WILEY | Hindawi

Research Article

A Mountain Summit Recognition Method Based on Improved Faster R-CNN

Yueping Kong [ID]**, Yun Wang** [ID]**, Song Guo** [ID]**, and Jiajing Wang**

School of Information and Control, Xi'an University of Architecture and Technology, Xi'an 710055, China

Correspondence should be addressed to Yueping Kong; kongyp@xauat.edu.cn

Mountain summits are vital topographic feature points, which are essential for understanding landform processes and their impacts on the environment and ecosystem. Traditional summit detection methods operate on handcrafted features extracted from digital elevation model (DEM) data and apply parametric detection algorithms to locate mountain summits. However, these methods may no longer be effective to achieve desirable recognition results in small summits and suffer from the objective criterion lacking problem. Thus, to address these problems, we propose an improved Faster region-convolutional neural network (R-CNN) to accurately detect the mountain summits from DEM data. Based on Faster R-CNN, the improved network adopts a residual convolution block to replace the traditional part and adds a feature pyramid network (FPN) to fuse the features with adjacent layers to better address the mountain summit detection task. The residual convolution is employed to capture the deep correlation between visual and physical morphological features. The FPN is utilized to integrate the location and semantic information in the extracted feature maps to effectively represent the mountain summit area. The experimental results demonstrate that the proposed network could achieve the highest recall and precision without manually designed summit features and accurately identify small summits.

1. Introduction

Mountain summits are essential topographic feature points that are widely utilized in military and nonmilitary domains, such as biodiversity assessment [1], landslide risk analysis [2], and glacier and snow-covered [3] summit analysis. The summit is the area with the maximum elevation from sea level area. Summits are usually located in a complex and giant topographic system, with complex structural and functional differences [4]. Automating the summit detection process will greatly advance and enrich our geospatial knowledge; thus, it is valuable to study effective methods for the automatic detection of mountain summits.

In the literature, there are two main streams of methodologies, heuristic-based methods and data-driven methods, which have been extensively discussed in summit detections. A common trait of the heuristic-based methods is that they rely on features selected by the algorithm designer and landform recognition rules that depend on parameters configured by the user. In a prior work [5], a fuzzy set theory was applied to terrain analysis, which computes the fuzzy membership of each digital elevation model (DEM) pixel to six different morphometric classes, Pass, Pit, Plane, Ridge, Channel, and Peak, which are obtained through the evaluation at multiple scales. In another study [6], a multiscale and multisemantic method, which combines landform attributes and the surrounding environment to compute the membership value of each grid around the mountain summit, was proposed to detect mountain summits. The author considers the mountain to be a fuzzy entity with various attributes, such as topographic relief, average slope, and relative altitude. In another work [7], an accurate summit detection method based on morphological analysis was presented to detect summits, in which the author concluded that the summit should be located in a nonflat area and should be the highest point with respect to the eight adjacent grids around it. Moreover, the author further illustrated that different summits should be

separated from a certain level horizontal and vertical distance. Thus, a 3×3 sliding window is applied on the DEM to find the highest point in a local area and regard it as a summit candidate. Then, the relative distance between neighbors of the candidate is analyzed to more accurately locate the summit. Although the methods above can well address the false detection and missing detection problem of mountain summit detection tasks, most of these methods require manually designed features, and, thus, their representation abilities are limited. Furthermore, it is nontrivial and cumbersome to manually select parameters especially when multiple parameters are involved.

With the rapid development of remote sensing technology, high-resolution DEM data have become easily accessible, which is characterized by complex backgrounds, diverse feature structures, and rich details. Easily accessible high-resolution DEM data have provided an incredible opportunity to study summit detection from a data-driven perspective. The reported data-driven methods can be classified into two categories: machine learning methods and deep learning (DL) methods. Recent surveys of the applications of DL in remote sensing can be found in areas such as scene classification [8], object detection [9, 10], land use, and land cover analysis [11]. In a prior work [12], 446 recorded landslides and landslide-related conditioning factors were acquired, stored, and analyzed through remote sensing and geographic information system technologies. Then, the landslide susceptibility of Ningdu County was predicted using supervised machine learning models (support vector machine and chi-squared automatic interaction detection models) and unsupervised machine learning models (K-means and Kohonen models) based on 11 conditioning factors. In another study [13], three machine learning models, boosted regression tree (BRT), classification and regression tree (CART), and random forest (RF), were compared to produce groundwater spring potential maps.

Recently, DL has been widely used in various computer vision applications, where it can automatically conduct feature selection from data samples. Convolutional neural networks (CNNs) have been successfully used to perform object detection and image recognition [14–16], and CNNs contain a series of mathematical operations, such as convolution, pooling, and thresholding, to automatically learn the target features from low-level semantics. Due to its strong ability to capture the spatial correlation and the more advanced mechanism of feature extraction, CNNs, as well as DL technology, are hot topics in geography. In a prior work [17], a DL method was developed to detect terrain features, including craters, which combines the Faster region CNN (R-CNN) model with a ZF-net architecture to recognize some common cases, such as multiple separated but very close craters and very small craters. In another study [18], a DL approach was proposed for automatic terrain feature identification from remote sensing images, which extends the Faster-RCNN architecture with deep CNNs and adopts ensemble learning to detect nine different types of terrain features. Torres et al. [19] proposed an automatic summit recognition method based on DL. This method regards the summit recognition task as a classification problem and

performs well compared to traditional methods. However, the sliding window makes the network only focus on local features, which ignores the summit's overall shape and spatial structure. With appropriately selected network structures, DL methods provide flexible options for better addressing various scenarios of terrain feature identification. However, the study of DL methods in terrain feature identification is still in its infancy, and further exploration is needed to discover its full potential.

