Research on High Resolution Optical Remote Sensing Image Technology Based on Internet Architecture

Gengyi Xiao*
Department of Mathematics and Computer Technology, Guilin Normal College, Guilin, China

*Corresponding author: gy_xiao@glnc.edu.cn

Abstract. Target detection technology in high-resolution optical remote sensing images is of great significance in both civil and military fields. Based on the Internet architecture, this paper proposes a multi-class target detection method for high-resolution and high-resolution optical remote sensing images. The fused multi-layer features are used for detection. In view of the challenges of complex background and deformation of the target, the deformable convolution network can be used for reference to extract the characteristics of the target itself and reduce the background interference. In order to reduce the storage space required by the deep convolution neural network model and increase the portability of the network, I also put forward a method of lightening the network. Experiments show that our proposed method is feasible. Compared with the popular target detection method based on deep convolution neural network, our method has great advantages in precision and recall rate.

Keywords: Internet architecture; High score optical remote sensing image; Target detection, deep learning.

1. Introduction
In recent years, with the rapid development of optical remote sensing imaging technology, the spatial resolution of high-resolution optical remote sensing images has been continuously improved. Compared with other types of remote sensing images such as synthetic aperture radar (SAR), high-resolution high-resolution optical remote sensing images have unique advantages [1]. In the civil and military fields, the research and application of high-resolution optical remote sensing images are receiving extensive attention.

With the increasing sources of high-score optical remote sensing images, we can get more and more high-score optical remote sensing images, unlike the previous one, which can only get a few or even one. Traditional machine learning (machine learning of shallow structure) is difficult to extract features, and the generalization ability of network structure is poor. Workers who detect targets in high-score optical remote sensing images are gradually abandoning the method based on shallow machine learning [2-3]. In recent years, deep learning [4], especially the super-high performance of deep convolution neural network in image processing, has attracted a lot of attention [5]. Using deep convolution neural network to detect targets in high-resolution optical remote sensing images can
avoid the process of artificially designing specific algorithms to extract target features according to different targets in shallow machine learning. Deep convolution neural network directly extracts the corresponding features of targets from the network itself, so deep convolution neural network has greater adaptability [6-7]. The successful application of deep convolution neural network to target detection in high-resolution optical remote sensing images is a big step towards intelligence in remote sensing image processing.

The purpose of this paper is to improve the target detection technology of visible high-resolution optical remote sensing images based on Internet architecture, and to study multi-class target detection methods based on deep learning, so as to improve the problems of high false alarm rate and easy missed detection of small targets in the complex background of high-resolution optical remote sensing images. At the same time, the proposed method has the advantages of strong multi-scale adaptability, good affine invariance and good universality, and can be used to detect various ground targets such as airplanes, vehicles, surface ships, oil tanks and so on.

2. Distributed storage of remote sensing data

2.1. HBase database
HBase is a distributed storage system with high reliability, high performance and column orientation, which can access a large amount of data efficiently by using several cheap PCs, and provides a reliable method for realizing distributed storage of data. HBase is the same as Oracle and other databases we know, providing the platform with the functions of storing and reading data. However, HBase is different from traditional database in that it stores data through unstructured database. The access interface of HBase is very simple, which can not support the access of complex data, and can not provide structured query of relational database. It can only be indexed by row keys, and the data distribution and query function depend on rowkey.

HBase is similar to MongoDB's sharding mode, which distributes data to each node through key values. The difference is that MongoDB locates and distributes data to each node through configserver, while Hbase accesses Zookeeper and obtains the address of ROOT table, thus obtaining the information of table and the address of data storage.

2.2. Design of data storage structure
HBase can manage sparse tables and provide I/O interfaces for them. With this advantage, the HBase table (Table 1) is designed to support the writing of spatial data. Based on the table structure of HBase [9], the storage table structure of remote sensing image is designed by integrating with OGC protocol. Similar to the traditional spatial data table, each type of feature corresponds to a table. Columns of Hbase table are designed according to the characteristics of features, and each piece of geographic information corresponds to a row key, which is composed of GeoHash index and feature ID.

| Line key | Column cluster | Geometric field | Attribute 1 | … | Attribute n |
|----------|----------------|----------------|-------------|---|-------------|
| GeoHash_1 | TS: geometry 1 | TS: attribute 1 | … | … |
|          |                | value 1        |             |   |
| GeoHash_2 | TS: geometry 2 | TS: attribute 2 | … | … |
|          |                | value 2        |             |   |
| …        |                | …              | …           | … |

2.3. Design of storage architecture
The heterogeneous data collected from different data sources are converted into text data by data format and stored in HDFS of data union module. This section combines the advantages of HDFS and HBase to design a remote sensing data storage architecture, as shown in Figure 1:
In this section, a spatial data import method based on MapReduce is designed, which converts the data format into Hfile file, and can import and process spatial data quickly. Operating steps are:

1. Storing remote sensing data in HDFS;
2. Create a spatial data storage table [10] with the above table structure;
3. Using Hadoop framework to analyze remote sensing data, and storing the data in Hfile after calculating GeoHash index;
4. Store HFile file in the corresponding HBase table. The remote sensing data is converted into text data and stored in HDFS. The data in HDFS is read by the combination of GeoHash and HBase, and converted into new data to save the ground object information of remote sensing data.

