

Article

Scene Recognition for Construction Projects Based on the Combination Detection of Detailed Ground Objects

Jian Pu ^{1,2} , Zhigang Wang ^{1,2,*}, Renyu Liu ³, Wensheng Xu ^{1,2}, Shengyu Shen ^{1,2}, Tong Zhang ³  and Jigen Liu ^{1,2}

¹ Changjiang River Scientific Research Institute of Changjiang Water Resources Commission, Wuhan 430010, China

² Research Center on Mountain Torrent & Geologic Disaster Prevention of the Ministry of Water Resources, Wuhan 430010, China

³ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China

* Correspondence: wangzg@mail.crsri.cn; Tel.: +86-150-7240-0980

Abstract: The automatic identification of construction projects, which can be considered as complex scenes, is a technical challenge for the supervision of soil and water conservation in urban areas. Construction projects in high-resolution remote sensing images have no unified semantic definition, thereby exhibiting significant differences in image features. This paper proposes an identification method for construction projects based on the detection of detailed ground objects, which construction projects comprise, including movable slab houses, buildings under construction, dust screens, and bare soil (rock). To create the training data set, we select highly informative detailed ground objects from high-resolution remote sensing images. Then, the Faster RCNN (region-based convolutional neural network) algorithm is used to detect construction projects and the highly informative detailed ground objects separately. The merging of detection boxes and the correction of detailed ground object combinations are used to jointly improve the confidence of construction project detection results. The empirical experiments show that the accuracy evaluation indicators of this method on a data set of Wuhan construction projects outperform other comparative methods, and its AP value and F1 score reached 0.773 and 0.417, respectively. The proposed method can achieve satisfactory identification results for construction projects with complex scenes, and can be applied to the comprehensive supervision of soil and water conservation in construction projects.

Keywords: construction projects; remote sensing image; detailed ground objects; soil and water conservation



Citation: Pu, J.; Wang, Z.; Liu, R.; Xu, W.; Shen, S.; Zhang, T.; Liu, J. Scene Recognition for Construction Projects Based on the Combination Detection of Detailed Ground Objects. *Appl. Sci.* **2023**, *13*, 2578. <https://doi.org/10.3390/app13042578>

Academic Editor:
Edyta Plebankiewicz

Received: 13 January 2023
Revised: 6 February 2023
Accepted: 9 February 2023
Published: 16 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, along with the rapid pace of urbanization, built-up areas are developing rapidly [1]. During urbanization development, human disturbances, mainly construction activities, intensifying environmental degradation [2], as well as water and soil loss, call for immediate action to strengthen its supervision [3]. Construction projects often come in various types, featuring scattered points, long lines, wide areas, and rapid changes. Traditional manual detection methods based on high-resolution remote sensing images can hardly meet the precise supervision requirements of full coverage and multi-temporal phases [3–5]. Since high-resolution remote sensing images have rich ground object information and fewer spectral bands, remote sensing image classification methods that are based on pixel spectral feature statistics and are “object-oriented” are prone to the “salt and pepper effect”, “different objects with same spectrum” and “same object with different spectrum”, posing greater difficulties for high-resolution image analysis and processing [6,7]. How to process high-resolution images rapidly, efficiently and automatically has become a hot research topic and challenge in the field of remote sensing.

With the development of computer vision technologies and remote sensing image interpretation methods, significant progress has been made in target detection and semantic

segmentation based on high-resolution optical remote sensing images [8,9]. The target detection of buildings, planes, vehicles, ships and other objects has achieved high accuracy, mainly because the semantic definition of these objects is relatively clear, their boundaries are well-defined, and their diversity is relatively limited. The construction projects in high-resolution remote sensing images have no unified semantic definition and their scenes include many man-made and natural ground objects with highly non-structured features and significant differences in image features [10–12]. At present, the identification of construction projects in high-resolution remote sensing images mainly adopts manual interpretation and drawing, but this presents the problem of a large workload of manual interpretation [13]. The proposed object-oriented identification method that selects the optimal segmentation scale also faces the problem of different identification results under different remote sensing image sources and different surface coverage characteristics [14]. However, the existing construction project target detection methods cannot be directly applied to the detection of complex semantic scenes, and the generalization ability of the methods that directly train the detector cannot meet the requirements of accurate extraction [15]. The disadvantage of the previous work is that the semantics of construction projects are very complex, and the research on detection methods for complex scenes is still in its early stage, making it difficult to cope with the exceptionally complex urban scenes of construction projects.

To solve these problems, the goal of this study is to propose a target detection method for construction projects in complex semantic scenes, which can automatically and accurately identify construction projects in high-resolution remote sensing images. Compared with the traditional target detection methods, this method can improve the identification results by merging the candidate boxes, training the highly informative detailed ground objects and adjusting the confidence of the construction project candidate box, thus improving the detection results. This method can improve the detection performance for construction projects to a certain extent without significantly raising the complexity of training. Furthermore, it can identify the construction period of the construction project (such as topping out or completion) and the implementation of soil and water conservation measures (such as whether the spoil and slag are covered with dust screens) by the features of detailed ground objects.

2. Materials and Methods

2.1. Technical Process

This method was inspired by semantic event detection in video data in the field of computer vision. It has always been challenging to detect semantic events with complex definitions (such as weddings, gatherings, etc.) from videos, which are characterized by unclear definitions, large intra-class differences, and the inclusion of complex visual features [16–20]. Studies found that the events in videos can be composed of more specific underlying concepts; for example, wedding events can be composed of wedding dresses, cakes, MCs, and other concepts that are easy to detect by target detection [21–24]. Thus, this method corresponded the complex events of construction projects to the complex events in videos, and the various types of detailed ground objects in the scene to the underlying concepts in the video. It further used the target detection method and the combination of detailed ground objects in the complex scene to improve detection results.

