover and uncertainty of light and dark intensity of remote sensing images, we expand the dataset by code with data enhancement of the labeled images. The original data volume was expanded by 10 times, i.e. more than 2000 landslide images, by transforming the images left and right, flipping up and down, optical transformation, Gaussian blur, affine transformation and bounding box transformation. The expanded data set is divided into the training set and the test set in an 8:2 manner, and the training set and the verification set are divided in a 9:1 manner in the training set. Adam is adopted in the optimization algorithm. The first 20 epochs are normal training after freezing training,

<sup>4</sup> <http://host.robots.ox.ac.uk/pascal/VOC/>

with the freezing initial learning rate of 0.0002 and the normal training initial learning rate of 0.00002. After starting training, the learning rate attenuation strategy is adopted, with the attenuation coefficient of 0.94.

### 3.2. Comparison of Objective Indicators of Detection Effect

The objective comparison indicators selected in this paper is AP (Average Precision) values, which is defined as follows:

$$P = TP / (TP + FP) \quad (18)$$

$$R = TP / (TP + FN) \quad (19)$$

$$AP = \int_0^1 P(R) dR \quad (20)$$

Among them, TP is the number of positive images with correct predictions, FP is the number of positive images with incorrect predictions, FN is the number of negative images with incorrect predictions, P is the precision rate, and R is the recall rate. APs for different IOU thresholds and APs of different sizes for detecting target objects are specifically defined in the Figure 6.

Average Precision(AP)	
$AP(\%)$	AP at IOU= .50 : .05 : .95 (primary challenge metric)
$AP_{50}$	AP at IOU= .50 (PASCAL VOC metric)
$AP_{75}$	AP at IOU= .75 (strict metric)
AP Across Scales:	
$AP_S$	AP for small objects: area < 32 <sup>2</sup>
$AP_M$	AP for medium objects: 32 <sup>2</sup> < area < 96 <sup>2</sup>
$AP_L$	AP for large objects: area > 96 <sup>2</sup>

**Fig. 6.** The above 6 metrics are used for charaterzing for the performance of obejct detecor

**Ablation Experiment of Faster R-CNN** In this chapter, the different components of the model in this article based on the improvement of the original Faster R-CNN are disassembled for ablation experiments. The Faster R-CNN with ResNet34, ResNet50, and ResNet101 as the backbone network is used as the Baseline and compared with the Faster R-CNN with the backbone network of ResNet34+ERCA, ResNet50+ERCA, and ResNet101+ERCA. The objective comparison indexes of the experimental results are shown in Table 1.

It can be drawn from Table 1 that the  $AP_M, AP_L, AP_{75}$  and final  $AP$  values of the Faster R-CNN using ResNet34+ERCA, ResNet50+ERCA and ResNet101++ERCA as the backbone network have been improved to different degrees compared to the Faster

**Table 1.** Ablation experiment

Model	AP(%)	AP50	AP75	APS	APM	APL	Params
ResNet34	0.656	0.976	0.827	0.614	0.641	0.678	82.47M
ResNet34+ERCA	0.684	0.988	0.847	0.619	0.667	0.709	82.48M
ResNet34+ERCAMA	0.697	0.988	0.868	0.630	0.693	0.715	82.48M
ResNet50	0.716	0.988	0.881	0.647	0.680	0.764	108.12M
ResNet50+ERCA	0.728	0.988	0.925	0.655	0.703	0.766	108.13M
ResNet50+ERCAMA	0.750	0.988	0.928	0.712	0.729	0.783	108.13M
ResNet101	0.725	0.988	0.903	0.645	0.708	0.758	180.83M
ResNet101+ERCA	0.740	0.988	0.933	0.733	0.724	0.767	180.84M
ResNet101+ERCAMA	0.764	0.985	0.940	0.763	0.749	0.788	180.84M

R-CNN using the original ResNet34, ResNet50 and ResNet101 as the backbone network, and there is only a small increase in the model parameters. The algorithm incorporated into ERCA proved to be improved compared to Baseline, and the ERCA in this paper can improve the original network.

