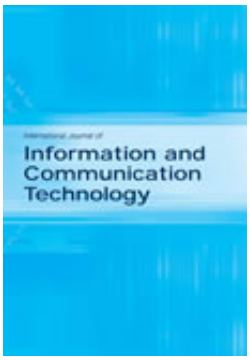


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Open-pit coal mine detection in large-scale remote sensing images

Qiang Liu, Xiaoliang Yang, Chongyang Wei, Kenan Cheng, Junzheng Wu

DOI: 10.1504/IJICT.2025.10072770

Article History:

Received:	14 April 2025
Last revised:	11 June 2025
Accepted:	11 June 2025
Published online:	26 August 2025

Open-pit coal mine detection in large-scale remote sensing images

Qiang Liu, Xiaoliang Yang, Chongyang Wei, Kenan Cheng and Junzheng Wu*

Northwest Institute of Nuclear Technology,
Xi'an, Shaanxi, 710024, China
Email: liuqiang2020@stu.xjtu.edu.cn
Email: yangxiaoliang@nint.ac.cn
Email: weichongyang@nint.ac.cn
Email: chengkenan@nint.ac.cn
Email: jasonzheng0404@126.com
*Corresponding author

Abstract: Satellite remote sensing is crucial for large-scale, regular monitoring of mineral mining. It helps grasp the mining status and fight illegal activities. However, challenges like scarce training datasets, low detection efficiency, fragmented targets, and inaccurate positioning exist. This paper focuses on open-pit coal mines. It presents a saliency-guided image-cutting algorithm and an improved non-maximum suppression-based object-relocation algorithm. These are integrated with a deep-learning object-detection model to form a deep-learning-based mine-detection framework. Tests on 10 large-scale images show the framework achieves 85.22% recall and 45.73% precision efficiently, outperforming pure deep-learning models.

Keywords: coal mine; open-pit mining area; deep learning; saliency detection; detection framework.

Reference to this paper should be made as follows: Liu, Q., Yang, X., Wei, C., Cheng, K. and Wu, J. (2025) 'Open-pit coal mine detection in large-scale remote sensing images', *Int. J. Information and Communication Technology*, Vol. 26, No. 31, pp.25–39.

Biographical notes: Qiang Liu is an Assistant Research Fellow, M.S., Graduated from Xi'an Jiaotong University in 2022. Worked in Northwest Institute of Nuclear Technology. Her research interests include Computer Applications.

Xiaoliang Yang is an Assistant Research Fellow, received his PhD and graduated from National University of Defense Technology in 2010. He worked in Northwest Institute of Nuclear Technology. His research interests include image interpretation.

Chongyang Wei is an Assistant Research Fellow, received his PhD and graduated from National University of Defense Technology in 2016. He worked in Northwest Institute of Nuclear Technology. His research interests include image processing.

Kenan Cheng is an Assistant Research Fellow, received her MS and graduated from Xi'an Jiaotong University in 2019. She worked in Northwest Institute of Nuclear Technology. Her research interests include image processing.

Junzheng Wu is an Associate Research Fellow, received his PhD and graduated from National University of Defense Technology in 2022. He worked in Northwest Institute of Nuclear Technology. His research interests include object change detection.

1 Introduction

China is rich in mineral resources, with all kinds of large, medium and small mines all over the country. Using satellite remote sensing technology to regularly monitor the mining situation in a large scale is an important means to timely grasp the development and utilisation of mineral resources and availably combat the illegal mining. Up to date, researches of mine monitoring using remote sensing images mainly focus on the identification and classification of mining features such as tailings pond, open-pit, buildings in mining area, etc. The existing methods can be roughly divided into three categories: visual interpretation, target detection based on artificial feature and traditional classifiers, and target detection based on depth convolution neural networks (DCNN).

The accuracy of visual interpretation can be guaranteed, which is the main means for the national ministries and commissions to carry out relevant investigations and digital product production. However, it requires the participation of a large number of experienced interpreters, which makes it time-consuming, laborious and costly (Zhang et al., 2022). Methods of the second category can achieve high accuracy in some specific scenarios with clear steps and obvious interpretability of features. Whereas, the applicability of these methods need to be further improved. The target detection based on DCNN mainly focuses on the detection of open pit and tailing ponds. For example, Gallwey et al. (2020) adopted the depth CNN model to detect small manual mining areas on Sentinel-2 multispectral images, which obtained the error less than 8%. Zhang (2019) used CNN to automatically extract open pit on high spatial resolution images, which can obtain a recall rate of 91.3%. Lyu et al. (2021) used the YOLO-4 model to detect tailing ponds on high spatial resolution image, and the accuracy reached 89.7%. On the whole, although some preliminary applications have achieved promising results, most existing deep learning based methods only evaluate the accuracy on the constructed datasets, which can be hardly extend to the applications on large-scale remote sensing images.