This process is divided into the following steps: (1) Construction of sample data set; (2) Selection and expression of detailed ground objects; (3) Target detection of construction projects and detailed ground objects; (4) Post-processing. The flow chart is shown in Figure 1, with the red boxes highlighting the innovation points of this paper.

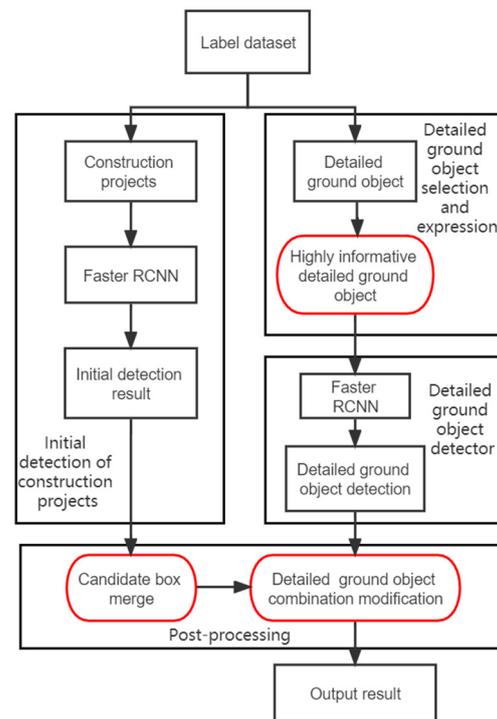


Figure 1. The method flow chart.

2.1.1. Selection and Expression of Detailed Ground Objects

For the complex semantic scenes of a construction project, it is necessary to find out which detailed ground object can enhance its detection, that is, to find the type of detailed ground object that can best represent the construction project. The selection is made based on expert experience and the area and quantity of various detailed ground objects in the construction project sample.

It is assumed that there are some types of common detailed ground objects in the area where the sample is located. However, with the large number of types of detailed ground objects, and limited information provided by some objects, the use of low-informative detailed ground objects will make it even more complex. Therefore, it is necessary to select certain types of detailed ground objects which can provide the corresponding amount of information for identifying complex semantic scenes of construction projects. The information amount of the i th detailed ground object can be calculated by the following equation:

$$C_i = N_i * w_i, \tag{1}$$

where N_i is the amount of detailed ground objects in all sample boxes of the construction project. Then, w_i can be calculated according to the following equation of term frequency-inverse document frequency (TF-IDF) [25,26]:

$$w_i = TF_i * IDF_i = \frac{N_i}{N} * \lg \frac{Y + 1}{Y_i}, \tag{2}$$

where TF_i is the frequency of occurrence for the i th detailed ground object, and its value is the ratio of the number of its occurrence in the construction project label box to the number of its occurrence in all labeled boxes. The inverse document frequency, IDF_i , is the ratio of the average area of construction projects in the sample area, Y , to the average area of the i th detailed ground object in the sample area, Y_i (taking logarithm for calculation). (Note: In Equation (2), the denominator of $\lg \frac{Y+1}{Y_i}$ is taken as $Y + 1$ to ensure that the calculated IDF_i will be greater than 0, and the effect on the actual result can be neglected.)

2.1.2. Selection and Expression of Detailed Ground Objects

Both detectors of this method used the classic two-stage target detection algorithm Faster RCNN [27–30], with its network schematic diagram shown in Figure 2. The algorithm is mainly composed of four parts: (1) The feature extraction part uses the classical convolutional neural network, that is, to obtain the feature map of the input image through several convolutional layer-activation function-pooling layer units as the basis of the target detection network. (2) Region proposal network, a network structure for generating candidate boxes, has the same function as the traditional algorithm in generating candidate boxes. The only difference is that it adopts the method of deep learning. (3) ROI pooling maps the candidate box generated in the region proposal network to the feature map of the image, and combines the relevant information to obtain the feature map of the corresponding candidate box. (4) Classification and regression network use the feature map of the candidate box to classify the candidate frame and conduct bounding regression to obtain the detection result.

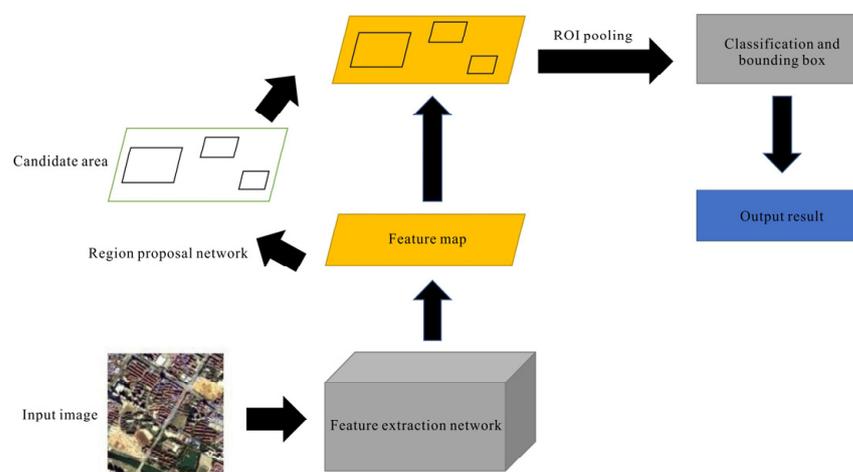


Figure 2. Schematic diagram of Faster RCNN network.