In this paper, we focus on how to apply a DL model to summit detection to achieve high accuracy without manually designed features. A mountain summit recognition approach based on the Faster R-CNN framework is proposed for more effective mountain summit detection. The proposed approach borrows ideas from residual convolution [20] and feature pyramid network (FPN) [21] to automatically extract the feature of the mountain summit and directly output the summit's location in an end-to-end manner without setting parameters.

The main contributions of this paper can be summarized as follows: (1) We formalize summit detection as an image processing task to train the DL model with DEM data and locate the boundary coordinates of the summit. (2) We propose an advanced method for identifying summits from DEM data, which uses the residual structure to improve the convolutional layer and merges features of different levels. (3) We created a new summit detection data set, including its location and boundaries, to build the proposed model. (4) A computational experiment demonstrated that the proposed method could outperform the benchmarks, especially in terms of detecting small mountain tops and pseudosummits.

2. Methodology

According to the spatial, scale, and controlling area characteristics, the mountain summit can be classified into several types. Each type of summit has distinctive geological properties that are not easy to represent in a single model. Faster R-CNN is the most representative CNN for object detection. The region proposal network (RPN) is presented for efficient and accurate region proposal generation. It is possible to use a very deep network to improve the overall object detection accuracy by sharing convolutional features with a downstream detection network. However, some limitations of Faster R-CNN, such as poor feature extraction ability and inefficient feature utilization mechanism, result in the tendency to miss small summits. Therefore, the hierarchical structure of FPN is applied in the proposed method to integrate features of different scales to more accurately locate and identify summits. The overall improved Faster R-CNN is illustrated in the schematic diagram shown in Figure 1.

As shown in Figure 1, the improved Faster R-CNN consists of four parts components: a feature extractor, an FPN, an RPN, and a summit classifier. The DEM is fed to the feature extractor to shrink its size and increase the number of channels through a sequence of stacked convolution layers. The output of the feature extractor is a series of feature maps that represent the summit from different
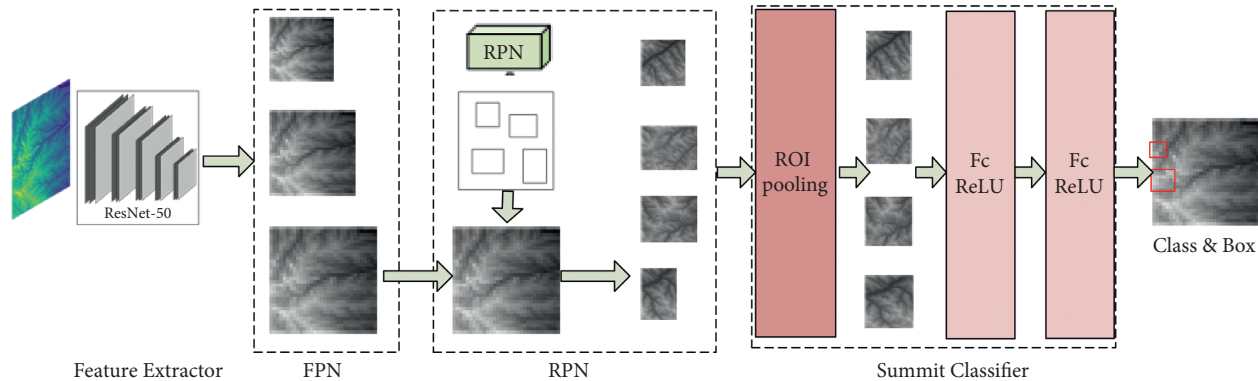
FIGURE 1: The framework of improved faster R-CNN.

perspectives learned from the data. Next, these feature maps are sent to the FPN to generate several fused feature maps containing the summit's semantic information and location information. Finally, the RPN generates the location of the summit through the fused feature maps. Meanwhile, the parts of the fused feature maps are sent to the summit classifier to determine whether it is a summit.

The performance of a neural network increases with the depth of the network layers. However, neural network models with many layers are subject to problems during training, including gradient vanishing and gradient exploding. The ResNet model effectively addresses these problems by introducing a deep residual framework. In this work, we evaluated the performance of the ResNet-50 and ResNet-101 architectures. ResNet-101 achieved only a 0.1% accuracy improvement, while the computational cost increased significantly. This is because the summits in the DEM are relatively small and their features may no longer be identified in those deeper network levels. We chose ResNet-50 as the feature extractor in the improved Faster R-CNN framework to balance accuracy and computational complexity.

Faster R-CNN only uses RPN to perform region suggestion operations in the last convolutional layer, while the semantic information displayed by small targets in the high-level features is very limited. It is not easy to obtain more comprehensive information to predict the summit location. Figure 2 shows the feature maps extracted by ResNet-50. We can see that shallow features identify edges by comparing the brightness of adjacent pixels, while deeper features can find a specific set of contours and corners to detect the entire part of the summit and finally identify the summit in the image. However, as the number of layers in the network increases, the semantic information in feature maps becomes increasingly prominent, and the location information is gradually blurred. To find all the possible summit-like regions for subsequent inferring, the FPN structure fuses semantic and location information of the mountain summit so that the features at each scale have wealthy semantic information. The improved structure is depicted in Figure 3.

We combine the feature from pyramid levels 4 (P4), 3 (P3), and 2 (P2) to generate the finest feature map. Since the summit is so small that it cannot be retained at this level, the

output from the fifth convolutional layer (C5) is excluded for proposal detection. Afterward, the feature maps P2, P3, and P4 are used as an input of the RPN. Based on the location regression layer of the RPN, a regional suggestion box is generated to determine the possible locations of the mountain summit, and the classification layer of the RPN determines the probability of the existence of the summit area in the box. In Faster R-CNN, three anchor boxes of different scales and aspect ratios are predefined manually according to the PASCAL VOC data set. These anchor boxes are used as the reference bounding boxes for the algorithm to predict the target position for the first time. It should be noted that, if we use the default anchor box, the convergence speed of the bounding box regression slows down during the training process of Faster R-CNN. Moreover, once an error occurs in the RPN, it is difficult for the summit classifier to correct because they share some features between them. Therefore, considering the size of the summit areas in the SUMMIT-DEM data set, k-means clustering [22] is used to adjust the size of the anchor box in the proposed network. Table 1 shows the anchor box information of the proposed and previous methods.