3. Target detection model based on deep learning

3.1. Method based on regional suggestion
The remarkable feature of the method based on region suggestion is "two steps". The candidate regions with possible targets, also called RoI, are given by a certain method, and then the features of RoI are extracted by convolution neural network. Finally, the target type is discriminated by classifier and the positioning is adjusted by frame regression network.

Among the regional suggested methods, Faster R-CNN proposed in 2015 is the first algorithm to realize end-to-end training and detection, which has achieved remarkable progress in speed, accuracy and consumption of computing resources [11]. Figure 2 shows its detection framework. Because the method of region suggestion has the advantage of high detection accuracy, there are many improvements and applications of this method. For example, Mask R-CNN[12] proposed based on this algorithm can complete two tasks of instance segmentation and target detection at the same time, showing good scalability and room for improvement.
The whole Faster R-CNN can be regarded as consisting of Fast R-CNN and Region Proposal Network (RPN). The basic feature extraction network uses deep convolutional neural networks CNN, such as ZFNet, AlexNet, VGGNet, GoogleNet, RestNet, etc., but the feature extraction only uses the convolutional layer in front of the convolutional neural network, excluding the subsequent fully connected classification. The obtained feature map is shared by the region suggestion network, candidate region classification and location regression network.

3.2. Target detection based on deep convolution neural network

With the continuous development of deep convolution neural network, related researchers began to try to apply deep convolution neural network to target detection. At present, target detection algorithms based on deep convolution neural network can be roughly divided into two types: one is a two-stage detector represented by Faster RCNN, and the other is a single-stage detector represented by YOLO and SSD.

3.2.1. Combined with context information. The influence of complex background information in remote sensing images on target detection can not be ignored. In this section, context information is introduced into the target detection model to improve the accuracy of target detection. Set the image size \( w, h, ROI \) area coordinate \((x_{min}, y_{min}, x_{max}, y_{max})\) and scale ratio \( s \), and define the corresponding context area:

\[
\begin{align*}
    x'_{\min} &= \max(1, x_{\min} - s(x_{\max} - x_{\min})/2) \\
    y'_{\min} &= \max(1, y_{\min} - s(y_{\max} - y_{\min})/2) \\
    x'_{\max} &= \min(1, x_{\max} - s(x_{\max} - x_{\min})/2) \\
    y'_{\max} &= \min(1, y_{\max} - s(y_{\max} - y_{\min})/2)
\end{align*}
\]

The activation values of context features are divided into \( m \) groups and constrained respectively (\( m = 10 \) in this paper), and the final target feature representation is obtained, which is called local-context feature in this paper, as shown in formula (2)

\[
F_{joint} = f_{joint} [\text{normalize}(F_{local} \oplus \lambda \ast F_{context})]
\]

In which \( f_{joint} \) stands for HRBM model; \( \lambda \ast F_{context} \) represents element-by-element multiplication within the group; \( \oplus \) stands for splicing operation.

3.2.2. Deformable convolution. In traditional CNN, the default conventional matrix sampling leads to poor affine transformation invariance, and the pooling operation only obtains certain translation and scaling invariance. In recent years, some researchers have proposed to adapt to the geometric diversity
of targets by irregular sampling to enhance the affine transformation invariance of CNN. When the target rotates or deforms, deformable convolution sampling is obviously more effective in collecting the pixels of the target itself. In addition, when multiple targets are closely attached, deformable convolution sampling can distinguish each target more effectively than conventional rectangular sampling, thus helping to reduce missed detection.