Some similarities exist between optical remote sensing images and natural images, which makes some detection algorithms for natural images can be used on remote sensing images. However, unique characteristics exist in remote sensing images, such as large size, diverse target directions, uneven target distribution, and complex backgrounds, which result in some new challenges. To solve the problem of large size, the most common method is to cut the image into blocks and then carry out detection on each block. However, it will produce a large number of blocks without targets as the proportion of targets is small and the distribution is uneven in large size images, resulting in low detection efficiency. Meanwhile, the targets located across blocks may be segmented, which would lead to low positioning accuracy. Wang et al. (2018) cut the image with a certain overlap rate, which improved the integrity of the target to a certain extent, while the number of image blocks increased, resulting in greater time redundancy. To reduce the time redundancy, R2-CNN (Pang et al., 2019) designed a strategy of ‘judge first, locate later’ to filter out blocks without target by considering the fact that most

blocks are background and the target to be detected is only concentrated in small areas. Thus, the computational burden can be reduced, but the issue that targets may be segmented is still unsolved. YOLT (Etten et al., 2018) used multiple lightweight models to detect the blocks, and then the detected results were fused, which improved the speed with relatively high accuracy. Zhou et al. (2024) proposed an adaptive attention-guided framework that integrates the convolutional block attention module (CBAM) and Transformer networks to enhance feature discrimination capability in cluttered scenes. The method demonstrates higher accuracy in detecting small and fragmented targets. Overall, the aforementioned methods are based on the strategy of evenly dividing the image, which may not be optimal.

To sum up, the following problems still exist in large-scale and high-frequency mine monitoring using remote sensing images based on deep learning:

1 Relevant research is less, resulting in insufficiency of dataset that can be used for model training. Furthermore, most of related works are only conducted on small-scale remote sensing images, but no framework for large-scale images has emerged.

2 For remote sensing images with large coverage, the conventional method of first evenly cutting and then detecting block by block is not only inefficient, but also causes problem that the target may be segmented.

3 Due to the poor monomer and unclear boundary of mine, accurately positioning mine targets faces more challenges.

To solve these problems, this paper takes the open-pit coal mine as the research object, proposes an image cutting algorithm based on saliency guidance and a target relocation algorithm based on improved non-maximum suppression, and combines them with the deep learning object detection model to form a mine detection framework based on deep learning methods, which can realise the automatic detection of the open-pit coal mine in large-scale remote sensing images. Key contributions:

1 Proposed a saliency-guided image cutting algorithm to reduce background redundancy while preserving target integrity, reducing average per-image detection time by 57.62% for Faster R-CNN models.

2 Developed an improved non-maximum suppression (NMS) algorithm for nested target relocation, effectively resolving boundary ambiguity issues.

3 Introduced an integrated deep learning framework combining guided image cutting and target relocation, achieving 85.22% recall rate and 45.73% precision on large-scale imagery.

2 Design of the algorithm

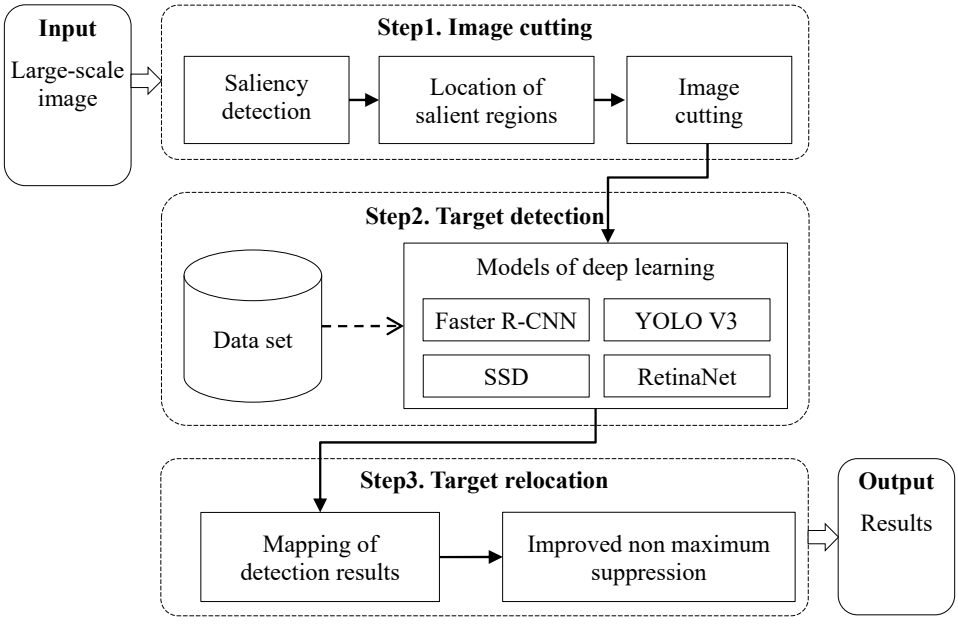
The proposed framework can be illustrated as Figure 1, which mainly contains the following steps:

1 the saliency guided image cutting algorithm is used to cut the large-scale remote sensing images and obtain the image blocks to be detected

2 the dataset of open-pit coal mine is used to train the model, and the target detection of the image blocks to be detected is completed

3 the target relocation algorithm based on improved non-maximum suppression was used to map and optimise the target, and the final detection results were obtained.