3. Experimental Data

Deep learning is much more potent than traditional approaches due to its ability to learn high-level and abstract features from data. Therefore, a large amount of data is needed. Although many large databases, such as VOC [23] and COCO [24], are available in object detection, few publicly available data sets for terrain elements detection are based on optical images.

We use the DEM data marked by NASA [25] to build the samples data set of summits area, named SUMMIT-DEM. The DEM avoids the influence of the illumination and viewing angle on experimental results, but it lacks many details representing the summit areas, such as morphology, orientation, and contrast. We render DEM data into different modes through different visualization technologies to enable the network to represent mountain summit area features better.

Firstly, different elevation values are assigned to different gray scales to achieve three-dimensional terrain expression
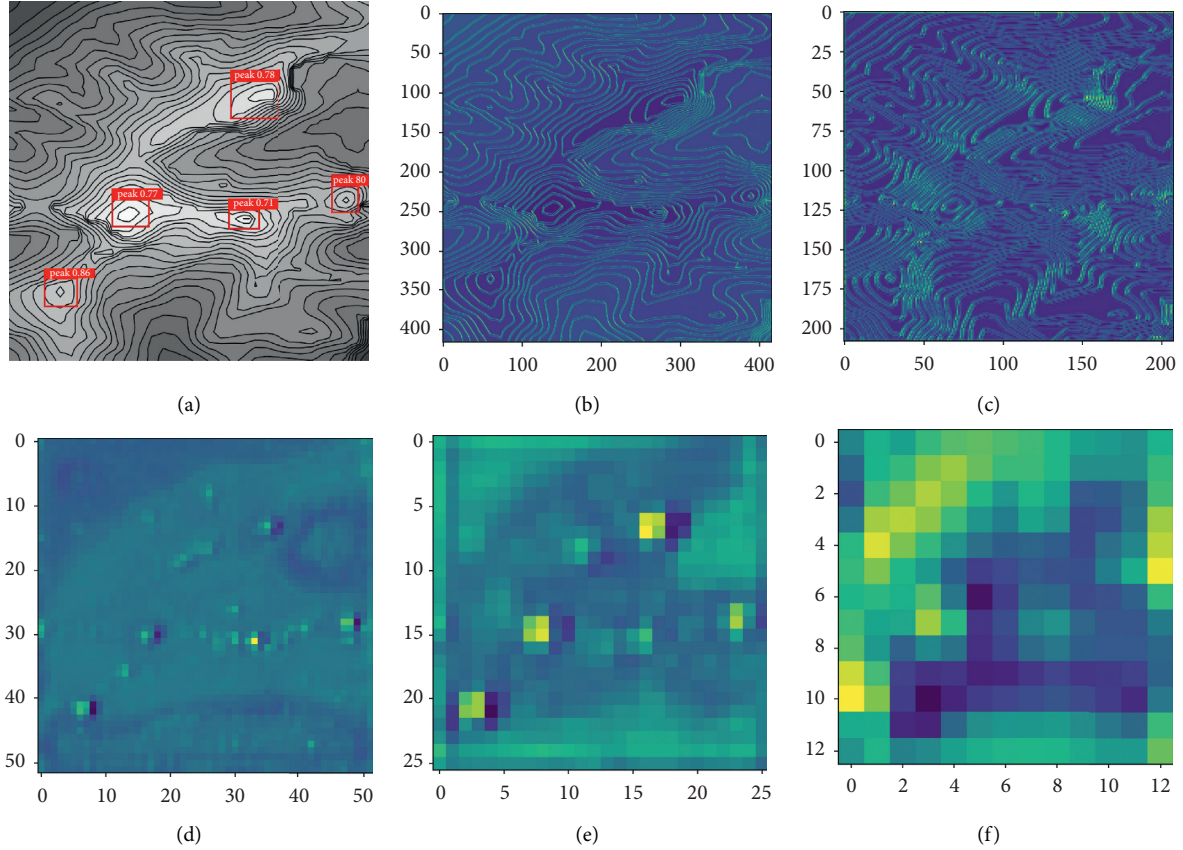
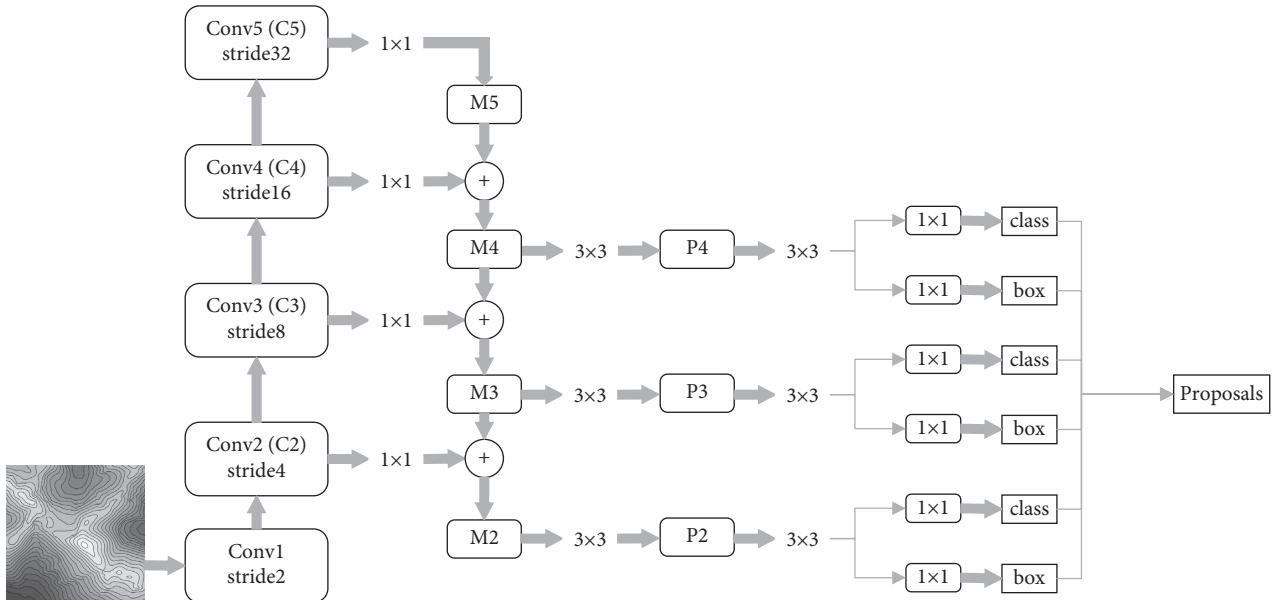Figure 2: Feature maps from the shallow to deep layers of the input image.

