

## Research and application of land illegal behavior monitoring based on video image recognition

Min Hu

Shanghai Surveying and Mapping Institute, Shanghai 200063, China

**Keywords:** Land illegal behavior, Farmland protection, Video monitoring, Space integration

### Abstract

Our country had increasingly high requirements for land management, especially for arable land, which was the most precious resource. Strict policies for protecting arable land are being implemented. The enforcement of satellite imagery in mega cities had certain limitations in terms of timeliness and accuracy. To meet the requirements of refined management, a high-altitude camera was constructed to form a near ground monitoring network covering the cultivated land area. Multi temporal image groups were obtained through video frame extraction. A monitoring sample library for typical behaviors of natural resources was established. Land objects closely related to illegal activities such as construction machinery and bulldozing areas were identified based on deep learning algorithms. By using the technology of bidirectional conversion between video spatial location and geographic coordinates, a fusion of spatial information video monitoring patterns was formed. Taking the 20000 acre permanent basic farmland area as an example, the recognition rate of illegal land related behaviors was about 81%. By combining approval information and conducting spatial analysis on the identified patterns, real-time warning information could be pushed. This application can effectively improve the timeliness of supervision and enhance the level of supervision.

### 1. Introduction

The protection of cultivated land is related to national food security, ecological security and social stability. For a long time, China has attached great importance to the protection of cultivated land resources, and resolutely stopped the "non-agricultural" and "non-grain" transformation of cultivated land. In order to coordinate the relationship between development and security, development and protection, consolidate the foundation of food security in an all-round way, firmly guard the red line of 1.8 billion mu of cultivated land, and emphasize high-quality land law enforcement work to help high-quality social development. As a super large city with limited land resources and continuously expanding population, Shanghai needs to strictly adhere to the red line of farmland protection. At the same time, Shanghai's work requirements of "detection and law enforcement" also emphasize the timeliness of monitoring land violations, and the level of monitoring and detection needs to be improved.

In recent years, the application of remote sensing and GIS related technologies has improved the level of land law enforcement supervision. Remote sensing technology combined with target image recognition has been continuously carrying out related research. In the machine algorithm, the sample target learning method has more advantages, such as neural network (NN) (Zhang B, 2018; Liu Y G and You Z, 2003; Tian Z Z et al, 2016), support vector machine (SVM) (Guo M W et al, 2014; Chen L and Qing Q Q, 2006), random forest (RF) (Xiang T et al, 2016) and other algorithms are applied. The machine learning model established by sample training has strong recognition accuracy. In the field of remote sensing, deep learning can achieve better image feature extraction in classification and recognition (Wang B and Fan D L, 2019). Convolutional neural network (CNN) reduces the complexity of the network model and has a good effect in the field of image processing. Therefore, there are many studies such as the use of DCNN to identify water bodies (Zhou W X, 2023), the use of semi-supervised classification technology based on hyperspectral remote sensing

images for accurate classification (Mao Y L, 2023), the classification method of heterogeneous convolutional neural network (HCNN) feature cascade based on high-resolution images, and the improved full convolutional neural network model to study land cover classification (Du J, 2017; Yang X, 2024). Through the analysis, it is found that the research on remote sensing image recognition is mainly applied to the three fields of scene classification, target monitoring, and image retrieval (Tian Z H et al, 2023; Heng X B et al, 2023; Tian Q C et al, 2023), which is more suitable for large-scale monitoring research. For the megacity of Shanghai, with a large urban area and relatively small and scattered cultivated land, the law enforcement requirements are highly granular and effective, and a single regulatory system cannot meet the current needs. With the development of digital twin city construction and AI artificial intelligence technology, the ability to recognize and interpret small image targets has also become stronger, and AI technology has been widely used in urban management, public security, transportation, and other fields for camera information extraction, character recognition, and behavior analysis and recognition (Ye P P, 2023; She Z M, 2023; Hu M, 2023; Zhang X Z, 2022; Xue F C, 2018). In this paper, the monitoring technology route introduces cameras as monitoring sensors for key areas on the ground, uses deep learning methods for test training, and selects the region-based faster regions with CNN feature (Faster-RCNN) algorithm. The algorithm has a good detection effect on targets with incomplete image features (Bao N S et al, 2023), which is suitable for automatic identification of land violations and its application to the construction of cultivated land protection system. Video image recognition, as a supplementary source of illegal clues for image recognition, improves the monitoring level of cultivated land protection.