Figure 1 Illustration of the proposed framework



2.1 The image cutting algorithm based on saliency guidance

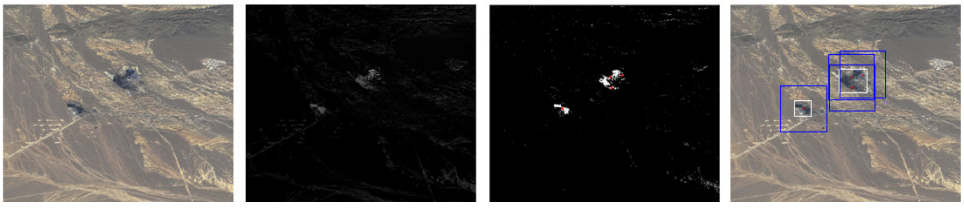
In the preprocessing of remote sensing images, background suppression and target region extraction are critical for improving detection efficiency. Wang et al. (2021) proposed a robust principal component analysis (RPCA) method for aluminum foil surface defect detection, which effectively separated defect features from background noise through low-rank and sparse decomposition. Inspired by this, our study employs a saliency-guided image cutting algorithm to focus on visually prominent regions, addressing the issue of target fragmentation in large-scale images more adaptively than traditional uniform cutting methods. The algorithm includes three steps: saliency detection, location of salient regions and image cutting.

2.1.1 Saliency detection

Saliency detection can quickly and effectively focus the salient regions with visual saliency differences in the image. Existing saliency detection algorithms can be roughly divided into two categories: bottom-up and top-down. The bottom-up algorithms are stimulus driven, which directly use the underlying visual features such as colour and brightness or various heuristic prior clues (such as contrast prior, background, connectivity prior, centre prior, etc.) to locate salient regions. The top-down algorithms

are task driven, which use training data with category labels to analyse task oriented visual attention, which require a large number of labelled data. In view of the visual characteristics of the open-pit coal mine and the lack of training data, a bottom-up IG algorithm based on global contrast is adopted to extract salient blocks in this paper. One example is shown in Figure 2(b).

Figure 2 Steps of obtaining candidate blocks based on saliency guidance, (a) original image (b) saliency map (c) salient region map (d) candidate image blocks (see online version for colours)



(a) (b) (c) (d)

Notes: The red dots in (c) represent the centroid of each salient region.
The blue wire frame in (d) is the cutting position and the white wire frame is the ground truth of the target.

IG algorithm (Achanta et al., 2009) is a frequency-tuned saliency detection method. The main idea is to obtain pixel level saliency maps by calculating the difference between the average colour value (RGB or Lab) of each pixel (representing low frequency) and the whole image (representing high frequency) after Gaussian filtering. The saliency map S of image I can be calculated as:

$$S(i, j) = \left\| I_{\mu} - I_{\omega_{hc}}(i, j) \right\| \tag{1}$$

where I_{μ} is the mean image feature vector, $I_{\omega_{hc}}(i, j)$ is the corresponding image pixel vector value in the Gaussian blurred version of the original image. Using the Lab colour space, each pixel is a $[L, a, b]^T$ vector, and $\| \|$ is the Euclidean distance.

2.1.2 Location of salient regions

The saliency maps are values within the range of [0, 255]. After using the Otsu algorithm binarisation, a map of a salient region can be obtained, which identifies the areas that are significant in the bottom visual features of the whole image, including the open-pit coal mine and some other objects. Then, according to the statistical prior of the target scales, eliminating the regions which are too small, the salient regions with a certain scale can be obtained, as shown by Figure 2(c).

The Otsu (Ostu, 1979) algorithm determines the best segmentation threshold by maximising the between-class variance of foreground and background. The method for computing the between-class variance g of a greyscale image I can be calculated as:

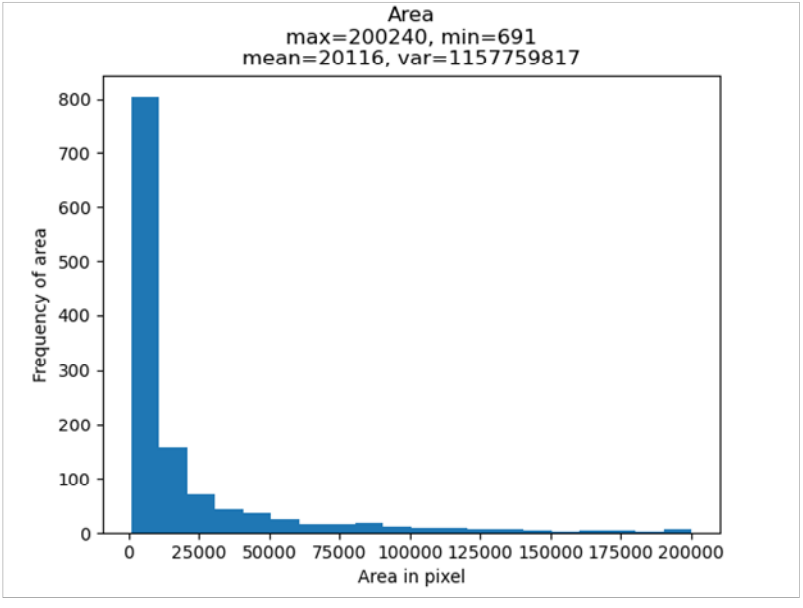
$$g = \omega 0 \times \omega 1 \times (\mu 0 - \mu 1)^2 \tag{2}$$

Here, g is defined as the inter-class variance between foreground and background, where:

- $ω0$: proportion of foreground pixels to total image pixels

- $μ0$: average greyscale of all foreground pixels
- $ω1$: proportion of background pixels to total image pixels
- $μ1$: average greyscale of all background pixels

Figure 3 Histogram of target area distribution in the coal mine dataset (see online version for colours)



The optimal segmentation threshold is determined by exhaustively searching for the threshold value that maximises g. Subsequently, we statistically analysed target areas in the coal mine dataset, generating the area distribution histogram shown in Figure 3. Using the minimum area value (691) as the threshold, we filtered out excessively small connected regions, retaining only salient connected regions of reasonable scale as candidate target areas [Figure 2(c)].