Figure 3: Feature pyramid module structure of the improved faster R-CNN.

on a two-dimensional plane through tonal differences. The range of elevation value is $[H_{\min}, H_{\max}]$, and the corresponding gray range is $[G_{\min}, G_{\max}]$. Then, for any elevation, the corresponding gray value G_i can be calculated through equation (1). After that, the gray value is normalized to

between 0 and 1, which can reduce part of the noise in the input data without changing the relative elevation between elements. The converted image is shown in Figure 4(a). Then, the contour lines are generated by a serial of elevation intervals from the DEM, which can scientifically reflect the

TABLE 1: Anchor information of the proposed RPN and the original RPN.

Method	Feature	Size	Ratio
Original RPN*	C5	128, 256, 512	0.5, 1, 2
Our RPN	P2	42	0.66, 0.58, 0.59
	P3	102	0.66, 0.58, 0.59
	P4	302	0.66, 0.58, 0.59

*The original RPN represents the RPNs of faster R-CNN.

primary geomorphological forms and changes such as ground elevation, mountain body, slope, slope shape, and mountain strike, as shown in Figure 4(b). Finally, to satisfy the input of the network, two kinds of data are superimposed to form the sample shown in Figure 4(c). Figure 4(d) shows a sample of the summit area after annotation.

By this way, we have made the sampling data set SUMMIT-DEM, which includes a total of 1000 images and 3,345 samples of the mountain summit area. The data set was divided into training, verification, and testing set with a ratio of $7:2:1$.

$$G_i = G_{\min} + \frac{G_{\max} - G_{\min}}{H_{\max} - H_{\min}} \times (H_i - H_{\min}). \tag{1}$$

4. Experiments and Discussion

Two experiments were conducted to evaluate the advantages of the improved Faster R-CNN. Ablation experiments were first conducted on three improved modules (feature extractor, FPN, and anchor box size) to find the contributions of each module. Then, different heuristic-based and DL methods, including Faster R-CNN [26], YOLOv3 [27], SSD [28], and Landserf Peak Classification (LPC) [29, 30], were compared to validate the effect of improved Faster R-CNN. The selection of the methods considered their relevance and heterogeneity along with the availability of the source code or a tool supporting their execution.

4.1. Evaluation Metrics and Parameter Selection. The precision, recall, F1 score, and average precision (AP) were selected to evaluate the performance of the improved Faster R-CNN. These metrics are defined as follows:

$$P = \frac{T_p}{(T_p + F_p)},$$

$$R = \frac{T_p}{(T_p + F_n)},$$

$$F_1 = \frac{2 \times P \times R}{(P + R)}, \tag{2}$$

$$AP = \frac{1}{11} \sum_{R \in \{0, 0.1, \ldots, 1\}} P(R),$$

where T_p, F_p, and F_n denote the true positive, false positive, and false negative rates, respectively. Let P be the precision and R be the recall, while F_1 balances P and R. Besides, AP

summarizes the shape of the precision and recall curves to avoid the problem that the threshold is difficult to evaluate the effect of the model absolutely, and it is defined as the mean precision at a set of eleven equally spaced recall levels $[0, 0.1, \ldots, 1]$.

Each method was executed with different parameters, the values were sampled from the parameter space, and all the resulting parameter combinations were tested. For the traditional method, LPC took the DEM and two parameters as input. The two parameters were as follows: (1) The minimum height that a point must have had to be considered as a candidate summit. For this parameter, we tested values from 400 m to 6,000 m with a step size of 100 m because these two values were the lowest and the highest elevations of the territory under evaluation. (2) The minimum distance that was the local maxima in a region. We tested values from 900 m to 30 m with a step size of 15 m. This yielded 3,363 configurations. Each configuration made the algorithm run independently once, for a total of 3363 runs. Deep learning models have only one parameter: the probability threshold value to determine if a point is a summit. We tested a value range from 0.01 to 1 with a step of 0.01, yielding 100 configurations. The DL algorithm only ran once, choosing different thresholds to obtain different precision and recall.

4.2. Ablation Experiment. We analyzed the contributions of each module in our method, namely, ResNet-50, feature fusion, and size of anchor box to the overall performance. The experimental results are given in Table 2.

By comparing the results of Faster R-CNN, the replacement of ResNet-50 brought performance improvements on the AP, with a margin of 1.98%. This means that ResNet-50 replaced VGG16 as a feature extractor that can better represent the summit feature. The scale and aspect ratio of the anchor boxes were adjusted in Improved 2, and then the AP increased to 92.97%. By comparing Improved 3 and Faster R-CNN, the addition of the FPN and the adjustment of the anchor box brought performance improvements on the AP, with a margin of 2.82%. This validates the effectiveness of our FPN and anchor box adjustment strategy. Finally, the proposed method (Row 6 in Table 2) was tested, and its AP reached 94.49%. The effectiveness of the three improvements, including the replacement of ResNet-50, the addition of FPN, and the adjustment of the anchor box, was consistently demonstrated.