### 2. Sample Data Source

#### 2.1 Identifying targets

The main purpose of illegal behavior monitoring is to detect abnormal changes in cultivated land as early as possible.

Specific traces of human intervention often appear before the cropland changes. These traces can be further subdivided into specific scenes and recognized objects, which can be automatically identified by the method of object detection in video images.

After analysis, we can broadly classify illegal behavior into two categories. One is a change in the land use properties of cultivated land in an area, which usually requires the operation of mechanical equipment or the construction of temporary construction sites nearby. The other is that the attributes of cultivated land do not change, and other piles occupy part of the area, resulting in a decrease in the actual cultivated land area. Therefore, three categories of identification scenarios and 13 specific objectives are proposed, as shown in Table 1. The start-up warning scenario is for the identification of machinery and equipment, the construction site early warning scenario identifies the characteristics of the construction site, and the groundbreaking early warning scenario is a common type for identifying the occupied cultivated land, as shown in Figure 1.

Table 1 Classification of warning scenarios

Categories	specific targets	Identifying the scene
The start-up warning scenario	Cranes, tower cranes, mixers, drilling machinery, large trucks	The nature of cultivated land has changed, and most of them require mechanical equipment
The construction site early warning scenario	Prefabricated sheds, piles of discarded building materials, fences, buildings under repair	The conversion of cultivated land into a construction site generally has typical characteristics
The groundbreaking early warning scenario	Waste accumulation, ground hardening, rocky stacking, land stacking	The non-cultivated area of cultivated land affects the area of actual cultivated land.



Figure 1. Schematic diagram of the scene (a1 and a2 shows The start-up warning scenario; b1 and b2 shows the construction site early warning scenario; c1 and c2 shows the groundbreaking early warning scenario)

## 2.2 Sample library production

After classifying the early warning scenarios, the land illegal behaviors have been refined into corresponding identification targets, that is, it is clear that the samples collected can be further made into a sample database. In the process of deep learning model calculation, sample library production is an important and complex task. Effective samples can improve the performance and generalization ability of the model, so as to better solve the problem and goal of characteristics.

The sample needs to have obvious features, such as appearance color, target texture, target shape, and other typical features. According to the identification requirements of the three types of early warning scenarios, the mechanical equipment in the groundbreaking early warning is a more typical identification target, and the texture is basically fixed. The samples of construction site early warning and groundbreaking early warning scenes are complex, the types are diverse, and the regularity of recognition features such as texture color is not significant, so the type and number of samples should be enriched when selecting.

The accuracy of sample selection affects the training results. There are a variety of methods for sample generation, including manual annotation, data augmentation, and synthetic data. Manual annotation is the most common and accurate method that manually labels samples in a dataset to generate input data and labels for responses through expert experience, but requires a lot of time and human resources. Data augmentation mainly processes the raw data through rotation, zoom, panning and other operations, which can enhance the diversity of samples, but excessive data augmentation may introduce unnecessary noise to the performance of the image model. Synthetic data algorithms augment datasets by simulating or generating algorithms with computer graphics to create images. Synthetic data is controllable and flexible, but its quality and realism can affect the performance of the model in real-world scenarios.

In order to select accurate and reliable samples, the results of image recognition for many years were used and the verified patches were screened and extracted. Due to the fact that the patches extracted from Shanghai based on images better than 0.5m satellite films, the patches are strictly close to the target object during extraction, so the number of coordinate point groups at the edge of the patches is too large, which is not conducive to the calculation and recognition of the model in the later stage, and the graph needs to be simplified first. In order to facilitate the calculation, the largest external rectangle was taken as the sample recognition frame, and the batch optimized patches were used as the samples.