2.1.3 Image cutting

In order to obtain candidate blocks, the connected areas in Figure 2(c) are numbered and their centroids are calculated one by one. Then, the image is cut according to the prior scale of targets to be detected with the centroid coordinate as the centre point to obtain candidate blocks. Based on the dataset statistics, the average target diameter is approximately 600 pixels. Thus, the image block size is set to 800 × 800 pixels to ensure target integrity within each block, as shown by Figure 2(d). The centroid position (x, y) of a connected area Ω can be calculated by:

$$\begin{cases} x = \left(\sum_{i,j;(i,j)\in\Omega} i * S(i,j)\right) \Big/ \left(\sum_{i,j;(i,j)\in\Omega} S(i,j)\right) \\ y = \left(\sum_{i,j;(i,j)\in\Omega} j * S(i,j)\right) \Big/ \left(\sum_{i,j;(i,j)\in\Omega} S(i,j)\right) \end{cases} \tag{3}$$

Different from the commonly used methods which directly cut the image with a certain overlap rate, our algorithm can not only filter most of background, but also ensure the integrities of targets in candidate blocks, which may improve the subsequent detection efficiency and accuracy.

2.2 Target detection based on deep learning models

According to whether the region of interest (ROI) is required, target detection methods based on deep learning can be divided into two categories: two stage and one stage target detection. Two stage methods use the region recommendation network (RPN) to extract regions of interest, and then intercept the local feature layer of each ROI for classification and regression prediction. Typical methods include R-CNN (Ross et al., 2016), fast R-CNN (Girshick et al., 2015), faster R-CNN (Shaoqing et al., 2017), R-FCN (Dai et al., 2016), and mask R-CNN (He et al., 2017). These methods can usually achieve high accuracy. However, efficiency is usually dissatisfactory due to the extraction of regions of interest which needs additional computation. One stage methods directly extract features from multi-layer deep feature maps for classification and position regression without extraction of regions of interest, such as SSD (Liu et al., 2016), the YOLO series (Redmon et al., 2016; Jo et al., 2017; Redmon and Farhadi, 2018; Bochkovskiy et al., 2020; Wang et al., 2023), SSD (Liu et al., 2016), and RetinaNet (Lin et al., 2017). Speeds of these methods can basically meet the requirements of real-time systems, but the accuracies are slightly lower than those of two stage methods. Additionally, object detection algorithms can be categorised into anchor-based and anchor-free methods based on whether predefined anchor boxes are required. Anchor-based algorithms necessitate predefining a series of anchor boxes with varying sizes and aspect ratios, decoding final detection results by predicting anchor offsets; these methods require statistical analysis of target dimensions and aspect ratios in training data to determine hyperparameters such as anchor scales, aspect ratios, anchor density, and IoU thresholds, resulting in limited generalisation capability. In contrast, anchor-free algorithms eliminate anchor boxes, directly predicting pixel-wise probabilities of target presence and bounding box coordinates to generate detections, offering superior generalisation and effectiveness for irregularly shaped targets, with representative methods including CenterNet (Zhou et al., 2019), FCOS (Tian et al., 2019), YOLOX (Ge et al., 2021), and FoveaBox (Kong et al., 2020).

This study selects mature and representative models from various object detection methodologies – faster R-CNN, SSD, YOLOv3, YOLOv5, RetinaNet, CenterNet, and YOLOX – to conduct detection experiments, enabling comparative analysis of their precision and applicability on our custom Open-Pit Coal Mine Target Detection Dataset. Sequentially, the model with highest performance is adopted to detect the candidate blocks obtained in Section 2.1.

2.3 The target relocation algorithm based on improved non-maximum suppression

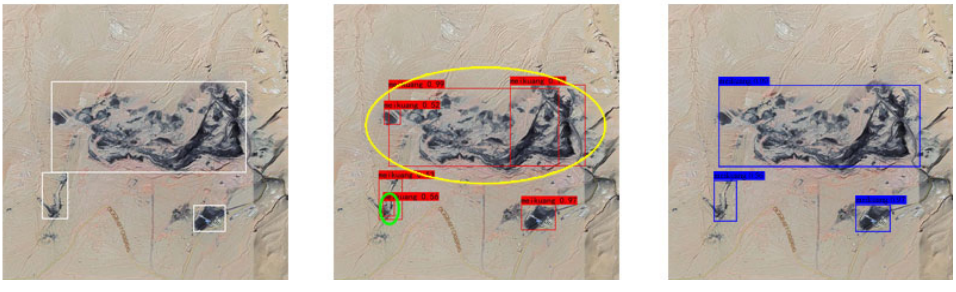
After extraction of deep detector, target classification and position regression results on all candidate blocks are obtained. Sequentially, these location boxes need to be mapped to the original large-scale image, and the nested overlapping objects need to be filtered and relocated to obtain the final detection results.