2.1.3. Post-Processing

The post-processing is conducted based on the construction project detection results obtained by the detector. The visual differences within the construction project are huge, with inconsistent coverage and no clear and well-defined boundaries, making it difficult to obtain satisfactory detection results by directly using conventional deep learning target detection methods such as Faster RCNN. However, when the visual differences in the detailed ground object samples are small, with basically the same coverage, faster RCNN can be used to obtain satisfactory detection results. Therefore, it is necessary to improve the detection results based on the preliminary detection results, as well as the characteristics of the construction project and the detection results of detailed ground objects. The post-processing in this paper adopted two steps: (1) merging the candidate boxes of the construction project; and (2) modifying the detailed ground object combination.

(1) Merging construction project candidate boxes. The preliminary detection results obtained by the construction project detector are multiple candidate boxes (hereinafter referred to as candidate boxes) and their corresponding confidence. Taking a construction project in Wuhan in Figure 3 as an example, based on the cognitive differences of different experts, the image can be identified as one construction project (a green box) or two construction projects (two red boxes). This difference in cognition will cause a large error in the target detection of construction projects, so this paper proposes a strategy of merging candidate boxes to mitigate the error.



Figure 3. Examples of construction projects.

In the construction project sample labeling, the green box in the above figure indicates the real situation on the ground, and the red candidate boxes are the detection result of the detector. The red candidate boxes overlap, so the detection results obtained by the construction project scene detector can be merged to match the ground truth (the green box). The specific implementation method is as follows: if the construction project scene detector detects that the candidate boxes A and B overlap, and the ratio of the overlapped area of A and B to the minimum area of the two is greater than the merging threshold α , and the confidence of both A and B is greater than the combined confidence threshold β , then A and B will be merged into the candidate box C, the area of the candidate box C will be the smallest bounding rectangle of A and B, and the confidence of the combined candidate box will be:

$$Conf_t = (Conf_1 * S_1 + Conf_2 * S_2) / S_3, \quad (3)$$

where S_1 , S_2 , and S_3 are the areas of A, B and the merged candidate box C obtained from the construction project scene detection, and $Conf_1$, $Conf_2$, and $Conf_t$ are the confidence of A, B, and C, respectively.

(2) Detailed ground object combination modification. The difficulty of the detection of construction project scenes is reflected in the difficulty of identifying the range of candidate boxes, as well as the fact that the confidence of the detected candidate boxes is not accurate enough. The detailed ground objects have similar characteristics to general target detection categories, such as single features and clear boundaries. Therefore, the detailed ground object candidate box obtained by the detailed ground object detector (hereinafter referred to as the detailed ground object box) was used to assist the construction project scene detector, and the overall visual features of the construction project and the detailed ground objects inside were combined to obtain the comprehensive expression confidence of the construction project. The confidence calculation equation of the candidate box after the combination is as follows:

$$Conf = \min(1, \gamma * Conf_t + (1 - \gamma) * \sum_{i=1}^N (Conf_i * IoP_i)), \quad (4)$$

where γ is a hyperparameter used to adjust the weight of the construction project scene detector and the detailed ground object detector. N is the number of detailed ground objects in the candidate box obtained by the construction project scene detector. $Conf_i$ is the confidence of the i th detailed ground object box. IoP_i (intersection over proposal) is the

ratio of the intersected area of the i th detailed ground object box and candidate box to the area of the candidate box, and is defined as follows:

$$IoP_i = \frac{\text{area}(B \cap B_i)}{\text{area}(B)}, \quad (5)$$

B and B_i represent the candidate box and the i th detailed ground object box, respectively.

2.2. Experiment Data

2.2.1. Sample Data Set

The role of the sample data set is to provide training samples for the construction project detector and detail ground object detector. Therefore, it is necessary to label the construction project with the smallest bounding box, and use the smallest bounding box to label various detailed ground objects in the construction project. This data set is manually labeled. Based on the collected scope of responsibility in the soil and water conservation plan for construction projects in Wuhan and the overall layout of the project, the construction project and its six detailed ground objects (bare soil (rock), dust screen, construction roads, moveable slab houses, buildings under construction and constructed buildings) were selected for labeling as shown in Figure 4. The labeling of this data set is based on the Gaofen-1 satellite remote sensing image with a resolution of 2 m, RGB format, a size of 600×600 Pixels, and labellmg as the tool (<https://github.com/tzutalin/labellmg>) (accessed on 2 June 2020).

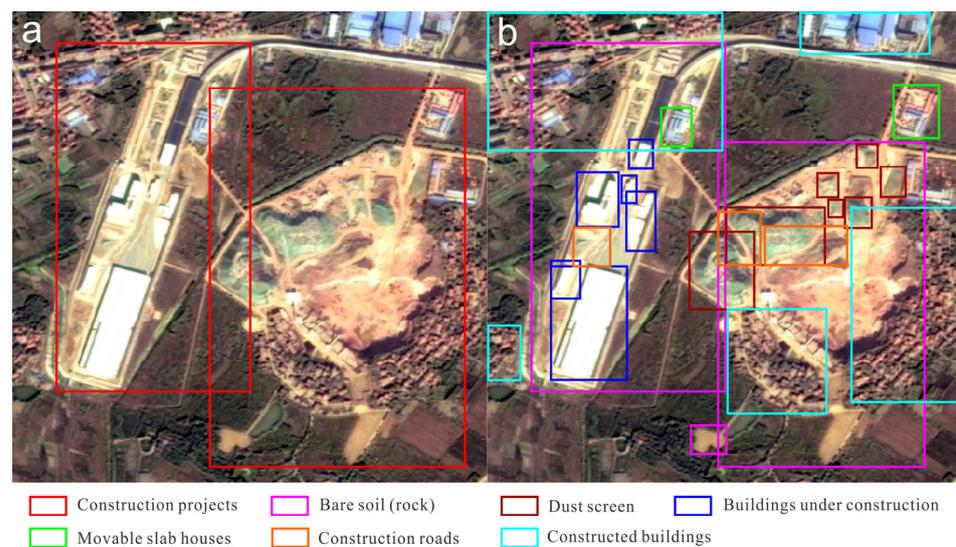


Figure 4. Construction projects in (a) and detailed ground objects in (b) labeling examples.