The sampling position of deformable convolution has its own offset direction, which is no longer limited to the inside of the block. Deformable convolution sampling has better scale invariance and rotation invariance than conventional rectangular convolution. After using deformation convolution, it is as shown in formula (3).

\[ y(p_0) = \sum_{p_n \in R} W(p_n) \cdot X(p_0 + p_n + \Delta p_n) \]  

Since the offset \( p_n \) is usually not an integer, and the coordinates of the pixel position of the actual sampling point must be an integer, the direct, simple and rude rounding will lead to obvious errors. Here, the pixel value of each point is obtained by bilinear interpolation, and the formula of bilinear difference is shown in Formula (4):

\[ X(p) = \sum_q G(q, p) \cdot X(q) \]  

In which \( p = p_0 + p_n + \Delta p_n \) and \( q \) represent all integer positions within the range of \( X \), for convenience of calculation, the two-dimensional convolution kernel \( G(q, p) \) can be decomposed into two one-dimensional convolution kernels \( g(q, p) \), as shown in formula (5), wherein the one-dimensional convolution kernel is calculated according to formula (6).

\[ G(q, p) = g(q_x, p_x) \cdot g(q_y, p_y) \]  

\[ g(q, p) = \max(1, 1 - |q - p|) \]  

The schematic diagram of the 3*3 deformable convolution operation process is shown in Figure 3. The original conventional convolution process is divided into two paths, and one more path is added to learn the offset of the full connection layer.
In this method, deformable convolution is introduced to improve the basic network Res Net-50, and 3x3 deformable convolution is used to replace the traditional convolution in the last residual block in each group of residual units, and the number of channels remains unchanged.

3.2.3. Lightweight design of network model. There are many methods to reduce the network model, such as "pruning", which is a method to change the storage space required by the model in software, mainly by artificially removing some parameters in the network model.

Firstly, a threshold value is set artificially, and when the parameters in the storage matrix of deep convolutional neural network are less than this value, it is directly set to 0; Then, all the bytes corresponding to the 0 parameter are deleted to change the storage mode of the deep convolution neural network storage matrix; After the above two steps are completed, the accuracy will drop, so finally, a network training will be conducted to restore the accuracy. Another commonly used method is to introduce a large number of 1*1 convolution operations for network decomposition and dimension reduction, so as to reduce the size of the model. However, introducing a large number of new layers will reduce the convergence ability of the network, resulting in a decline in accuracy.

In this paper, the fully connected layer is changed into convolution layer and pooling layer, which greatly reduces the storage space required by deep convolution neural network. The process of modifying the fully connected layer into pooling layer and convolution layer is shown in Figure 4 (for simplicity, only the modified part of the fully connected layer is drawn).
There are several reasons for our modification in this way: 1. The parameters required by the full connection layer far exceed those of the convolution layer; 2. In Faster RCNN, in order to ensure the variable scale of the input image, ROI-Pooling layer is introduced to fix the resolution of the features output by the last convolution layer. However, the introduced ROI-Pooling layer will lead to feature deformation to a certain extent, so it is necessary to pick up a convolution layer for feature extraction, and then introduce a feature pooling layer for feature sampling. Directly connecting the convolution layer to the feature pooling layer will lead to a sharp decline in accuracy.

Compared with the network before improvement, the network improved by lightening the network needs much less storage space, which is beneficial to the transplantation and expansion of the network. At the same time, the light detection speed has also increased to a certain extent.

4. Experimental analysis

4.1. Parallelization classification accuracy
Under the Spark distributed environment, the classification accuracy of the residual network distributed computation proposed in this paper is 98.9%, and the classification accuracy is shown in Figure 5. With the framework of Tensor Flow On Spark, the SS-CNN algorithm is implemented to classify remote sensing images with an accuracy of 96.7%, which is compared with the distributed residual network.
In addition, the time consumed by distributed computing is not half of that consumed by single-machine computing. The reason is that in distributed computing, after each iteration, two working nodes need to transmit the gradient to the reference server, update it, and then redistribute the updated parameters to the two working nodes. During this period, the information transmission between machines consumes part of the time, resulting in a waste of time.

4.2. Lightweight analysis of network model

The purpose of lightening the network model is to reduce the size of the finally generated convolutional neural network model, thereby increasing the portability and improving the running speed. In the corresponding design, we use a convolution layer and a feature pooling layer instead of the full connection layer to reduce the size of the storage space required by the model and improve the speed.

It can be seen from Table 2 that convolution layer has very strong learning ability, and it is feasible to use convolution layer after RoI-Pooling operation. By adjusting the number of channels in convolution layer, we can further adjust the size and speed of neural network model appropriately. When the number of channels is reduced to 10, the lightweight structure does not affect the accuracy and requires less storage space, and the testing speed is faster. Further reducing the number of channels will reduce the accuracy and the testing speed and required storage space will not change again. The more intuitive change is shown in Figure 6 (the required storage space has been normalized).