The most commonly used method for eliminating repeatedly detected results is non-maximum suppression which comprehensively considers the located box score and the intersection over union (IOU) of each located box, and removes the non-maximum score results whose intersection over union are greater than a certain threshold. However, because the open-pit coal mines have the characteristics of unclear boundary, poor monomer, nesting and overlapping relationships between targets, this method can't be effectively applicable for our study. For instance, when a small target is nested in a large one, the small one can't be eliminated due to the small intersection over union, as marked by the green circle in Figure 3(b). Besides, when a target is located by multiple boxes, the maximum range covered by each box is the most accurate location result, as shown by the yellow circle in Figure 3(b).

Table 1 Detailed description of target relocation algorithm

Algorithm step	*Specific operation*
Step 1: Sorting	• Sort all Bboxes by area (descending) • Set flags $f_1, f_2, \ldots, f_n = 0$ for all Bboxes
Step 2: Grouping	For *m* Bboxes where $f_i = 0$: • Take the 1st Bbox as reference box for group *k* • For subsequent *m – 1* Bboxes: ‣ Set flag f_i = *k* if its centre lies within the reference box
Step 3: Merging	For group *k*: • Merge all Bboxes to generate NewBbox: ‣ Top-left: (min(x), min(y)) ‣ Bottom-right: (max(x), max(y)) ‣ Score: max(score) in group
Loop	Iteratively execute Steps 2-3 until all flags $f_1, f_2, \ldots, f_n \neq 0$

Figure 4 Target relocation based on improved non-maximum suppression, (a) ground true (b) results of non-maximum suppression (c) results of improved non-maximum suppression (see online version for colours)



(a) (b) (c)

Notes: The green circle in (b) represents the redundant target that has not been eliminated due to the small intersection ratio, while the yellow circle indicates the location boxes to be merged.

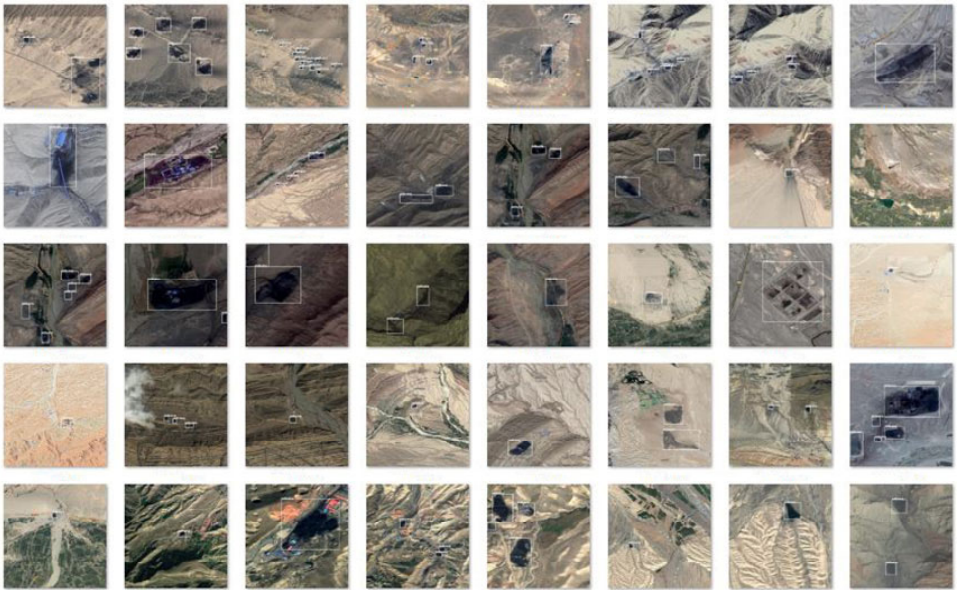
In this paper, an improved non-maximum suppression algorithm is proposed, which can further eliminate redundant nested targets and optimise location results. The location boxes obtained by the traditional non-maximum suppression are firstly sorted in descending order of area. Then, the location boxes whose centre points fall within the range of the current location box with maximum area are classified as a subset. Sequentially, the maximum and minimum corner coordinates among all location boxes in the subset are used to obtain the new location box with the highest score among all boxes in the subset. After traversing all the location boxes in turn, the new location box is the final result, as shown in Table 1. The final localisation results obtained by our algorithm are more consistent with the Ground truth, as shown in Figure 4(c).

3 Experimental results and analysis

3.1 Data and experimental setting

The research area in this paper is China's northwest region, where coal mines are widely distributed, with low density and various scales. In order to construct a dataset with higher applicable, 1,282 open-pit coal mines are labelled by visual interpretation in 398 images with the resolution range of 1 m to 8 m and the size of 800 * 800 from Map World. Figure 5 shows some samples of the dataset. As can be seen, the open-pit coal mines appear with great differences in scale and shape. Meanwhile, some ground objects such as shadows, residential areas, and vegetation coverage areas are similar to the open-pit coal mine to some extent, which makes the data better reflect the real scene, and also poses a great challenge to detection methods.