The number of labeled construction projects, bare soil (rock), dust screens, construction roads, movable slab houses, buildings under construction, and constructed buildings is, respectively, 752, 763, 154, 82, 372, 292, and 278.

2.2.2. Experimental Conditions

In the experiment of this paper, the hardware parameters and the software development environment configuration used in training and testing are shown below: the CPU is Intel(R) Xeon(R) CPU E5-2665, the GPU is GeForce RTX 2080 Ti, the memory is DDR4 10 G, the operation system is Ubuntu 18.04, the programming language is Python 3.6, the deep learning framework is Pytorch 1.0, the operation platforms are CUDA10 and cuDNN7.5.0.

2.2.3. Experimental Setting and Evaluation Indicators

In this experiment, based on expert experience, the six types of detailed ground objects of bare soil (rock), dust screen, construction roads, movable slab houses, buildings under

construction and constructed buildings were preliminarily selected. According to the selection and expression methods of detailed ground objects, the amount of information about the six types of detailed ground objects was 18.81, 20.90, 9.93, 44.82, 28.77 and 8.22, respectively.

It is theoretically possible to take all detailed ground objects into account and train all detailed ground object detectors. However, some detailed ground objects have a limited amount of information, producing minor improvement effects for the detection of the construction project and causing increased complexity. Thus, when determining the best detailed ground objects, it is necessary to select the highly informative ones. In this study, four types of detailed ground objects of movable slab houses, buildings under construction, dust screen, and bare soil (rock) were selected to characterize the construction project, and the detectors for these four detailed ground objects were trained.

In this experiment, both detectors randomly selected 60% of the images in the data set as the training set, 20% of the images as the verification set, and another 20% of the images as the test set. The relevant parameters of the Faster RCNN network are as follows: the initial learning rate is set at 0.001 and drops to 1/10 of the original every 10 rounds of iteration, and the number of training traversals is set at 50. The pre-trained ResNet101 is used as the network structure, the gradient optimization algorithm adopts SGD, the momentum is set to 0.9, and the attenuation coefficient is 0.0005.

The performance evaluation indicator of this experiment adopts the F1 score, precision-recall curve (PR curve), and single category average precision (AP). The F1 score, also known as the balance F score, is the reconciled average of precision and recall rate. The AP is the area formed by the PR curve, the x and y axis. The PR curve can evaluate the overall situation of the detection results of the construction project, and the AP can objectively evaluate the average quality of the detection of the construction project to a certain extent. The larger the value, the better the target detection model will be.

In remote sensing object target detection, in general, *IoU* (intersection over union) is used as an indicator to evaluate the correctness of the detection of a candidate box. However, due to the particularity of the construction project, there may exist cases where a single candidate box corresponds to multiple ground truths, and $IoU > 0.5$ is not suitable as the indicator to evaluate the correctness of the candidate boxes. Therefore, based on expert knowledge, a one-to-many approach was adopted, where a single candidate box corresponds to multiple ground truths intersecting with it, that is, satisfying $\sum_i^N IoU > 0.5$. The equation is:

$$\sum_i^N IoU = \sum_i^N \frac{area(D \cap D_i)}{area(D \cup D_i)}, \quad (6)$$

where D denotes the candidate box obtained for the detector, and D_i is the i th ground truth that intersects with the detector.

3. Results

This experiment was conducted on the same training set and test set, using Faster RCNN to train the construction project detector, and perform post-processing. As mentioned in Section 2.1.3, three hyperparameters α , β , and γ are involved in post-processing. In the experiment where Faster RCNN served as the control group, β was set at 0.005 by the original author. If the confidence level was lower than this value, the detection results of the candidate box would be considered untrustworthy. Therefore, in our experiment, β was also fixed at 0.005, so as to make a better comparison with the control group experiment. According to the step size of 0.25, the superparameters α and β were searched in the range of 0 to 1, and the experimental results were obtained as follows:

As can be seen from the table, the experimental performance is the best when $\alpha = 0.25$, and the highest AP value is obtained when $\gamma = 0.75$. This group of experiments is the method mentioned below.

Three sets of experimental comparisons were also made, using Faster RCNN and Yolo v5x (ref.) [31], to directly train the construction project detector (hereinafter referred to as

Faster RCNN and Yolo), and construction project candidate box merge based on Faster RCNN. (This set of experiments is the results of $\alpha = 0.25$, $\gamma = 1$ in Table 1, hereinafter referred to as the variant.)

Table 1. AP values of the experiment under different hyperparameters α and γ .

	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
$\gamma = 0$	0.527	0.694	0.677	0.649	0.599
$\gamma = 0.25$	0.512	0.730	0.701	0.651	0.624
$\gamma = 0.5$	0.497	0.728	0.707	0.699	0.618
$\gamma = 0.75$	0.551	0.773	0.759	0.707	0.697
$\gamma = 1$	0.523	0.754	0.721	0.672	0.672

Table 2 shows the comparison results of the AP and F1 score of the four experiments. The change curve of loss in the training process of the four groups of experiments is shown in Figure 5. The PR curves of the four experiments are shown in Figure 6.

Table 2. AP and F1 score results of the four experiments.

