

Region Change Detection Model for Remote Sensing Images Based on U-net

Lijie Gu

Department of Civil Engineer, Hefei University of Technology, Hefei, China.

2856285278@qq.com

Abstract. This paper presents a region-based convolutional neural network based on the U-net model and region change detection method, aiming to address the shortcomings of traditional remote sensing change detection methods such as poor feature characterization and long processing time. By employing a region-based object detection method and data augmentation strategy, a more generalized network model is obtained, and attempts are made to further improve training efficiency using U-net++. Experimental results demonstrate that the U-net and its derived models exhibit faster convergence and higher training efficiency, proving their effectiveness. The research methods and experimental results of this paper have important theoretical and practical significance for remote sensing image change detection..

Keywords: Remote sensing change detection; U-net model; Regional change detection; Feature extraction.

1. Introduction

Change detection in remote sensing images refers to a set of methods, algorithms, and tools used to monitor and analyze surface or scene changes. These techniques are widely applied in fields such as remote sensing image analysis, environmental monitoring, disaster management, urban planning, among others. The framework of change detection technology encompasses crucial steps including data acquisition, preprocessing, feature extraction, change detection algorithms, analysis and interpretation, as well as visualization and presentation. These steps aim to achieve monitoring and analysis of surface or scene changes.

Remote sensing images exhibit diverse spatial resolutions and cover various scenes, encompassing a wide range of artificial and natural features such as buildings, roads, farmland, vegetation, rivers, lakes, wetlands, and more. The heterogeneity of data sources, application scenarios, and the substantial increase in dataset sizes necessitate the development of efficient and versatile methods for change detection. With the advent of the big data era, enhanced computing capabilities have alleviated issues like inefficient training, paving the way for the emergence of deep learning. Building upon the foundation of deep learning, this study employs the U-net model for region change detection in remote sensing images. By leveraging rich feature representations, the model adapts to diverse scenarios in remote sensing image change detection. The end-to-end training process optimizes results, enhancing the accuracy of change detection.

The innovations of this paper are as follows:

(1) Addressing the shortcomings of poor feature characterization and long processing time in traditional remote sensing change detection, this paper adopts a region-based object detection method, based on the application of the U-net network, to rapidly understand the geometric and contextual spatial features of the targets, extract and identify targets in the scene in a short time, and achieve rapid model training.

(2) The training set of this paper uses random data cropping to perform spatial multi-scale sampling, flipping, noise addition, and other data augmentation strategies on existing data, resulting in a more generalized network model.

(3) Utilizing U-net++ for performance enhancement provides the convenience of using in_channels without the need to modify the input model, making it more user-friendly compared to U-net. Additionally, U-net++ introduces nested and dense skip connections, reducing the semantic gap between the encoder and decoder. After extensive experimentation, U-net++ has demonstrated

faster convergence and higher training efficiency, affirming the effectiveness of the framework.

The subsequent structure of this paper is as follows: Part II introduces recent change detection work, the use of U-net network, and data augmentation; Part III introduces the research methods, including an introduction to the U-net framework, feature processing, and loss function construction; Part IV provides specific experimental analysis based on the results; Part V summarizes the experimental results and future prospects.

2. Related Works

2.1 Change Detection Research

Remote sensing technology plays a significant role in advancing social progress, environmental protection, economic development, and national defense by analyzing Earth's changing information [1]. Satellite remote sensing enables timely analysis of remote sensing images at different times to promptly detect dynamic changes in land resources and predict natural disasters [2].

In recent years, scholars worldwide have been devoted to the research of change detection technology. The accuracy of traditional change detection methods heavily relies on difference images, which may lead to unstable detection results due to the loss of significant information during the process of generating difference images. For example, changes between images may only include changes in brightness or color, whereas in the real world, changes can be diverse, including the appearance, disappearance, and shape change of targets. Inaccuracies in feature extraction, data registration, algorithm assumptions, and data noise limit the precision of change detection results.

With the increasing complexity of land cover changes and diversity of remote sensing data, traditional methods have exhibited many limitations. For instance, although the classification comparison method [3] has resolved issues such as radiation correction and matching, it requires secondary image classification, reducing the accuracy of change detection due to seasonal changes, satellite mapping deviations, and atmospheric conditions. Additionally, changes between images may only include changes in brightness or color, whereas in the real world, changes can be diverse, including the appearance, disappearance, and shape change of targets. Inaccuracies in feature extraction, data registration, algorithm assumptions, and data noise limit the precision of change detection results. Even with object-based image analysis and transformation methods, extensive human computation and operational steps are required to reduce change.

As a result, new change detection methods and image processing algorithms continue to emerge [4]. For example, the use of change vector analysis and patch-based change detection methods using probabilistic statistical theory. However, no single change detection method is optimal according to practical situations [5]. These methods serve different purposes in various scenarios and each has its own applicability and limitations. With the rapid development of remote sensing data acquisition technology, the increasing variety of unique remote sensing data and their combinations pose new technical requirements for change detection.