Figure 5 Samples of open-pit coal mines in our dataset (see online version for colours)



To verify the effectiveness of the proposed framework, ten remote sensing images of Gaofen Series Satellites located in Inner Mongolia are acquired as test images with the size of 3,000 to 6,000 pixels and the resolution of 1 m to 8 m, in which 88 targets are labelled as the ground truth by visual interpretation.

3.2 Experiments and analysis

According to the extracted method of candidate regions shown by Figure 2, ten aforementioned large-scale remote sensing images are cut into blocks. The conventional uniform cutting method is adopted to compare with our cutting method, which cutting image with the size of 800 × 800 pixels, step of 512 pixels (i.e., 288 pixels overlap). Statistics of candidate image blocks obtained by the two methods is shown in Table 2.

Table 2 Comparison of two cutting methods

Method	*Number of blocks*	*Block with targets*		*Number of complete targets in a block*
		Number	*Ratio (%)*	
Uniform cutting method	560	131	23.39	86
Our cutting method	156	92	58.97	88

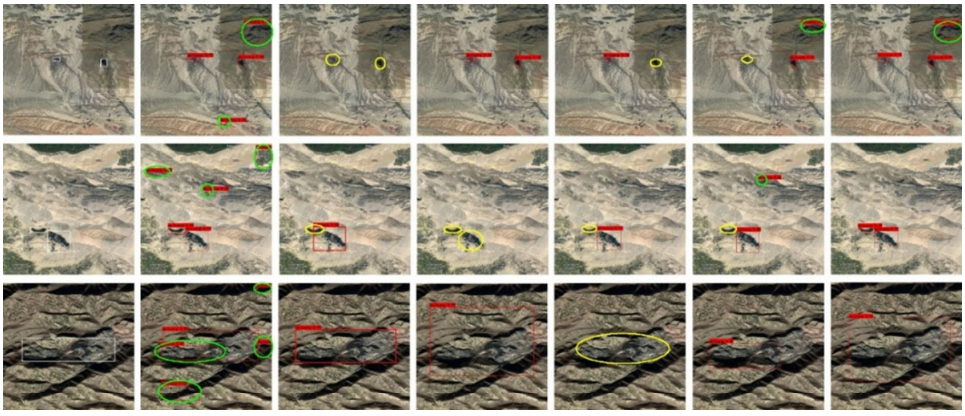
In terms of the number of blocks, 560 blocks are obtained by uniform cutting, of which only 131 contain targets, while 156 blocks are obtained by our cutting method, 92 of which contain targets. The rest blocks obtained by our method contain shadow of mountains or cloud, vegetation coverage areas and other ground objects which are similar to the coal mining area in colour and brightness. Thus, further detection model needs to differentiate targets and backgrounds. In terms of integrity of the objects, blocks obtained by uniform cutting contains 86 complete objects, while all 88 complete objects are contained in blocks obtained by our methods.

The above result indicates that our cutting method can effectively reduce the number of blocks and solve the problem of edge being segmented when used in cutting large-scale remote sensing images with certain differences in colour, brightness and other aspects.

Table 3 Performances of different models

Type	*Model*	*Backbone*	*mAP (%)*	*Recall (%)*	*Precision (%)*
Two-stage	Faster R-CNN	ResNet50	74.69	86.27	52.38
One-stage	SSD	VGG16	74.99	66.67	82.93
	YOLOv3	DarkNet-53	73.96	78.43	75.47
	YOLOv5-s	CSPDarknet	71.51	69.02	72.54
	YOLOv5-m	CSPDarknet	74.60	74.17	75.01
	RetinaNet	ResNet50	71.49	76.47	67.24
Anchor-free	CenterNet	ResNet50	75.13	79.46	71.35
	YOLOX	DarkNet-53	75.72	78.92	72.51
	YOLOX-s	CSPDarknet	73.08	70.22	74.88
	YOLOX-m	CSPDarknet	74.84	74.96	

Figure 6 Some results of different models, (a) GT (b) faster R-CNN (c) SSD (d) YOLOv3 (e) RetinaNet (f) CenterNet (g) YOLOX (see online version for colours)



(a) (b) (c) (d) (e) (f) (g)

The constructed dataset is randomly divided into training set, verification set and test set using the ratio of 8:1:1. Online data enhancement methods such as distortion, scaling, flipping, and colour gamut transformation are used to increase the diversity of training samples. Detection experiments are conducted using Faster R-CNN, SSD, YOLO V3, YOLO V5, RetinaNet, CenterNet and YOLOX models, respectively. The performance of each model on the coal mine dataset is presented in Table 3, with partial detection examples shown in Figure 6. Undetected targets are marked with yellow circles, while false positives are indicated by green circles. In terms of detection recall, Faster R-CNN detected more targets, whereas other models exhibited varying degrees of missed detections – primarily for small-sized targets and those with low contrast against the background. Regarding detection accuracy, Faster R-CNN's results included nested redundant objects and false alarms for visually similar targets such as shadows, vegetation-covered areas, and small dark surface features. YOLOX achieved the highest mean average precision (mAP).