Category	Faster RCNN	Yolo	Variation of Our Method	Our Method
AP	0.755	0.693	0.754	0.773
F1 score	0.415	0.361	0.405	0.417

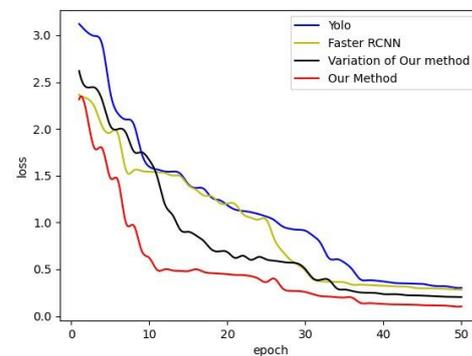


Figure 5. Loss curves of the four experiments.

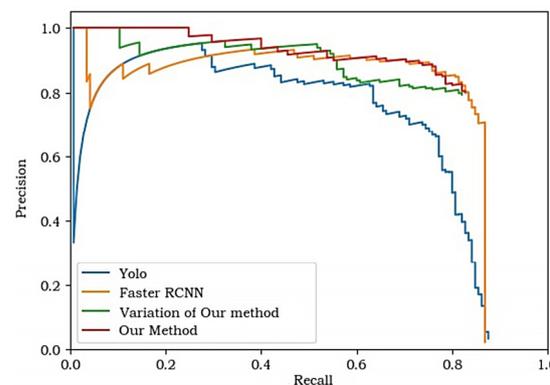


Figure 6. PR curves of the four experiments.

As can be seen from Table 2, the AP value and F1 score of our method are higher than those of the other three experiments, indicating that this method has the best detection performance. The PR curve shows that the overall performance of Yolo is lower than the other groups, and in the low recall stage (or high precision stage, i.e., the left side of

the curve), the curve of our method is significantly higher than the other three groups. This is also the most important improvement effect of this method compared with other experimental groups. By introducing the confidence of the detailed ground object detector, the confidence of the candidate box of the construction project is improved, and the false detection rate in the low recall rate stage is reduced, thus improving the detection results to a large extent.

4. Discussion

4.1. The Improvement Effect of Detailed Ground Objects on Detection and Its Limitations

We used the typical target detectors Yolo and Faster RCNN as control experiments. Our method, based on the Faster RCNN algorithm, introduced the combined detection optimization method of construction project candidate box merge and detailed ground object combination modification. This paper classified construction projects according to their size and construction cycle, compared the detection results of this method and the Faster RCNN algorithm, and discussed the differences in detection results and the effect of detailed ground.

The detection performance of this method in the low recall rate stage (left side of the PR curve) is significantly better than that of the Faster RCNN algorithm, that is, the false detection rate of this method at this stage is lower than that of Faster RCNN, but in the high recall rate stage (right side of the PR curve), the difference in detection performance is not significant.

Figure 7 shows some of the detection results of the two experiments. The first two sets of detection results show that, due to the unclear boundaries of the construction project, although there is only one ground truth in the image, the Faster RCNN algorithm detected two high-scoring candidate boxes, resulting in poor results for both candidate boxes. The construction project candidate box merge can better address the feature of unclear construction project boundaries, making the detection results better match the ground truth, and lowering the false detection rate. Detailed ground objects usually have clear features and well-defined boundaries, so directly using Faster RCNN for detection will yield satisfactory detection results. When there is a certain detailed ground object box in the candidate box, the confidence of the candidate box will be improved, otherwise, it will be lowered, which is also the core idea of detailed ground feature combination modification in this method. In the third set of detection results, there exists false detection, and the confidence of the false detection box is high. Since there is no detailed ground object detected by the detailed ground object detector near the false detection box, it can be easily known from Equation (4) that the confidence of this candidate box in this method is lower than that in the Faster RCNN algorithm. The results of this set further show that for false detection boxes, the confidence of this method for such candidate boxes is lower than that of the Faster RCNN algorithm. Although this step cannot completely eliminate false detection results, it can limit the confidence of the false detection box and improve the confidence of the correct detection box to a certain extent, thus improving the detection performance as a whole.

The introduction of detailed ground object detection is not only conducive to improving the confidence of conventional deep learning target detection methods, but also identifying the construction cycle and the implementation of soil and water conservation measures. In the first detection group in Figure 7 there is a large piece of bare soil (rock) in the construction project that can be detected through detailed ground object detection. Based on this, it can be judged that the construction project is in the stage of project leveling at the initial stage of construction. When detailed ground objects are detected in buildings under construction, the characteristics and completion of the building can be used to determine the type of construction project and whether the main body of the project is completed. At the same time, it can also judge whether soil and water conservation measures have been taken by the existence of the detailed ground object of a dust screen in the construction project.