The feature maps on the SUMMIT-DEM data set are shown in Figure 5. The input images, detection results, and features of the last two layers of Faster R-CNN and improved
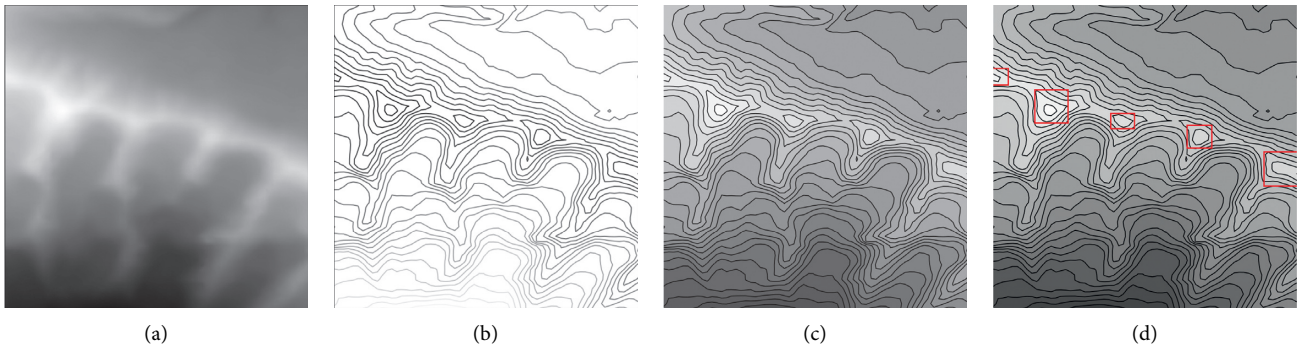
FIGURE 4: The sample construction process of mountain summits area. (a) Grayscale image based on elevation data. (b) Contour line image. (c) Superimposed images of (a) and (b). (d) Ground-truth images. The red box is the label of the mountain summit area.

TABLE 2: Results of the ablation experiments.

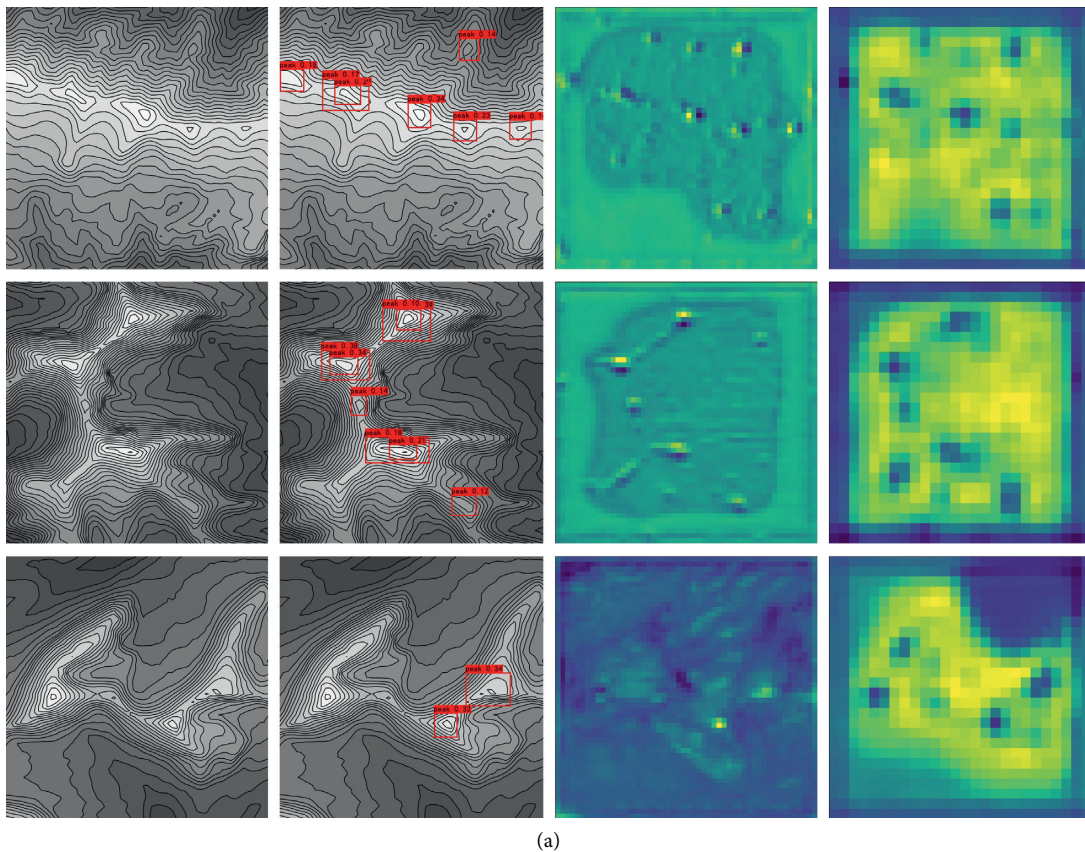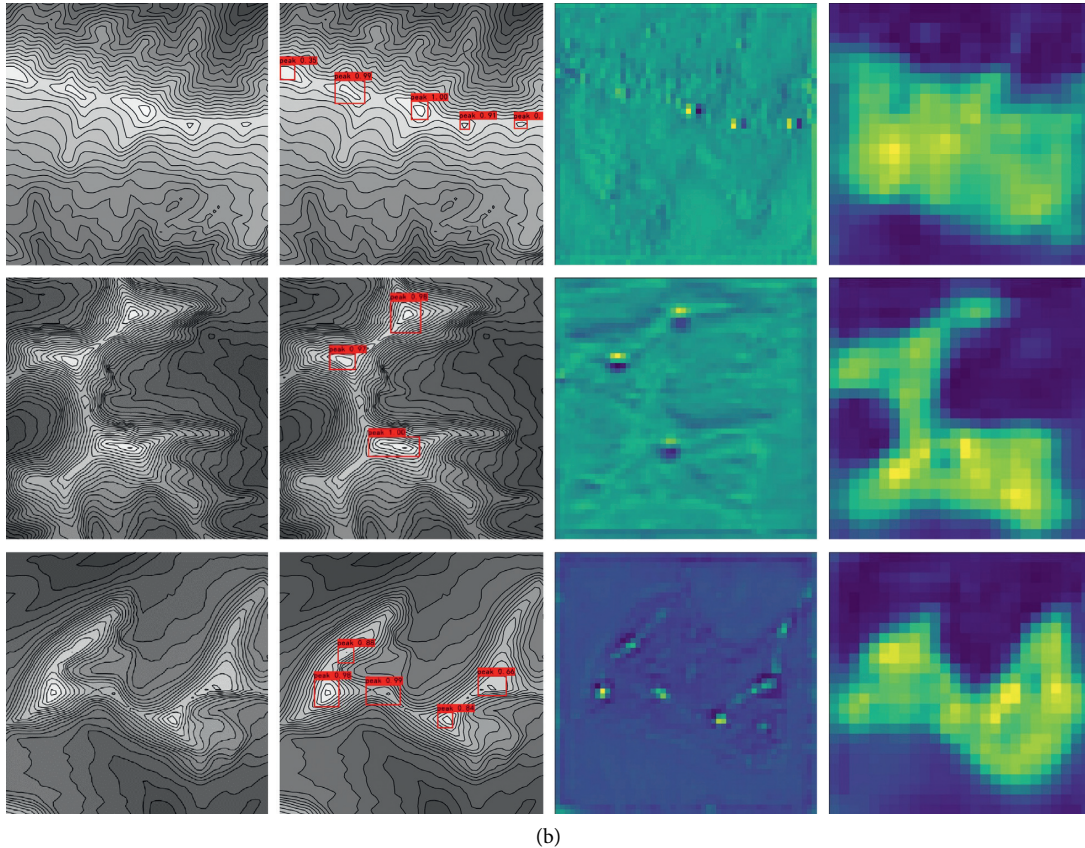| Methods	Feature extractor	Feature fusion	Anchor size	AP (%)
Faster R-CNN	VGG16	RPN	(128, 256, 512) {0.5, 1, 2}	90.55
Improved 1	**ResNet-50**	RPN	(128, 256, 512) {0.5, 1, 2}	92.53
Improved 2	VGG16	RPN	**(42, 102, 302) {0.66, 0.58, 0.59}**	92.97
Improved 3	VGG16	**FPN + RPN**	**(42, 102, 302) {0.66, 0.58, 0.59}**	93.37
Ours	**ResNet-50**	**FPN + RPN**	**(42, 102, 302) {0.66, 0.58, 0.59}**	**94.49**