Table 4 Average times of detecting single image

Models	*Average times of detecting single image (s)*	
	Uniform cutting	*Our cutting*
Faster R-CNN	6.89	2.92
SSD	1.10	1.31
YOLO V3	1.74	1.48
Retinanet	2.96	1.82
CenterNet	3.48	1.97
YOLOX	1.93	1.54

To study the running speeds, the trained models are used to detect the candidate blocks obtained by the two cutting methods. The average times of detecting single image are shown in Table 4. The experimental results show that our cutting method has great difference in the improvement effect of the average times of detecting single image when

using different detection models. The speed optimisation is the most obvious when using Faster R-CNN model detection, and the average times of detecting single image is shortened by 57.62%. This is because the detection speed of Faster R-CNN is the slowest among the detection models, and our cutting method greatly reduces the detection time by greatly reducing the number of candidate image blocks. At the same time, since the saliency detection in our cutting method is also time-consuming, the average times of detecting single image increases for the SSD model with faster detection speed.

In order to test the framework proposed in this paper, ablation experiments are conducted to test the optimisation effect of the two improved aspect: saliency guided cutting and target relocation. The results are shown in Table 5. As can be seen, saliency guided cutting and target relocation have significant effects on increasing the number of correct detections and reducing the number of false detections. Specifically, the saliency guided cutting method directly eliminates a part of the background without targets, thus the number of false detection is reduced. Meanwhile, considering the centroid of a salient region as the centre of a block can ensure the integrity targets in the block. The target relocation based on the analysis of the spatial distribution characteristics of open-pit coal mines can eliminate some redundant nested sub targets, and the small targets that conform to an inclusion relationship are combined into a large one, which may reduce the number of false detections. Additionally, both proposed algorithms demonstrate strong portability across diverse object detection models.

Table 5 Performance of different improved methods

Algorithm combination	*TP*	*FP*	*Recall (%)*	*Precision (%)*	*F1-Score (%)*
Uniform Cutting + Faster R-CNN	63	199	71.59	24.08	36.04
Guided Cutting + Faster R-CNN	68	123	77.27	35.60	48.75
Guided Cutting + Faster R-CNN + Relocation	75	89	85.22	45.73	58.36
Uniform Cutting + SSD	49	154	55.68	24.14	33.67
Guided Cutting + SSD	49	98	55.68	33.33	41.76
Guided Cutting + SSD + Relocation	51	72	57.95	41.16	47.67
Uniform Cutting + YOLOv3	53	161	60.23	24.77	34.70
Guided Cutting + YOLOv3	54	102	61.36	34.62	43.28
Guided Cutting + YOLOv3 + Relocation	60	80	68.18	42.86	52.27
Uniform Cutting + RetinaNet	52	169	59.09	23.53	33.17
Guided Cutting + RetinaNet	52	118	59.09	30.59	40.37
Guided Cutting + RetinaNet + Relocation	57	82	64.77	41.01	50.27
Uniform Cutting + CenterNet	59	166	67.05	30.67	41.26
Guided Cutting + CenterNet	60	118	68.18	33.71	45.58
Guided Cutting + CenterNet + Relocation	65	82	73.86	44.22	54.77
Uniform Cutting + YOLOX	61	175	69.32	25.85	36.86
Guided Cutting + YOLOX	62	124	70.45	33.33	44.72
Guided Cutting + YOLOX + Relocation	68	79	77.27	46.26	57.04

4 Conclusions

In order to realise coal mine detection oriented to large-scale remote sensing images and improve detection efficiency and accuracy, this study proposes an image cutting algorithm based on saliency guidance and a target relocation algorithm based on improved non-maximum suppression, and combines them with the deep learning object detection model to form a mine detection framework based on deep learning methods. Experiments on 10 large-scale remote sensing images show that compared with the object detection model alone, the recall can be improved from 71.59% to 85.22%, and the precision can be improved from 24.08% to 45.73% by using our framework. Besides, our image cutting algorithm can reduce the number of image blocks, improve the efficiency, and solve the problem of edge being segmented on large-scale remote sensing images with differences in colour, brightness. In practical applications, the large coverage area and complex backgrounds of remote sensing images introduce visually similar features to coal mines – such as mountain or cloud shadows, isolated vegetation patches, dark cultivated land plots, water bodies, and artificial ponds – which frequently trigger false positives in optical remote sensing-based detection. Consequently, both detection accuracy and recall rates on large-scale imagery experience varying degrees of degradation compared to test dataset performance. Future work will explore multi-source image fusion techniques (e.g., Synthetic Aperture Radar with optical imagery, thermal infrared with optical imagery) to mitigate these challenges. Furthermore, to mitigate the lack of multi-source sample data, the simulation-based approach for remote sensing image generation could be employed – exemplified by Dong et al.'s (2025) simulated SAR image generation method specifically designed for target recognition task, which server as an effective data augmentation solution.

Declarations

The author declares that it does not have any known interests or personal relationships that could potentially influence the reported work in this paper.

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

References

Achanta, R., Hemami, S., Estrada, F. and Susstrunk, S. (2009) 'Frequency-tuned salient region detection', *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1597–1604, IEEE.