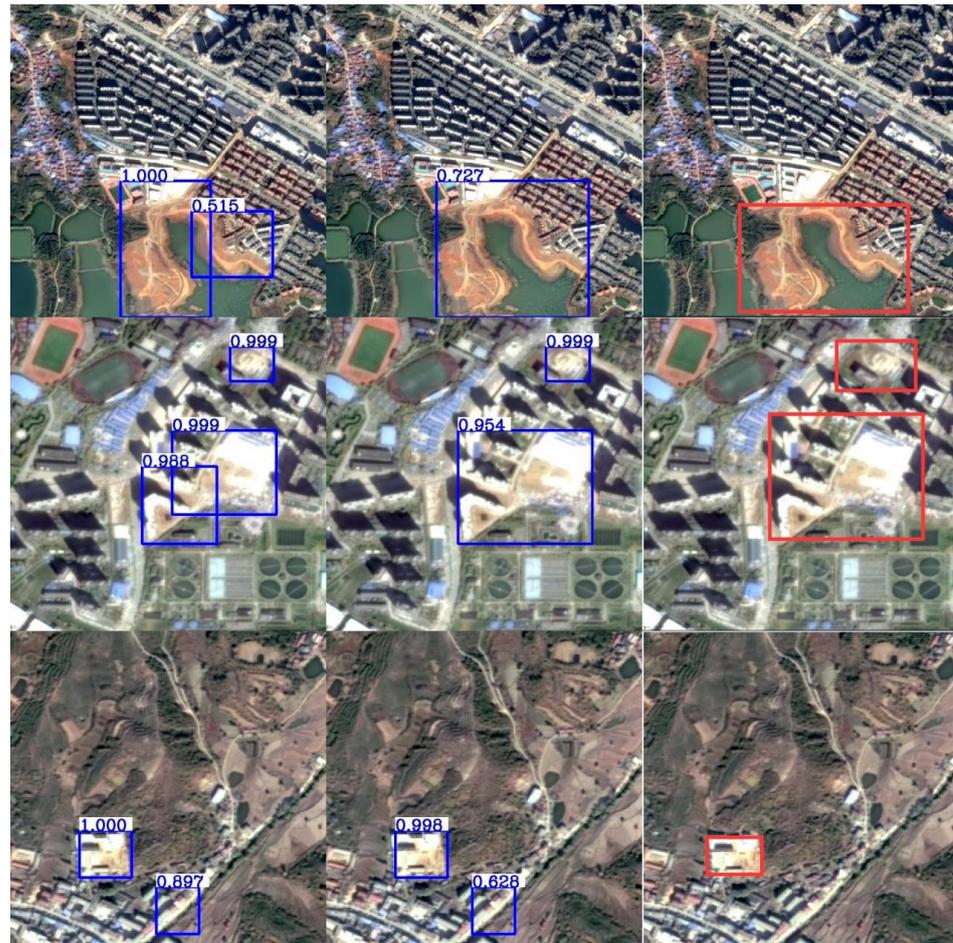


Figure 7. Detection results of the Faster RCNN method and this method. Each row in the figure is from the same image as a group for detection. From left to right, the purple candidate boxes in the first and second columns are detection results of Faster RCNN and this method respectively, and the red boxes in the third column are ground truth.

In [22], the author proposed to learn a higher-level semantic descriptor named a “concept bundle” to facilitate a video search for complex queries; mAP@1000 increased 0.015 and 0.013, respectively, over the “TV08” and “YouTube” data sets compared to not using the “concept bundle”. In our paper, it is found that adding detailed ground objects can improve the detection effect to a certain extent, but it is limited. Further research is needed on how to improve the detection effect more effectively with several types of detailed ground objects. As can be seen from Table 1, the selection of hyperparameters will largely detect the effect, which may vary in different data sets, and finding the appropriate hyperparameters requires a traversal search, which will take time. More detailed parameter analysis will be discussed in Section 4.3.

4.2. Comparison of Detection Results of Different Construction Project Types

One of the main characteristics of construction projects is large intra-class differences and no unified visual feature information. The main reasons for the large intra-class difference are the large difference in the area of construction projects, the long construction cycle, and significant visual differences during different stages. For example, in the early stage, the visual feature of the construction project tends to be bare soil (rock), while in the later stage, it tends to be more architectural. Therefore, the construction projects are classified according to their area size and construction cycle, and the detection performance of the Faster RCNN algorithm and this method under different categories is analyzed. This is essential for the understanding of detailed ground detectors, and even the adjustment

effects of different detailed ground objects on the confidence of construction project candidate boxes. Therefore, according to the size of the construction project, and taking the average area of the construction projects in the data set as the boundary, the construction projects are classified as those larger than average area and those smaller than average area. They are also divided into those in the early stage and those in the later stage according to the construction cycle and based on the standard that no buildings under construction exist in the early stage, and these buildings do exist in the later stage. Based on these two classification methods, the *IoU* value of each type of ground truth and its corresponding detection box is calculated, and the results are shown in Tables 2 and 3.

Table 3. *IoU* of the Two Experiments Classified by Area of Construction Projects.

Method	Less than Average Area	Larger than Average Area
Faster RCNN	0.803	0.825
Our method	0.819	0.857

It can be seen from Tables 3 and 4 that for both classification methods, the *IoU* value of this method is higher than that of the Faster RCNN algorithm. In the classification according to the area of construction projects, the improvement of the *IoU* value of this method was even greater in the “larger than average area” category compared with that of the Faster RCNN algorithm. That is, in the ground truth with a large area, this method can realize the greatest improvement in the detection results by the candidate box merging strategy, indicating that candidate box merging mostly exists in the candidate box corresponding to the ground truth with a large area. In the classification according to the construction cycle, it is found that this method has no huge difference in the improvement of the two categories, indicating that the difference in the construction cycle has no major effect on the improvement of the *IoU* value of this method.

Table 4. *IoU* of the Two Experiments Classified by Cycle of Construction Projects.

Method	Early Stage	Later Stage
Faster RCNN	0.820	0.814
Our method	0.835	0.826

4.3. Parameter Analysis

In this method, there are mainly three parameters involved, namely the merging threshold α and confidence threshold β in Equation (3), and the weight parameter γ in Equation (4). Using a grid search, the optimal solution for the three parameters is obtained in the validation set. The merging threshold α is used to determine the merging conditions. When α is too high, the conditions for candidate box merging will be harsh. While, when α is too low, only a few intersected candidate boxes will be merged, which will lower the detection performance. It can be found in Table 1 that when $\alpha = 0$, compared with other experiments, the AP value of the model is very low and the detection effect is poor. In Figure 8, the two candidate boxes corresponding to the two ground truths will be merged into one, creating a false candidate box.