**Table 2** Accuracy, test time and required storage space corresponding to different channel numbers

| Detector          | Number of channels | Required storage space (M) | Precision | Test time (s) |
|-------------------|--------------------|----------------------------|-----------|---------------|
| Faster RCNN       | 802                | 56.3                       | 0.620     | 0.102         |
|                   | 86                 | 66.1                       | 0.891     | 0.077         |
| Lightweight network | 60                 | 55.3                       | 0.801     | 0.067         |
|                   | 32                 | 51.6                       | 0.881     | 0.063         |
|                   | 10                 | 49.7                       | 0.809     | 0.060         |
From Figure 6, we can more intuitively see that with the increase of the number of convolution kernel channels with new convolution layer, the required test time, storage space and accuracy change. When the number of channels increases to a certain value, its accuracy will not change, which indicates that the number of channels at this time is enough to extract features. At the same time, when the number of channels is further reduced, its speed will no longer increase and the required storage space will no longer decrease. Because we only improve the fully connected layer, the main structure of the network has not changed much, so after the number of channels in a single convolution layer drops to a certain value, the required storage space will not change significantly.

5. conclusion
High-precision target detection in high-resolution optical remote sensing images is of great significance in both military and civil fields. With the increasing sources of high-resolution optical remote sensing images, how to efficiently use these images for high-precision target detection has become a topic of great concern to related researchers. In this paper, the classic deep learning target detection model Faster R-CNN is improved based on Internet architecture. In addition, semantic segmentation is used to reduce the false detection rate through semantic information. After a series of improvements, the detection performance has been greatly improved, but the finally generated convolutional neural network model needs a lot of storage space. Therefore, we once again put forward a method of lightening the network model to compress the model. After this method, the storage space required by the network model is drastically reduced, which greatly enhances the portability of the network model. The method is suitable for multi-class target detection tasks of high-resolution optical remote sensing images, and has strong robustness and high precision.

Acknowledgments
This work was financially supported by Guangxi university scientific research fund: Research on the key technologies of vehicle-mounted network based on CPS (2013YB286).

References
[1] Wang M, Yuan T, Cheng Y. Development of On-orbit Geometric Calibration for High Resolution Optical Remote Sensing Satellite [J]. Geomatics and Information Science of Wuhan University, 2017, 42(11):1580-1588.
[2] Hui-Li W, Ming Z, Chun-Bo L, et al. Ship detection in optical remote sensing image based on...
[3] Awad M M. Toward Robust Segmentation Results Based on Fusion Methods for Very High Resolution Optical Image and LiDAR Data [J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2017, 10(5):2067-2076.

[4] Bi F, Chen J, Zhuang Y, et al. A Decision Mixture Model-Based Method for Inshore Ship Detection Using High-Resolution Remote Sensing Images [J]. Sensors, 2017, 17(7):1470.

[5] Lan Z, Yang L. Study on Multi-Scale Window Determination for GLCM Texture Description in High-Resolution Remote Sensing Image Geo-Analysis Supported by GIS and Domain Knowledge [J]. Isprs International Journal of Geo Information, 2018, 7(5):175.

[6] Yehia A, Safy M, Amein A S. Fusion of High-Resolution SAR and Optical Imageries Based on a Wavelet Transform and IHS Integrated Algorithm [J]. International Journal of Engineering Research in Africa, 2021, 52.

[7] Wendt J, Rivera A, Wendt A, et al. Recent ice-surface-elevation changes of Fleming Glacier in response to the removal of the Wordie Ice Shelf, Antarctic Peninsula [J]. Annals of Glaciology, 2017, 51(55):5530-5557.

[8] Scott G J, Hagan K C, Marcum R A, et al. Enhanced Fusion of Deep Neural Networks for Classification of Benchmark High-Resolution Image Data Sets[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15:1451-1455.

[9] Masse A, Lefevre S, Binet R, et al. Denoising Very High Resolution Optical Remote Sensing Images: Application and Optimization of Non Local Bayes method [J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2018:691-700.

[10] Oh J, Lee C. Extraction of digital elevation model using stereo matching with slope-adaptive patch transformation [J]. KSCE Journal of Civil Engineering, 2017, 20(7):1-8.

[11] Haut J M, Fernandez-Beltran R, Paoletti M E, et al. Remote Sensing Image Superresolution Using Deep Residual Channel Attention[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, PP (99):1-13.

[12] Jiao J, Wang B, Wang C, et al. STUDY ON HIGH RESOLUTION MEMBRANE-BASED DIFFRACTIVE OPTICAL IMAGING ON GEOSTATIONARY ORBIT [J]. Isprs International Archives of the Photogrammetry Remote Sensing & Spatial Information Sciences, 2017, XLII-1/W1:371-375.