Bochkovskiy, A., Wang, C-Y. and Liao, H-Y.M. (2020) 'YOLOv4: optimal speed and accuracy of object detection', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp.1–14, https://doi.org/10.48550/arXiv.2004.10934.

Dai, J., Li, Y., He, K. and Sun, J. (2016) 'R-FCN: object detection via region-based fully convolutional networks', in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS 2016)*, pp.379–387, Curran Associates Inc., https://doi.org/10.48550/arXiv.1605.06409.

Dong, C., Zhang, Y., Meng, X. and Guo, L. (2025) 'Application evaluation of simulated SAR image for target recognition', *Modern Applied Physics*, Vol. 16, No. 1, pp.011317-1–011317-7.

Etten, A.V. (2018) 'You only look twice: rapid multi-scale object detection in satellite imagery', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp.1–14, https://doi.org/10.1109/CVPR.2018.00008.

Gallwey, J., Robiati, C., Coggan, J., Vogt, D. and Eyre, M. (2020) 'A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: applying deep learning to shallow mining', *Remote Sensing of Environment*, Vol. 248, pp.111970-1–111970-15, https://doi.org/10.1016/j.rse.2020.111970.

Ge, Z., Liu, S., Wang, F., Li, Z. and Sun, J. (2021) 'YOLOX: exceeding YOLO Series in 2021', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, pp.1–10, https://doi.org/10.48550/arXiv.2107.08430.

Girshick, R. (2015) 'Fast R-CNN', *IEEE International Conference on Computer Vision*, pp.1440–1448.

He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017) 'Mask R-CNN', *Proceedings of the IEEE International Conference on Computer Vision*, pp.2961–2969.

Jo, K., Im, J., Kim, J. and Kim, D.S. (2017) 'A real-time multi-class multi-object tracker using YOLOv2', *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp.507–511, IEEE.

Kong, T., Sun, F., Liu, H., Jiang, Y., Li, L. and Shi, J. (2020) 'FoveaBox: beyond anchor-based object detection', *IEEE Transactions on Image Processing*, Vol. 29, No. 12, pp.7389–7398, https://doi.org/10.1109/TIP.2020.3002345.

Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P. (2017) 'Focal loss for dense object detection', *Proceedings of the IEEE International Conference on Computer Vision*, pp.2980–2988.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C. (2016) 'SSD: single shot multibox detector', *Computer Vision-ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, 11–14 October, Proceedings, Part I 14, pp.21–37, Springer International Publishing.

Lyu, J., Hu, Y., Ren, S., Yao, Y., Ding, D., Guan, Q. and Tao, L. (2021) 'Extracting the tailings ponds from high spatial resolution remote sensing images by integrating a deep learning-based model', *Remote Sensing*, Vol. 13, No. 4, p.743.

Ostu, N. (1979) 'A threshold selection method from gray-histogram', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp.62–66.

Pang, J., Li, C., Shi, J., Xu, Z. and Feng, H. (2019) 'R^2-CNN: fast tiny object detection in large-scale remote sensing images', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 8, pp.5512–5524, https://doi.org/10.1109/TGRS.2019.2899955.

Redmon, J. and Farhadi, A. (2018) 'YOLOv3: an incremental improvement', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.1–14, https://doi.org/10.1109/CVPRW.2018.00008.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) 'You only look once: unified, real-time object detection', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.779–788.

Ross, G., Jeff, D. and Trevor, D. (2016) 'Region-based convolutional networks for accurate object detection and segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 1, pp.142–158.

Shaoqing, R., Kaiming, H. and Ross, G. (2017) 'Faster R-CNN: towards real-time object detection with region proposal networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp.1137–1149.

Tian, Z., Shen, C. and Chen, H. (2019) 'FCOS: fully convolutional one-stage object detection', *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.9627–9636.

Wang, C., Bai, X., Wang, S., Zhou, J. and Ren, P. (2018) 'Multiscale visual attention networks for object detection in VHR remote sensing images', *IEEE Geoscience and Remote Sensing Letters*, Vol. 16, No. 2, pp.310–314.

Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. (2023) 'YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.7464–7475.

Wang, H., Gao, C. and Ling, Y. (2021) 'A deep learning-based method for aluminium foil-surface defect recognition', *International Journal of Information and Communication Technology*, Vol. 19, No. 3, pp.231–241.

Zhang, C., Li, J., Lei, S., Yang, J. and Yang, N. (2022) 'Progress and prospect of the quantitative remote sensing for monitoring the eco-environment in mining area', *Met. Mine*, Vol. 51, No. 1, pp.1–27.

Zhang, F.J. (2019) *Research on Deep Learning Extraction Method in Open Mining Area Based on Multi-Source Remote Sensing Images*, Anhui University, Hefei.

Zhou, W., Huang, S., Luo, Q. and Yu, L. (2024) 'Research on a ship target detection method in remote sensing images at sea', *International Journal of Information and Communication Technology*, Vol. 25, No. 12, pp.29–45.

Zhou, X., Wang, D. and Krähenbühl, P. (2019) 'Objects as points', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.7464–7475, https://doi.org/10.1109/CVPR.2019.00721.