At the same time, in order to take into account the angle of high-altitude cameras, multi-angle images such as aviation, satellite, and drone cameras were selected, and some scene pictures automatically generated by AI models were added (Figure 2).



Figure 2. Multi-angle and multi-type sampled images (a,b,c and d are selected from aerial imagery, satellite imagery, video camera, and AI model, respectively)

### 3. Deep Learning Methodology

#### 3.1 Faster R-CNN model

Convolutional Neural Network (CNN) is a common deep learning model, which has been widely used in automatic image recognition, target object detection, and other fields. It is to extract the features of an image by using convolutional layers and pooling layers, so as to achieve the classification and recognition of images. With the continuous optimization, the region-based convolutional neural network model has also been proposed.

The Faster R-CNN model (Ren S Q et al, 2017; Wang J C, 2018) is a region-based convolutional neural network model, which is mostly used for object detection. In this model, Region Proposal Network (RPN) is introduced to replace the original selective search (SS) and other methods, which greatly shortens the detection efficiency time (Uijlings et al, 2013).

The main idea of the model is to divide the object detection task into RPN region generation and target classification and bounding box regression (Figure 3). First, the candidate region is generated through the RPN network. RPN is a fully convolutional network that obtains a fixed-size window on the input feature map and outputs multiple proposals for each window position. Specifically, RPN predicts  $k$  anchor boxes at each window position, and uses the size and position of the anchor box to generate candidate boxes, and RPN screens each candidate box by binary classification and regression, binary classification is to clarify the foreground or background, and regression is mainly to adjust the size and position of the box, and finally obtain the final candidate area. Secondly, after

screening out the candidate regions, Faster R-CNN uses the ROI pooling layer to extract the features of each candidate region. ROI pooling maps candidate regions of different sizes into fixed-size feature maps, and transforms target classification and bounding box regression tasks into fixed-size features. Then, the targets are divided into two categories by the fully connected layer and the softmax classifier, and the bounding box is fine-tuned by the regressor, and the final output is the category where the recommended area is located and the exact position of the image. The main advantages of Faster R-CNN are end-to-end training, shared convolutional features, accurate position and size regression, so the detection speed is fast, the accuracy is high, and it is easy to implement and train, which has become one of the mainstream algorithms in the field of object detection.

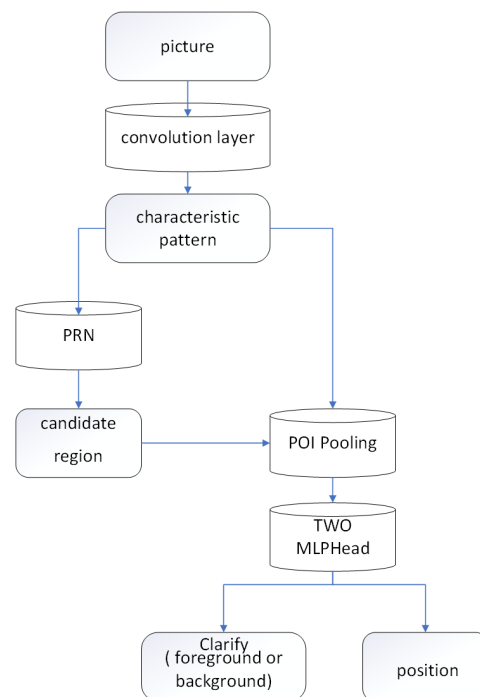


Figure 3. Structure Diagram of Faster R-CNN model

#### 3.2 Model training

The Faster R-CNN model has many parameters, including network structure, anchor scale, learning rate, etc. Trained on the open-source framework Pytorch, tuned to our 3 categories of target audiences and a sample library of multiple data sources. ResNet50 has a deep network, which has a good detection effect as a feature extraction network (Li D Z et al, 2018; Li X J et al, 2024; Zhou S H et al, 2024), which is conducive to the refinement of abstract and deeper target features, and is more suitable for identifying targets with less obvious texture features in the types of construction site warning and scene warning. Considering the cost of running time and overfitting, the number of epochs is not easy to be too much, and the learning rate is mainly set according to experience, too large a value will cause the parameters to be optimized to fluctuate near the minimum value, and too small a value will cause the convergence speed to slow down, and 0.001 is selected for testing according to the general situation. The biggest feature of Faster R-CNN is that RPN training generates an anchor to improve the detection speed. Considering that there are different sizes of indiscriminate occupation and hardening of

identification targets in the target groundbreaking warning scenario, the size of the anchors is set to (32<sup>2</sup>, 64<sup>2</sup>, 128<sup>2</sup>, 256<sup>2</sup>, 512<sup>2</sup>) in order to be compatible with the identification of small targets.