**Comparison of Different Channel Attention Models** In order to verify the effectiveness of the efficient residual channel attention mechanism network (ERCA) proposed in this paper, three different channel attention mechanisms SENet, ECA, ERCA,ERCAMA are used in the backbone network combined with ResNet34, ResNet50, ResNet101 as the backbone network for comparison in this section. ECA uses the Github source code <https://github.com/BangguWu/ECANet> disclosed by the original author, and SENet uses <https://github.com/moskomule/senet.pytorch>, the warehouse with the highest number of stars in the pytorch version on Github. The objective comparison indexes of the experimental results are shown in Table 2.

**Table 2.** Ablation experiment

Model	AP(%)	AP50	AP75	APS	APM	APL	Params
ResNet34+ECA	0.663	0.987	0.817	0.542	0.645	0.697	82.48M
ResNet34+SENet	0.661	0.987	0.807	0.651	0.641	0.683	85.67M
ResNet34+ERCA	0.684	0.988	0.847	0.619	0.667	0.709	82.48M
ResNet34+ERCAMA	0.697	0.988	0.868	0.630	0.693	0.715	82.48M
ResNet50+ECA	0.719	0.987	0.879	0.613	0.684	0.765	108.13M
ResNet50+SENet	0.723	0.988	0.913	0.709	0.692	0.765	159.24M
ResNet50+ERCA	0.728	0.988	0.925	0.655	0.703	0.766	108.13M
ResNet50+ERCAMA	0.750	0.988	0.928	0.712	0.729	0.783	108.13M
ResNet101+ECA	0.731	0.988	0.900	0.660	0.717	0.755	180.84M
ResNet101+SENet	0.730	0.988	0.905	0.695	0.710	0.760	277.25M
ResNet101+ERCA	0.740	0.988	0.933	0.733	0.724	0.767	180.84M
ResNet101+ERCAMA	0.764	0.985	0.940	0.763	0.749	0.788	180.84M

It can be seen from Table 2 that the algorithm ERCA in this paper is basically the same as compared to the ECA model parameters, which are greatly reduced compared

to SENet. The performance results of the three different attention mechanism networks on Faster R-CNN with ResNet34 and ResNet50 as the backbone networks show that the ERCA model outperforms the ECA model with the same parameters in all metrics, and the  $AP_S$  of ERCA is slightly lower than SENet, but the rest of the metrics are higher or the same than SENet, indicating that ERCA has stronger robustness compared to SENet in detecting objects containing different sizes. The three attention mechanism networks have equal  $AP_{50}$  on the Faster R-CNN with ResNet101 as the backbone network, and the remaining metrics ERCA model outperforms the ECA model and SE model. The efficient residual channel attention mechanism in this paper is experimentally proven to have good results.

**Comparison with other Target Detection Algorithms** In order to quantitatively analyze the detection performance of the Faster-RCNN algorithm after adding ERCA, three classical target detection networks, Faster-RCNN, YOLOv3[28] and YOLOv4[29], are selected for experimental comparison with the algorithm in this paper. From Table 3, we can see that the Faster R-CNN with ECRA is better than YOLOV3 and YOLOV4 except for the AP50 index, which is slightly lower than YOLOV4.

**Table 3.** Ablation experiment

Model	AP(%)	AP50	AP75	APS	APM	APL
Faster R-CNN	0.725	0.988	0.903	0.645	0.708	0.758
YOLOV3	0.653	0.978	0.811	0.659	0.645	0.666
YOLOV4	0.667	0.990	0.856	0.620	0.660	0.678
Faster R-CNN+ERCA	0.740	0.988	0.933	0.733	0.724	0.767
FasterR-CNN +ERCAMA	0.764	0.985	0.940	0.763	0.749	0.788













### 3.3. Display of the Subjective Effect Detection

Table 4 shows the subjective detection results of ResNet50+SENet, ResNet50+ECA, ResNet50+ERCA and ResNet50+ERCAMA as the backbone of Faster R-CNN. For the image of a single landslide as Picture 1, the detection results of the three algorithms are basically the same. For images of multiple landslides with complex backgrounds as in Picture 2, all three algorithms show different degrees of misses and misjudgments, among which ResNet50+SENet and ResNet50+ECA show both misses and misjudgments, while ResNet50+ERCA and ResNet50+ERCAMA only shows misses. For images with multiple landslides and a single background as in Picture 3, the detection results of the three algorithms are basically the same with no misses or misjudgments.