(a)

FIGURE 5: Continued.

(b)

FIGURE 5: The feature maps of Faster R-CNN and the improved Faster R-CNN. (a) From left to right, the first to fourth items are the input images, detection results, feature maps of C4, and feature maps of C5. (b) From left to right, the first to fourth items are the input images, detection results, feature maps of P3, and feature maps of P2.

Faster R-CNN are shown in Figure 5. In Column 3, the filter could learn the summit locations in the form of blobs. As expected, the improved Faster R-CNN had more information at its disposal to distinguish false summits and find small summits because it preserved the complete location and semantic features via the FPN. Furthermore, the improved Faster R-CNN could exploit the "context" of a location. In Column 4, the improved Faster R-CNN preserved the correlations among the different locations comprised in the adjacent points, that is, the network paid more attention to the surrounding mountains, which affected the computation of the summit locations and, ultimately, the accuracy.

4.3. Evaluation of Traditional and DL Methods.

We compared the proposed method with different methods, including Faster R-CNN (FR), YOLOv3, SSD, and LPC on the SUMMIT-DEM data set.

Figure 6 shows the summit detection results of the different methods. As shown, the proposed model was the closest to the ground truth in various summit categories, including minor, submajor, and major. More importantly, the proposed model could well identify the

pseudosummits and find the small summits, demonstrating the effectiveness of the proposed improved Faster R-CNN.

Table 3 reports the recall, precision, F1 score, and AP of the proposed method compared with the other methods. The improved Faster R-CNN achieved excellent results on all the evaluation metrics. The SSD method had the worst performance because of its limited ability to extract shallow features and a hierarchical prediction mechanism, which made the predicted feature maps have a low utilization rate. In particular, comparing Faster R-CNN (Column 3) and the improved Faster R-CNN (Column 2), the recall improved with a margin of 6.45%, which means more summits were discovered. These results clearly illustrate the superior performance and robustness of the improved Faster R-CNN.