#### 4. Application & Practice

##### 4.1 Target recognition experiments

Based on the sample database of illegal land behavior monitoring, the Faster R-CNN model was used as the basic model, and the learning rate and iteration number were set to identify the targets of construction early warning, construction site early warning and ground breaking early warning. Evaluate the performance of the model on the test dataset, mainly based on the calculation of major metrics such as Precision, Recall, and Mean Average Accuracy (mAP). Precision is the ratio of the number of correct targets detected by the model to the total number of targets on the test picture. Recall is the ratio of the number of correct targets detected by the model to the total number of targets on the test picture. mAP is the area difference enclosed by the two coordinate axes of precision and recall, and then the average value is calculated.

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \int_0^1 P(r)dr \quad (3)$$

where  $TP$  = the number of positive targets identified as positives

$FP$  = the number of negatives targets identified as positives

$FN$  = the number of positive targets identified as negatives

$P$  = Precision

$R$  = Recall

mAP = Mean Average Accuracy

The test point is selected to cover more than 20,000 acres of basic farmland area, the camera deployment height is about 20~40m height, the resolution is 200 dpi, and the test is carried out through video frame extraction. Through experiments, it is found that the three types of target recognition detection are construction early warning > construction site early warning > groundbreaking early warning (Table 2). The early warning is mainly for the identification of mechanical targets, and the model recognition results are relatively good, which can achieve 90% accuracy and 95% recall rate, and the mAP index is 0.85. The construction site early warning and groundbreaking early warning have reached more than 81% after many iterations of training, and the construction site early warning is mainly for the identification of construction site characteristic objects such as sheds, with an accuracy rate of 84%, a recall rate of 90%, an mAP index of 0.8, an accuracy rate of 81% and a recall rate of 88%, and an mAP index of 0.75.

Table 2. Table of training results

Categories	Precision	Recall	mAP
The start-up warning scenario	90%	95%	0.85
the construction site early warning scenario	84%	90%	0.8
the groundbreaking early warning scenario	81%	88%	0.75

On the whole, the Faster R-CNN model has good target detection efficiency and certain effectiveness in three aspects: construction early warning, construction site early warning, and ground breaking early warning (Figure 4). However, there are some differences in the previous identification of different categories. The early warning effect of construction is the highest, which may be that the characteristics of mechanical targets are more obvious, the contour colors are more fixed, and the detection targets are more specific. For example, the outline shape of fences and piles of waste is not fixed, the texture is different under the influence of light, and the background interference is large, so the relative recognition is poor.



Figure 4. Diagram of identifying monitoring points

##### 4.2 Cultivated land protection applications

At present, the supervision of cultivated land protection is mainly based on law enforcement and manual inspection, which cannot meet the regulatory needs of high timeliness response. By deploying high-altitude cameras in areas where farmland is concentrated, video images of farmland can be obtained. Then we obtain video information in real time through video transmission, and regularly extract frames as pictures as needed, and identify the behaviors of land illegal construction early warning, construction site early warning and ground breaking early warning through the trained model, which can effectively identify different types of illegal behaviors.

Through the target recognition of the model, a real-time supervision platform has been established, as shown in the Figure 5, the model is identified and immediately pushed through spatial positioning, and after the business system and the business approval data are approved and compared, the suspected illegal patches are sent to the inspectors, and the inspectors will feedback the inspection situation and disposal information to the platform after the on-site inspection, and the feedback inspection results will be regularly supplemented to the sample library to train and improve the model algorithm.