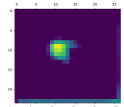
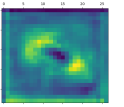
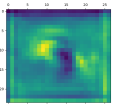
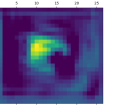
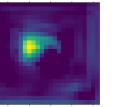

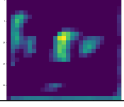
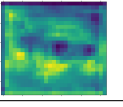
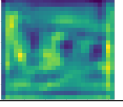
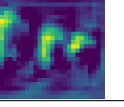
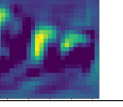
### 3.4. Feature Map Visualization

Table 5 shows the results of the detection using the feature map visualization to verify the validity of the detection results, where the warmer color indicates the higher attention

**Table 4.** Three model results diagram.

Model	Picture 1	Picture 2	Picture 3
ResNet50+SENet			
ResNet50+ECA			
ResNet50+ERCA			
ResNet50+ERCAMA			

**Table 5.** Feature map visualization

Original image	ResNet101	ResNet101 +ECA	ResNet101 +SENet	ResNet101 +ERCA	ResNet101 +ERCAMA
					
					

of the deep learning network. From the results in Table 5, it can be seen that the highlighted areas can cover part of the landslide area when no attention is added. And the highlighted areas increase significantly after adding the channel attention mechanisms ECA and SENet respectively, but they also cover many areas that are not landslides, which enhances the landslide features and also enhances part of the redundant features. After adding ERCA and ERCAMA attention mechanisms on the backbone network, the highlighted areas can cover the landslide areas more accurately and comprehensively.

### 3.5. Application Effect Display

The display site in Yingxiu Town, Wenchuan, in the Sichuan Province, Southwest of China, is selected to present its application. The 18-layer satellite image slices of Yingxiu town were downloaded by bigemap software, and then batch tested based on the program. Some of the test results are as follows.

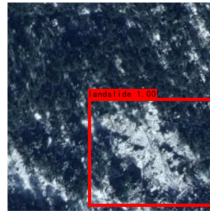


**Fig. 7.** Landslide



**Fig. 8.** Landslide

Figures 7 and 8 show better detection results, but there Figures 9 are still some problems worthy of improvement in actual detection. For example, Figure 8 mistakenly identifies snow mountains as landslides, Figure 10 mistakenly identifies open spaces of human buildings as landslides.



**Fig. 9.** Snow Mountain



**Fig. 10.** Open space

In the follow-up practice, more landslide data will be marked, and various terrains and buildings that are easily judged as landslides will be distinguished.

#### 4. Conclusion

This paper uses deep learning Faster R-CNN network for landslide detection research, for the diversity and complexity of landslides this paper proposes an efficient residual channel attention mechanism network (ERCA), ERCA has high portability and can be incorporated into Resnet34, Resnet50, Resnet101 and other networks. It is incorporated with Resnet34, Resnet50, Resnet101, etc. as the backbone network of Faster R-CNN, and the objective index and subjective detection of the algorithm proposed in this paper have achieved good results in landslide image detection experiments, and the experiments prove that the algorithm proposed in this paper has some practicality. The landslide detection model in this paper mainly focuses on different landslide detection in different complex environments. In future work we will use the model in this paper as a sub-model combined with other environmental data such as rainfall, geological conditions, earthquake levels, etc. to build a more comprehensive landslide detection model.

**Acknowledgments.** This research was funded by the Basic Research Project of Department of Science and Technology of Sichuan Province (NO. 2021YJ0335).