The precision-recall curves are shown in Figure 7. The PR curves of the improved Faster R-CNN, represented by the straight blue lines, consistently outperformed all the other methods. Compared with DL methods, the heuristic-based method, represented by the straight purple lines, was more sensitive to parameter changes. At point precision = 0.5 and recall = 0.5, small parameter changes

(a)   (b)   (c)   (d)   (e)   (f)
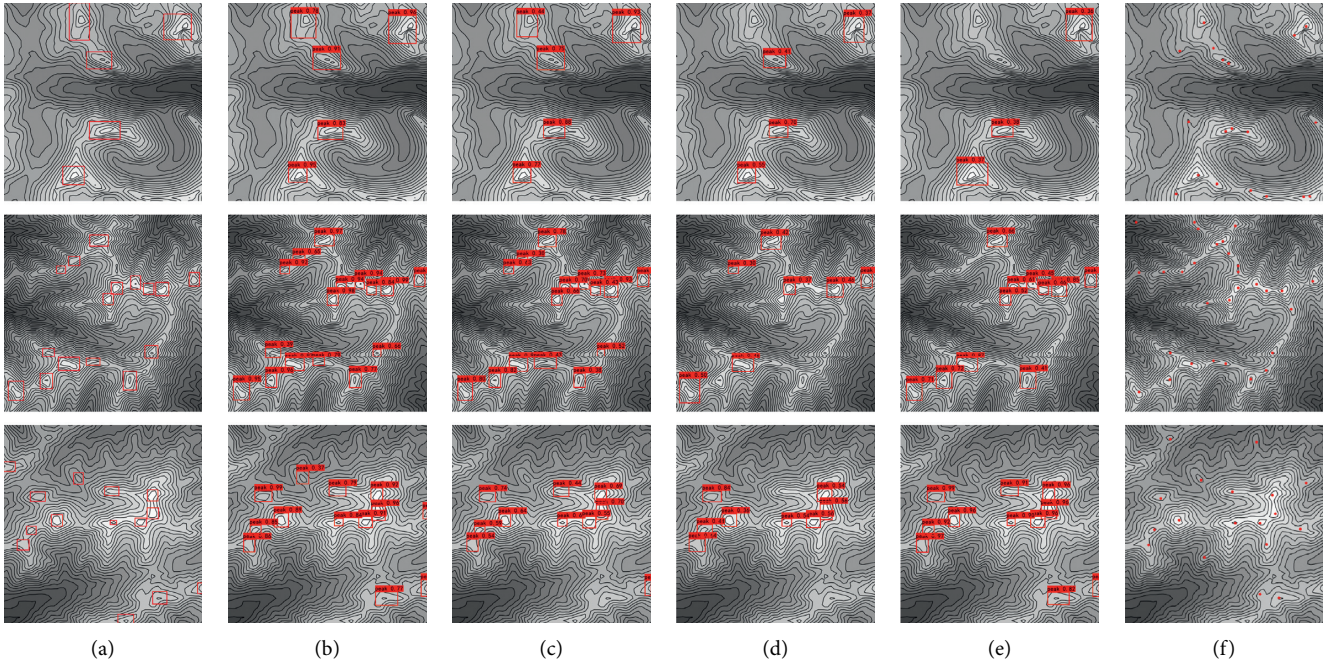
FIGURE 6: Mountain summit area recognition results of different networks. (a) Ground-truth images. The red boxes denote the ground-truth mountain summits area. (b) The recognition results of the proposed improved Faster R-CNN. (c) The recognition results of YOLOv3. (d) The recognition results of SSD. (e) The recognition results of the network were designed by prior work [26]. (f) The recognition results of LPC.

TABLE 3: The recognition results of different network models on the test data set.

Measures	Improved FR	FR-based	YOLOv3	SSD	LPC
Recall	**86.63%**	80.18%	81.89%	66.41%	53.79%
Precision	**93.98%**	91.55%	91.16%	90.57%	77.58%
F1 score	**0.90**	0.85	0.86	0.77	0.64
AP	**94.49%**	90.53%	90.69%	85.59%	59.62%

*The boldface in the table is the best result of the evaluation index.
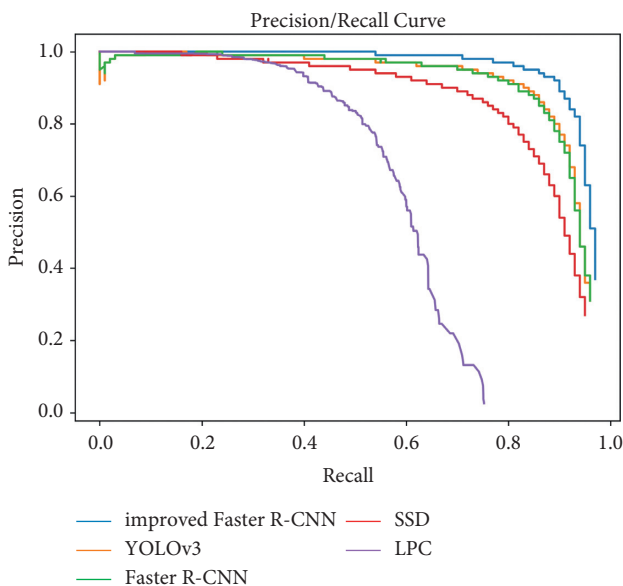

FIGURE 7: The P-R curve of each algorithm.

could lead to very different results in precision and recall. These results convincingly demonstrate the effectiveness of the proposed model.

5. Conclusions

In this paper, a novel DL method, improved Faster R-CNN, was proposed for summit detection without manually designed summit features. In the improved Faster R-CNN, ResNet-50 is used as the feature extractor to obtain better features of the summit, and the hierarchical structure of the FPN is applied to integrate features of various scales. Benefiting from these two improved modules, efficient information communication across multiple layers is conducted, reducing the information loss during RPN anchor box generation, which leads to more accurate summit detection results. In experimental studies, SUMMIT-DEM data set was used to study the performance of the improved Faster R-CNN. Experiments were conducted in different popular DL and heuristic-based methods, demonstrating the effectiveness and robustness of the improved Faster R-CNN.

Our future work will pursue several directions: (1) Topographic elements are often symbiotic, such as when the saddle is between two summits and the ridge is the connection between the summits; thus, we will improve the model by applying other elements and some methods [31] to process a set of objects simultaneously through interaction between their appearance feature and topology, which could allow modeling of their relations. (2) Multi-information fusion is also a problem worthy of attention. The DEM avoids the influence of the illumination and viewing angle on the experimental results, but it loses many detailed features, such as color, texture, and contrast. We will try to use our method to experiment on remote sensing images in the future and find a method to combine DEM features and remote sensing image features to recognize topographic elements. (3) Due to the conventional nature of cartography, which often only contains prominent mountains for morphological, historical, and cultural reasons, a data set may omit many locations with summit-like characteristics. Therefore, some of the output classified as false positives may indeed be true positives under a complete ground truth. We will apply semisupervised learning [32] to improve the quality of summit data sets, specifically, combining labeled and unlabeled data to change the learning behavior of the network.