On the one hand, the platform monitors the farmland area in real time, and automatically identifies the relevant behaviors of land violations through the trained Faster R-CNN model, which can quickly identify suspected illegal behaviors, which has a certain degree of real-time and effectiveness, and improves work efficiency.

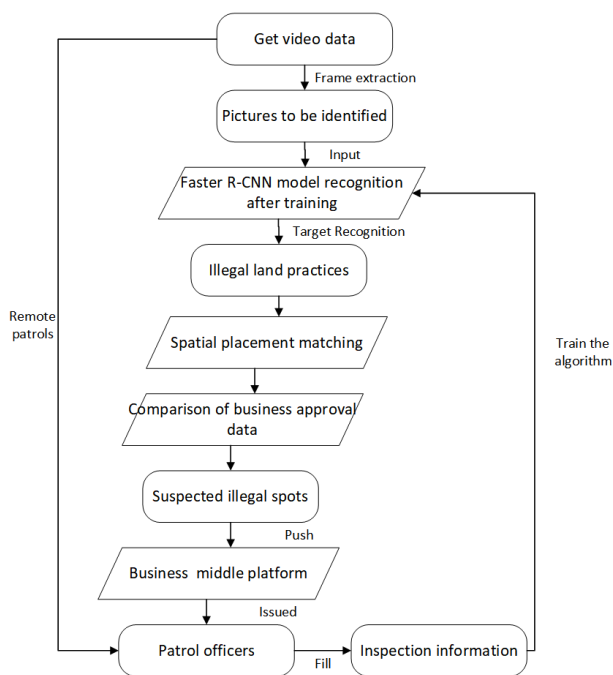


Figure 5. Design ideas of the real-time supervision platform

## 5. Conclusion

In order to better protect cultivated land, this paper studies the classification of early warning scenarios of land violations, and proposes three types of scenarios: early warning of construction, early warning of construction site, and early warning of ground breaking. For each scenario, we performed detailed target recognition and selected the corresponding samples to make a training sample library. Then, we used the Faster R-CNN model to train the samples and form a training model.

Through experiments, we found that the accuracy of the three types of scene recognition was more than 81%, which had a certain effectiveness, and the effect of groundbreaking early warning was the best, reaching 90%. Therefore, we apply this model to the application management of cultivated land protection, and improve the existing level of land law enforcement management by establishing a real-time supervision platform.

In the next study, we will further improve the sample bank and model training by analyzing the inspection results of the patches after the early warning is pushed. We will continue to optimize the model to improve the recognition rate of construction site warning and ground breaking warning.

## References

- Zhang B, 2018: Remotely Sensed Big Data Era and Intelligent Information Extraction. *Geomatics and Information Science of Wuhan University*, 43(12):1861-1871. DOI : 10.13203 / j.whugis20180172.
- Liu Y G, You Z S, 2003: An neural network for image object recognition and its application to car type recognition. *Computer Engineering*, 29(3):30-32
- Tian Z Z, Zhan R H, Hu J M, et al. SAR ATR based on convolutional neural network. *Journal of Radars*, 5(3): 320-325.
- Guo M W, Zhao Y Z, Xiang J P, et al, 2014: Review of object detection methods based on SVM. *Control and Decision*, (2):193-200
- Chen L, Qing Q Q, 2006: Samples selection for RS object detection based on SVM. *Computer Engineering and Applications*, 42(9):212-214.
- Xiang T, Li T, Zhao X Z, et al, 2016: Random forests for accurate object detection. *Application Research of Computers*, 33(9):2837-2840
- Wang B, Fan D L, 2019: A review of the research progress of deep learning in remote sensing image classification and recognition. *Mapping Bulletin*, (2) : 99-102. DOI : 10.13474 / j.cnki.11-2246.2019.0052.
- ZHOU W X, 2023: Research on remote sensing image retrieval based on deep learning features. *Acta Geodaetica et Cartographica Sinica*, 52(1): 167. DOI: 10.11947 / j.AGCS. 2023. 20210106.
- Mao Y L, 2023: Comparative study of target detection algorithms based on different convolutional neural networks. *Standardization of Surveying and Mapping*, 39(04) : 39-43. DOI : 10.20007 / j.cnki.61-1275 / P.2023.04.08.
- Du J, 2017: UAV remote sensing image water body recognition based on deep learning. *Jiangxi Science*, 35(01) : 158-161 + 170. DOI : 10.13990 / j.issn1001-3679.2017.01.03. DOI : 10.13990 / j.issn1001-3679.2017.01.03.
- Yang X, Fang L Y, Yue J, 2024: Advances in semi-supervised classification of hyperspectral remote sensing images. *Journal of Remote Sensing*, 28 (01) : 20-41.
- Tian Z H, Chang P, He X H, et al, 2023: A land cover classification of high resolution remote sensing image based on CNN-GCN. *Surveying and Mapping Science*, 48(06) : 59-72. DOI : 10.16251 / j.cnki.1009-2307.2023.06.008.
- Heng X B, Xu F, Tang L, et al, 2023: Research on land cover classification method based on improved fully convolutional neural network model. *Journal of Earth Information Science*, 25(03) : 495-509.