#### References

1. Roback, K., Clark, M. K., West, A. J., Zekkos, D., Li, G., Gallen, S. F., Godt, J. W.: The size, distribution, and mobility of landslides caused by the 2015 Mw7.8 Gorkha earthquake, Nepal. *Geomorphology*, Vol. 301, 121–138. (2018)

2. Parker, R. N., Densmore, A. L., Rosser, N. J., De Michele, M., Li, Y., Huang, R., Petley, D. N.: Mass wasting triggered by the 2008 Wenchuan earthquake is greater than orogenic growth. *Nature Geoscience*, Vol. 4, No. 7, 449–452. (2011)
3. Mondini, A. C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M., Ardizzone, F.: Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sensing of environment*, Vol. 115, No. 7, 1743–1757. (2011)
4. Hong, Y., Adler, R. F., Huffman, G.: An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 6, 1671–1680. (2007)
5. Ouyang, C., Zhou, K., Xu, Q., Yin, J., Peng, D., Wang, D., Li, W.: Dynamic analysis and numerical modeling of the 2015 catastrophic landslide of the construction waste landfill at Guangming, Shenzhen, China. *Landslides*, Vol. 14, No. 2, 705–718. (2017)
6. Ji, S., Yu, D., Shen, C., Li, W., Xu, Q.: Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides*, Vol. 17, 1337–1352. (2020)
7. Shi, W., Zhang, M., Ke, H., Fang, X., Zhan, Z., Chen, S.: Landslide Recognition by Deep Convolutional Neural Network and Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 6, 2020, 4654–4672. (2020)
8. Yu, B., Chen, F., Xu, C.: Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. *Computers and Geosciences*, Vol. 135, 104388–104388. (2015)
9. Zhang, L., Zhang, L., Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote Sensing magazine*, Vol. 4, No. 2, 22–40. (2016)
10. Sameen, M. I., Pradhan, B.: Landslide Detection Using Residual Networks and the Fusion of Spectral and Topographic Information. *IEEE Access*, Vol. 7, 114363–114363. (2019)
11. Stumpf, A., Kerle, N.: Object-oriented mapping of landslides using Random Forests. *Remote sensing of environment*, Vol. 115, No. 10, 2564–2577. (2011)
12. Van Den Eeckhaut, M., Kerle, N., Poesen, J., Hervás, J.: Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. *Geomorphology*, Vol. 173, 30–42. (2012)
13. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, Vol. 60, No. 6, 84–90. (2017)
14. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision pattern recognition*, 3431–3440. (2015)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision pattern recognition*, 770–778. (2016)
16. Ding, A., Zhang, Q., Zhou, X., Dai, B.: Automatic recognition of landslide based on CNN and texture change detection. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 444–448. (2016)
17. Zhang, Yunling., Fu, Yuhao., Sun, Yu., Zeng, Doudou., Xu, Zeran., Wu, Hangbin.: Combining deep neural networks for landslide detection highway with high-resolution remote sensing images, 188–194. (2021)
18. Lei, T., Zhang, Y., Lv, Z., Li, S., Liu, S., Nandi, A. K.: Landslide inventory mapping from bitemporal images using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, Vol. 16, No. 6, 1–5. (2019)
19. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE conference on computer vision pattern recognition*, 2117–2125. (2017)
20. Singh, B., Najibi, M., Davis, L. S.: SNIPER: Efficient Multi-Scale Training. *Advances in neural information processing systems*, 9310–9320. (2018)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision pattern recognition*, 770–778. (2016)



22. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science*. (2014)
23. Uijlings, J. R., Van De Sande, K. E., Gevers, T., Smeulders, A. W.: Selective Search for Object Recognition. *International Journal of Computer Vision*, Voc. 104, No. 2, 154-171. (2013)
24. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11534-11542. (2020)
25. Zhao, M., Zhong, S., Fu, X., Tang, B., Pecht, M.: Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Transactions on Industrial Informatics*, Voc. 16, No. 7, 4681-4690. (2019)
26. Hu, J., Shen, L., Sun, G.: Squeeze-and-Excitation Networks. In *Proceedings of the IEEE conference on computer vision pattern recognition*, 7132-7141. (2018)
27. Lin, M., Chen, Q., Yan, S.: Network In Network. *Computer Science*. (2013)
28. Redmon, J., Farhadi, A.: YOLOv3: An Incremental Improvement[J]. *arXiv e-prints*. (2018)
29. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv e-prints*. (2020)
30. Mirghasemi, S., Lotfizad, M.: A target-based color space for sea target detection. *Applied Intelligence*, Voc. 36, No. 4, 960-978. (2012)
31. Li, X., Du, Z., Huang, Y., Tan, Z.: A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, Voc. 179, 14-34. (2021)
32. Gao, S., Guan, H., Ma, X.: A recognition method of multispectral images of soybean canopies based on neural network, Vol. 68, 101538-101538. (2021)
33. Masouleh, M. K., Shah-Hosseini, R.: Development and evaluation of a deep learning model for real-time ground vehicle semantic segmentation from UAV-based thermal infrared imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, Voc. 155, 172-186. (2019)
34. Xu, D., Zhang, N., Zhang, Y., Li, Z., Zhao, Z., Wang, Y.: Multi-scale unsupervised network for infrared and visible image fusion based on joint attention mechanism. *Infrared Physics and Technology*, Voc. 125, 104242-104242. (2022)
35. Amatya, P., Kirschbaum, D., Stanley, T.: Use of Very High-Resolution Optical Data for Landslide Mapping and Susceptibility Analysis along the Karnali Highway, Nepal. *Remote Sensing*, Voc. 11, No. 19, 2284-2284. (2019)
36. Yu, B., Xu, C., Chen, F., Wang, N., Wang, L.: HADeenNet: A hierarchical-attention multi-scale deconvolution network for landslide detection. *International Journal of Applied Earth Observation and Geoinformation*, Voc. 111, 102853-102853. (2022)
37. Zeng, Q., Geng, J.: Task-specific contrastive learning for few-shot remote sensing image scene classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, Voc. 191, 143-154. (2022)
38. Wang, H., Zhang, L., Wang, L., Fan, R., Zhou, S., Qiang, Y., Peng, M.: Machine learning powered high-resolution co-seismic landslide detection. *Gondwana Research*. (2022)
39. Barlow, J., Martin, Y., Franklin, S. E.: Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Canadian journal of remote sensing*, Voc. 29, No. 4, 510-517. (2003)
40. Wang, H., Zhang, L., Yin, K., Luo, H., Li, J.: Landslide identification using machine learning. *Geoscience Frontiers*, Voc. 12, No. 1, 351 - 364. (2021)

**Yabing Jin** is a professor who works for the Shenzhen Geological Bureau and an expert in geological hazard-related research.

**Ou Ou** received the Ph.D. degree from Chengdu University of Technology, Chengdu, China, in 2015. He is a professor in the School of Computer and Network Security of

Chengdu University of Technology. His current research interests are artificial intelligence and big data in geology.

**Shanwen Wang** received the B.S. degree in Software Engineering from the School of Information Science and Technology, Chengdu University, Chengdu, China, in 2020. He is pursuing the M.S. degree at Chengdu University of Technology. His current research interests are deep learning and computer vision.

**Yijun Liu** works at the Shenzhen Geological Bureau. His research interests are the area of geological hazard monitoring and prevention.

**Haoqing Niu** received the B.S. degree in Internet of Things Engineering from the School of Information Engineering, DaLian University, DaLian, China, in 2020. He is pursuing the M.S. degree at Chengdu University of Technology. His current research interests are small target detection.

**Xiaopeng Leng** received the Ph.D. degree from Chengdu University of Technology. He is an associate professor in the School of Computer and Network Security at Chengdu University of Technology. His research interests are directed towards the application of artificial intelligence in geology.

*Received: August 31, 2022; Accepted: December 25, 2022.*