Due to its unique characteristics, the Faster RCNN algorithm will produce a large number of low-scoring candidate boxes, and the sharp drop on the far-right end of its PR curve in the PR curve image is just the result of this. Therefore, candidate boxes with higher confidence must be selected for merging, and the confidence threshold β serves as the judgment condition, that is, when the two candidate boxes meet the condition that the intersected area is greater than the merging threshold α , the confidence of both of them also needs to be greater than the confidence threshold β , then, the merging mechanism can be triggered. In order to ensure the consistency of parameters with the Faster RCNN control group, $\beta = 0.005$ was fixed.



Figure 8. Faster RCNN Detection Instance.

The weight parameter γ of Equation (4) is used to determine the weight of the detailed ground object, with its value range set as 0–1. When $\gamma = 1$, the method in this paper will be the variant, that is, the candidate boxes will be merged without using the detailed ground object information to adjust the candidate box of the construction project. When $\gamma = 0$, only the detailed ground object information is used to obtain the confidence of the candidate box. After a parameter search on the validation set, the optimal value of γ is 0.75.

4.4. Potential Application

Construction projects are one of the main sources of environmental pollution in urban areas due to their pollution, such as harmful gases, noise, dust, solid and liquid wastes during construction [32,33]. Some construction companies engage in illegal behaviors, such as building without approval and disturbing beyond surveillance. It is necessary to use remote sensing technologies to detect the illegal situations by supervising the full coverage in a region or even a city. The application of this method can accurately and effectively identify construction projects under construction and determine whether the projects comply with regulations by checking with the soil and water conservation plans and authorized scope that have been registered in the National Soil and Water Conservation Supervision and Management System.

Multi-level image scene identification and understanding is a hot issue in current high-resolution earth observation technology [34]. In addition to identifying construction projects, this study can further investigate the contents of the automatic extraction of disturbance ranges, predicting soil loss from construction projects, and evaluating the distribution of regional soil and water loss. In addition, this method can also be used in other fields such as building detection [35] and illegal mining identification [36].

5. Conclusions

In this paper, complex scenes of construction projects were identified based on the high-resolution remote sensing images of Wuhan City. Construction projects and their detailed ground object data sets were first prepared and highly informative detailed ground objects were selected for object detection. Then, the Faster RCNN algorithm was used to detect the construction project and the highly informative detailed objects separately, and a post-processing procedure were used to jointly improve the confidence and accuracy of construction project identification results. The results show that this method can effectively reduce the false detection rate and improve the goodness matching between the detection

results and the ground truth. The application of this method can accurately and effectively identify the construction project under construction.

The introduction of detailed ground object combination detection for construction project scene identification has the following three advantages: (1) the detailed ground objects are consistent with the network structure of the construction project training; (2) it can improve the detection performance for construction projects to a certain extent without significantly raising the complexity of training; (3) it can identify the construction period of the construction project (such as topping out or completion) and the implementation of soil and water conservation measures (such as whether the spoil and slag are covered with dust screens) by the features of detailed ground objects. However, as the internal information of the construction project is extremely complex without any unified feature representation, the post-processing procedure based on the conventional deep learning method has a limited improvement effect on the detection results, and for some construction projects, the improvement effect of this method is not satisfactory. How to better integrate the detailed ground object information, and consider the co-occurrence and mutually exclusive relationship between the detailed ground object information, so as to obtain a compact construction project feature representation, is the direction and focus of future study.

Author Contributions: Investigation, formal analysis, visualization, and writing—original draft, J.P.; conceptualization and project administration, Z.W.; software, writing—review and editing, R.L. and S.S.; supervision, T.Z. and W.X.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dynamic Monitoring of Soil Erosion and Production Construction Project Supervision project of Wuhan in 2019 (Grant Number CKSK2019444/TB), Annual Dynamic Monitoring of Soil and Water Loss in Fujian Province in 2022 (Grant Number CKSK2022687/TB), and Decomposition and Evaluation of Long-term and Phased Targets of Soil and Water Conservation Rate in Cities and Counties of Fujian Province (Grant Number CKSK2022686/TB).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rafiq, W.; Musarat, M.A.; Altaf, M.; Napiyah, M.; Sutanto, M.H.; Alaloul, W.S.; Javed, M.F.; Mosavi, A. Life Cycle Cost Analysis Comparison of Hot Mix Asphalt and Reclaimed Asphalt Pavement: A Case Study. *Sustainability* **2021**, *13*, 4411. [[CrossRef](#)]
2. Andrade-Núñez, M.J.; Aide, T.M. Built-up expansion between 2001 and 2011 in South America continues well beyond the cities. *Environ. Res. Lett.* **2018**, *13*, 084006. [[CrossRef](#)]
3. E, J.-p. Strengthening supervision of engineering industry to strengthen weaknesses and strive to create a new situation of water conservancy in the new era: Speech at the 2019 National Conference on Water Conservancy Work (Abstract). *China's Water Conserv.* **2019**, *2*, 11. (In Chinese)
4. Pu, C.Y. Ideas and requirements for promoting soil and water conservation monitoring and information technology. *Soil Water Conserv. China* **2017**, *5*, 1. (In Chinese)
5. E, J.-p. Minister E Jingping put forward clear requirements for national soil and water conservation work in 2020. *Soil Water Conserv. China* **2020**, *2*, 2. (In Chinese)
6. Blaschke, T.; Strobl, J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *Proc. GIS-Z. Fur Geoinf.* **2001**, *6*, 12.
7. Bruzzone, L.; Carlin, L. A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2587. [[CrossRef](#)]
8. Aksoy, S.; Yalniz, I.Z.; Tasdemir, K. Automatic Detection and Segmentation of Orchards Using Very High Resolution Imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3117–3131. [[CrossRef](#)]
9. Li, K.; Wan, G.; Cheng, G.; Meng, L.Q.; Han, J.W. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS-J. Photogramm. Remote Sens.* **2020**, *159*, 296. [[CrossRef](#)]
10. Yu, Y.; Guan, H.; Ji, Z. Rotation-invariant object detection in high-resolution satellite imagery using superpixel-based deep Hough forests. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2183–2187. [[CrossRef](#)]