Tian Q C, Wu S Y, Ma Y N, 2023: An overview of optical remote sensing image analysis based on convolutional neural networks. *Computer applications and software*, 40(10) : 1-9 + 45.

Ye P P, 2023: Research on automatic extraction technology of remote sensing image natural resource information based on deep learning. *Surveying and mapping technology and equipment*, 25(01) : 36-40. DOI : 10.20006 / j.cnki.61-1363 / P.2023.01.007.

She Z M, Shen Y Z, Song J H, et al, 2023: Using the classical CNN network method to construct the automatic extraction model of remote sensing image of road elements in Guiyang. *Surveying and mapping bulletin*, 2023(04) : 177-182. DOI : 10.13474 / j.cnki.11-2246.2023.0126.

Hu M, 2023: Research and Application of Key Technologies of Highway Traffic Perception Based on AI Machine Vision. *Traffic Technology and Management*, 4 (08) : 7-9.

Zhang X Z, Zhao S Z, Lin Y, 2022: Construction and application of intelligent policing under the background of big data. *Journal of China People 's Public Security University ( Natural Science Edition )*, 28(02) : 70-81.

Xue F C, Song X Y, Tang B X et al, 2018: Video Surveillance of Urban Waterlogging Monitoring and Early Warning. *Surveying and Mapping Science*, 43(08) : 50-55 + 61. DOI : 10.16251 / j.cnki.1009-2307.2018.08.008.

Bao N S, Han Z S, Yu J X, et al, 2023: Remote sensing identification method of open-pit coal mining area based on improved Faster-RCNN. *Journal of Northeastern University ( Natural Science Edition )*, 44 (12) : 1759-1768.

Ren S Q, He K M, Girshick Ross, Sun J, 2017: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE transactions on pattern analysis and machine intelligence*, 2017.

Wang J C, Tan X C, Wang Z H, et al, 2018: Faster R-CNN deep learning network based object recognition of remote sensing image. *Journal of Geo-information Science*, 20(10):1500-1508. DOI:10.12082/dqxkx.2018.180237

Uijlings J R, Sande K E, Gevers T, et al, 2013: Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154-171.

Li D Z, Fan D Z, Su Y L, 2018: Building detection of remote sensing images combined with Faster R-CNN model. *Journal of Surveying and Mapping Science and Technology*, 35 (04) : 389-394.

Li X J, Quan L F, Liu D M, et al, 2024: Improved traffic sign detection algorithm based on Faster-RCNN. *Journal of Jilin University (Engineering Edition)*, (03): 1-10. [ 2024-03-27 ]. <https://doi.org/10.13229/j.cnki.jdxbgxb.20230553>.

Zhou S H, Fang X J, Liu X Y, et al, 2024: Aircraft target detection based on transfer learning and improved Faster-RCNN remote sensing image. *Electromechanical engineering technology*, (03): 1-8. [ 2024-03-26 ]. <http://kns.cnki.net/kcms/detail/44.1522.TH.20240222.1433.002.html>.