2.2 Application of U-net Model

The U-net network model is widely used for its time-saving, labor-saving, and versatility. Compared to traditional land feature extraction methods, convolutional neural networks have become mainstream for high-resolution image building extraction due to their excellent feature representation capabilities. Building upon the CNN model, Long et al. proposed the FCN model [6], replacing the fully connected layer at the end of the CNN with a convolutional layer, effectively addressing the CNN models inability to accurately extract building contours, and improving model training and prediction performance. Subsequently, Ronneberger et al. [7] proposed the U-net model, which incorporates skip connections for feature fusion.

In recent years, through nested connections in the U-net model, the U-net++ [8] neural network has been developed, reducing the loss of detailed information after the addition of neural layers. U-net++ has been applied in clinical medical imaging fields such as bacteria detection [9] and cell

segmentation [10]. In the remote sensing domain, the U-net network and its improved models have been widely used in land resource management, agricultural and forestry monitoring, and natural disaster assessment [11]. Although semantic segmentation networks like U-net have shown good segmentation effects, incomplete segmentation of target regions and inaccurate extraction of regions remain issues. Further enhancing segmentation accuracy is a critical issue being researched by scholars both domestically and internationally.

3. Method

3.1 Introduction of U-net Framework

Net is named for its "U" shape and is a neural network model used for image segmentation, as shown in Fig.1. Comprising an encoder and decoder, it exhibits excellent feature extraction and image segmentation capabilities. The model consists of contracting and expansive paths, symmetrically arranged: the contracting path captures local features, while the expansive path ensures precise localization. In remote sensing change detection, U-Net effectively extracts features from images at different periods, generating change maps. Adopting the classic encoder-decoder structure, the encoder of U-Net includes convolutional layers, pooling layers, and activation functions. The encoder progressively reduces the size of feature maps, extracting features at various abstraction levels from input images. Simultaneously, the decoder restores these features to the original image size for prediction using upsampling and skip connections. Upsampling enlarges the feature map to the original image size through interpolation, while skip connections link encoder and decoder feature maps to preserve detail information, enhancing segmentation accuracy. For scenarios with limited samples, pretrained U-Net models trained on large datasets can be considered to capture general features. Subsequently, fine-tuning with transfer learning on smaller sample sets can improve model performance. This approach accelerates model convergence and enhances performance by initially training the pretrained model on a large dataset to capture general features. These pretrained weights are then used as initial parameters, followed by fine-tuning on smaller sample sets to adapt to specific tasks.

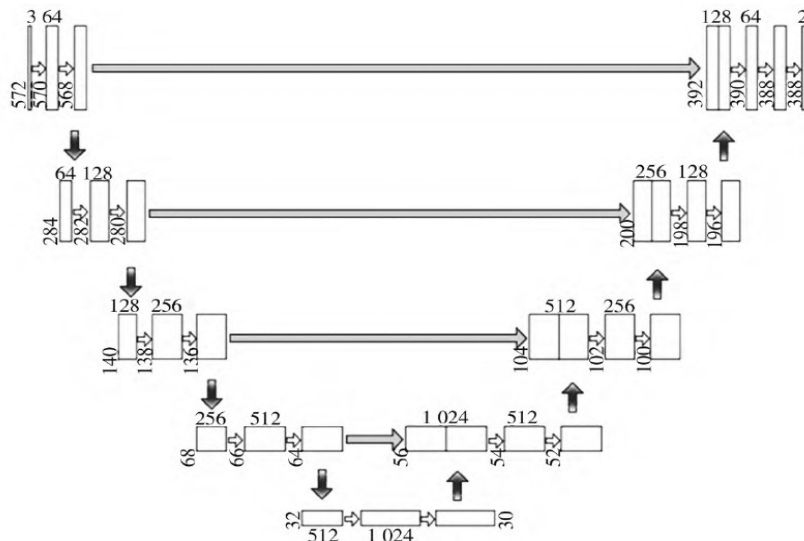


Figure 1 U-net network structure[12]

3.2 Loss Function

The U-Net model, when used for tasks such as image segmentation, utilizes binary cross-entropy loss (BCELoss) as the loss function to measure the difference between the predicted results and the ground truth labels.

The computation process of BCELoss is as follows:

(1) Suppose the U-Net model outputted a probability value between 0 and 1 through its final layer, representing the probability of each pixel belonging to the foreground (1).

(2) The ground truth label data should be a matrix or tensor of the same size as the output result, with each pixel corresponding to a binary label representing the foreground (1) or background (0).