Data Availability

The processed data used to support the findings of this study have not been made available because the data also forms part of an ongoing study.

Conflicts of Interest

There are no potential competing interests in our paper.

Acknowledgments

References

[1] C. Körner, W. Jetz, J. Paulsen, D. Payne, K. Rudmann-Maurer, and E. M. Spehn, "A global inventory of mountains for biogeographical applications," *Alpine Botany*, vol. 127, no. 1, pp. 1–15, 2017.

[2] C. De Jong and T. Barth, "Challenges in hydrology of mountain ski resorts under changing climatic and human pressures," in *Proceedings of the ESA Surface Water Storage and Runoff: Modeling, In-Situ Data and Remote Sensing*, Karlsruhe, Germany, September 2008.

[3] R. Fedorov, A. Camerada, P. Fraternali, and M. Tagliasacchi, "Estimating, ESA Proceedings, 2008 snow cover from publicly available images," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1187–1200, 2016.

[4] S. Yamada, "Mountain ordering: a method for classifying mountains based on their morphometry," *Earth Surface Processes and Landforms*, vol. 24, no. 7, pp. 653–660, 1999.

[5] P. Fisher and J. Wood, "What is a mountain? Or the Englishman who went up a boolean geographical concept but realised it was fuzzy," *Geography*, vol. 83, no. 3, pp. 247–256, 1998.

[6] Y. Deng and J. P. Wilson, "Multi-scale and multi-criteria mapping of mountain peaks as fuzzy entities," *International Journal of Geographical Information Science*, vol. 22, no. 2, pp. 205–218, 2008.

[7] T. Podobnikar, "Mountains'peaks determination supported with shapes analysis," *Geographia Technica*, vol. 5, pp. 111–119, 2010.

[8] R. Pires de Lima and K. Marfurt, "Convolutional neural network for remote-sensing scene classification: transfer learning analysis," *Remote Sensing*, vol. 12, no. 1, p. 86, 2020.

[9] K. Fu, Z. Chang, Y. Zhang, G. Xu, K. Zhang, and X. Sun, "Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 294–308, 2020.

[10] Z. Huang, B. Sui, J. Wen et al., "An intelligent ship image/video detection and classification method with improved regressive deep convolutional neural network," *Complexity*, vol. 2020, Article ID 1520872, 11 pages, 2020.

[11] X. Zhang, L. Han, L. Han, and L. Zhu, "How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery?" *Remote Sensing*, vol. 12, no. 3, p. 417, 2020.

[12] Z. Chang, Z. Du, F. Zhang et al., "Landslide susceptibility prediction based on remote sensing images and GIS: comparisons of supervised and unsupervised machine learning models," *Remote Sensing*, vol. 12, no. 3, p. 502, 2020.

[13] S. A. Naghibi, H. R. Pourghasemi, and B. Dixon, "GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran[J]," *Environmental Monitoring and Assessment*, vol. 188, no. 1, pp. 1–27, 2016.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[16] H. Sang, L. Xiang, S. Chen et al., "Image recognition based on multiscale pooling deep convolution neural networks," *Complexity*, vol. 2020, Article ID 618031, 13 pages, 2020.

[17] W. Li, B. Zhou, C. Y. Hsu et al., "Recognizing terrain features on terrestrial surface using a deep learning model: an example with crater detection," in *Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery*, pp. 33–36, Los Angeles Area, CA, USA, 2017.

[18] W. Li and C.-Y. Hsu, "Automated terrain feature identification from remote sensing imagery: a deep learning approach," *International Journal of Geographical Information Science*, vol. 34, no. 4, pp. 637–660, 2020.

[19] R. N. Torres, P. Fraternali, F. Milani et al., "A deep learning model for identifying mountain summits in digital elevation model data," in *Proceedings of the 2018 IEEE first international conference on artificial intelligence and knowledge engineering (AIKE)*, pp. 212–217, IEEE, Laguna Hills, CA, USA, September 2018.

[20] K. He, X. Zhang, S. Ren et al., "Deep Residual Learning for Image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[21] T. Y. Lin, P. Dollár, R. Girshick et al., "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, July 2017.

[22] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14, pp. 281–297, Los Angeles, CA, USA, June 1967.

[23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[24] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Computer Vision-ECCV 2014*, pp. 740–755, Prague, Czech Republic, May 2014.

[25] T. G. Farr and M. Kobrick, "Shuttle radar topography mission produces a wealth of data," *Eos, Transactions American Geophysical Union*, vol. 81, no. 48, pp. 583–585, 2000.

[26] S. Ren, K. He, R. Girshick et al., "Faster R-CNN: towards real-time object detection with region proposal networks," 2015, http://arxiv.org/abs/1506.01497.

[27] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, http://arxiv.org/abs/1804.02767.

[28] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multiBox detector," in *European Conference on Computer Vision-ECCV 2016*, pp. 21–37, Amsterdam, The Netherlands, October 2016.

[29] J. Wood, "The geomorphological characterisation of digital elevation models," Thesis, University of Leicester, Leicester, England, 1996.

[30] J. Wood, "Chapter 14 geomorphometry in LandSerf," *Developments in Soil Science*, vol. 33, pp. 333–349, 2009.

[31] H. Hu, J. Gu, Z. Zhang et al., "Relation networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3588–3597, Salt Lake City, UT, USA, June 2018.

[32] A. Tarvainen and H. Valpola, "Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results," 2017, http://arxiv.org/abs/1703.01780.