11. Chen, Z.; Chen, D.; Zhang, Y.; Cheng, X.; Zhang, M.; Wu, C. Deep learning for autonomous ship-oriented small ship detection. *Saf. Sci.* **2020**, *130*, 104812. [[CrossRef](#)]
12. Yang, X.; Sun, H.; Fu, K.; Yang, J.; Sun, X.; Yan, M.; Guo, Z. Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks. *Remote Sens.* **2018**, *10*, 132. [[CrossRef](#)]
13. Jiang, D.W.; Jiang, X.W.; Zhou, Z.L. Technical support of artificial intelligence for informatization supervision of soil and water conservation. *J. Soil Water Conserv.* **2021**, *35*, 1–6. (In Chinese)
14. Kang, Q.; Jiang, D.W.; Fu, Q.H.; Wang, X.G. On the identification of construction disturbance patches based on optimal segmentation scale. *Sci. Soil Water Conserv.* **2017**, *15*, 126–133. (In Chinese)
15. Dumitru, C.O.; Cui, S.; Schwarz, G.; Dancu, M. Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *8*, 1635–1650. [[CrossRef](#)]
16. Xu, Z.W.; Yang, Y.; Hauptmann, A. A discriminative CNN video representation for event detection. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1798–1807.
17. Fan, H.H.; Chang, X.J.; Cheng, D.; Yang, Y.; Xu, D.; Hauptmann, A.G. Complex event detection by identifying reliable shots from untrimmed videos. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 736–744.
18. Yu, J.; Lei, A.; Hu, Y. Soccer video event detection based on deep learning. In Proceedings of the International Conference on Multimedia Modeling, Thessaloniki, Greece, 8–11 January 2019; pp. 377–389.
19. Feng, X.; Jiang, Y.; Yang, X.; Du, M.; Li, X. Computer vision algorithms and hardware implementations: A survey. *Integration* **2019**, *69*, 309–320. [[CrossRef](#)]
20. Yadav, S.K.; Tiwari, K.; Pandey, H.M.; Akbar, S.A. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowl.-Based Syst.* **2021**, *223*, 106970. [[CrossRef](#)]
21. Chang, X.H.; Yang, Y.; Long, G.D.; Zhang, C.Q.; Hauptmann, A.G. Dynamic concept composition for zero-example event detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 3464–3470.
22. Yuan, J.; Zha, Z.J.; Zheng, Y.T.; Wang, M.; Zhou, X.D.; Chua, T.S. Learning concept bundles for video search with complex queries. In Proceedings of the 19th ACM International Conference on Multimedia, Scottsdale, AZ, USA, 28 November–1 December 2011; pp. 453–462.
23. Feng, L.; Bhanu, B. Semantic concept co-occurrence patterns for image annotation and retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 785–799. [[CrossRef](#)]
24. Fu, Y.; Xiang, T.; Jiang, Y.G.; Xue, X.; Sigal, L.; Gong, S. Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. *IEEE Signal Process. Mag.* **2018**, *35*, 112–125. [[CrossRef](#)]
25. Ramos, J. Using tf-idf to determine word relevance in document queries. In Proceedings of the First Instructional Conference on Machine Learning, Piscataway, NJ, USA, 3–8 December 2003; Volume 242, pp. 133–142.
26. Aizawa, A. An information-theoretic perspective of tf-idf measures. *Inf. Process. Manag.* **2003**, *39*, 45–65. [[CrossRef](#)]
27. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
28. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1440–1448.
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
30. Zhao, K.; Wang, Y.; Zhu, Q.; Zuo, Y. Intelligent Detection of Parcels Based on Improved Faster R-CNN. *Appl. Sci.* **2022**, *12*, 7158. [[CrossRef](#)]
31. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
32. Zolfagharian, S.; Nourbakhsh, M.; Irizarry, J.; Ressang, A. Environmental impacts assessment on construction sites. In Proceedings of the Construction Research Congress 2012, with the Theme Construction Challenges in a Flat World, West Lafayette, IN, USA, 21–23 May 2012; pp. 299–301.
33. Oke, A.; Aghimien, D.; Aigbavboa, C.; Madonsela, Z. Environmental sustainability: Impact of construction activities. In Proceedings of the 11th International Conference on Construction in the 21st Century, London, UK, 19 September 2019; pp. 229–234.
34. Li, D.R.; Tong, Q.X.; Li, R.X.; Gong, J.Y.; Zhang, L.P. Current issues in high-resolution earth observation technology. *Sci. China Earth Sci.* **2012**, *55*, 1043–1051. [[CrossRef](#)]

35. Shi, F.; Zhang, T. A Multi-Task Network with Distance-Mask-Boundary Consistency Constraints for Building Extraction from Aerial Images. *Remote Sens.* **2021**, *13*, 2656. [[CrossRef](#)]
36. He, D.K.; Le, B.T.; Xiao, D.; Mao, Y.C.; Shan, F.; Ha, T.T.L. Coal mine area monitoring method by machine learning and multispectral remote sensing images. *Infrared Phys. Technol.* **2019**, *103*, 103070. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.