(3) For each pixel, the computation of BCELoss is as follows: Firstly, the corresponding binary label for the pixel is obtained using the ground truth label data. Next, the probability value for the pixel corresponding to the output probability is calculated. Finally, the probability value is substituted into the BCELoss formula to compute the loss for that pixel.

(4) The overall loss value is obtained by summing or averaging all the pixels.

The formula for binary cross-entropy loss is as follows:

$$L = -[y * \log(p) + (1 - y) * \log(1 - p)]$$

Where:

-L represents binary cross-entropy loss

-y is the actual binary label

-p is the output probability value

-log denotes the natural logarithm function

By minimizing the binary cross-entropy loss, the model can gradually optimize, making the predicted results closer to the ground truth labels, thus achieving more accurate image segmentation tasks. In practical applications, the loss function can also be fine-tuned or weighted according to specific task requirements.

The U-Net model also utilizes the softmax function combined with the cross-entropy loss function to calculate the difference between the predictions and the ground truth labels, and optimize the model parameters through backpropagation, enabling the U-Net model to effectively perform image segmentation.

In the U-Net model, the softmax function is typically used for multi-classification problems to convert the model's output into a probability distribution.

The softmax function can be explained as an operation that maps an input vector to a probability distribution. The exponential function of each element and is a normalization factor that ensures the sum of the probabilities equals 1. The exponential function amplifies larger values in the input vector, increasing their corresponding probabilities, while reducing the probabilities of smaller values. In the U-Net model, the softmax function is usually used for the output of the last layer, converting the channels corresponding to each pixel into a probability distribution. This enables an intuitive understanding of the output results and ensures that the model's output conforms to the properties of probability theory, facilitating subsequent segmentation result analysis and post-processing.

3.3 Data Augmentation

Constructing a good training dataset is one of the direct and effective methods to improve the network's generalization. Common data augmentation strategies in image processing include flipping, rotating, scaling, cropping, adjusting contrast, adding noise, and more. The data augmentation strategy in this paper involves randomly sampling plots from the entire remote sensing area, setting selection criteria to remove exceptional data, and performing multi-scale sampling, random flipping, and noise addition in the spatial direction, resulting in a better generalized model.

Conventional training datasets are obtained directly through the sliding window approach, where data blocks are extracted using fixed sizes and steps. Setting a too large step may lead to information loss, while setting a too small step may result in redundant data, wasting computational resources without improving training quality. In this paper, data blocks are randomly sampled from the entire satellite image data, and exceptional data are removed. To address different sampling rates caused by different data, this paper considers a multi-scale spatial sampling data augmentation method. Horizontal flipping of the training data enriches the dataset, while adding minimal noise disturbance helps prevent model overfitting.

4. Experiment and Results Analysis

4.1 Introduction of Dataset

The dataset consists of training and validation data required for change detection, as well as pairs of test images. It includes 637 pairs of ultra-high-resolution (VHR, 0.5m/pixel) Google Earth (GE) image blocks, along with corresponding binary change labels. Each temporal image contains RGB three bands (brightness values quantized to 0-255), with a size of 1024×1024 pixels. The temporal span of these dual-state images ranges from 5 to 14 years, exhibiting significant land use changes, particularly in building growth. It covers various types of buildings including villas, high-rise apartments, small garages, and large warehouses. The experiment focuses on changes related to buildings, including building growth (changes from soil/grass/hardened ground or under construction buildings to new building areas) and building decline.

4.2 Training Settings

4.2.1 Data preprocessing

Samples are obtained from true color remote sensing images from two different time periods and different sensors, and these images are unified into the same coordinate reference system and resampled to a spatial resolution of 0.5m. The samples range from vegetation to buildings, bare land to buildings, and farmland to roads. Each type has approximately 500 samples. To meet the GPU memory requirements, the original image is segmented into images with a size of 1024×1024 pixels, with most samples having lengths and widths between 200 and 800 pixels. Secondly, the two different temporal images are preprocessed to unify the geographic coordinate system and image spatial resolution. The difference image is obtained by subtracting the later period image from the earlier period image, resulting in a series of rectangular regions. The intersecting regions of the same change type are merged, and then the model is used to segment the changed regions.

4.2.2 Network training

The basic process involves first merging the two temporal images, and then selecting a 3×3 neighborhood for each position pixel to obtain a one-dimensional vector of length 54, which is then input into the corresponding deep learning model for change detection. During the training process, the number of training rounds is set to 100, the batch size is set to 2, the optimizer is AdamW, and CosineAnnealingDecay is used to adjust the learning rate.

4.3 Experimental Results

The data coverage range is $1669m \times 1368m$, with main land cover categories including buildings, wasteland, vegetation, and roads. Three sets of change detection results are extracted from the test set, as shown in Fig 2. Where Fig 2(a) represents the earlier temporal image, Fig 2(b) represents the later temporal image, and Fig 2(c) represents the change detection results.



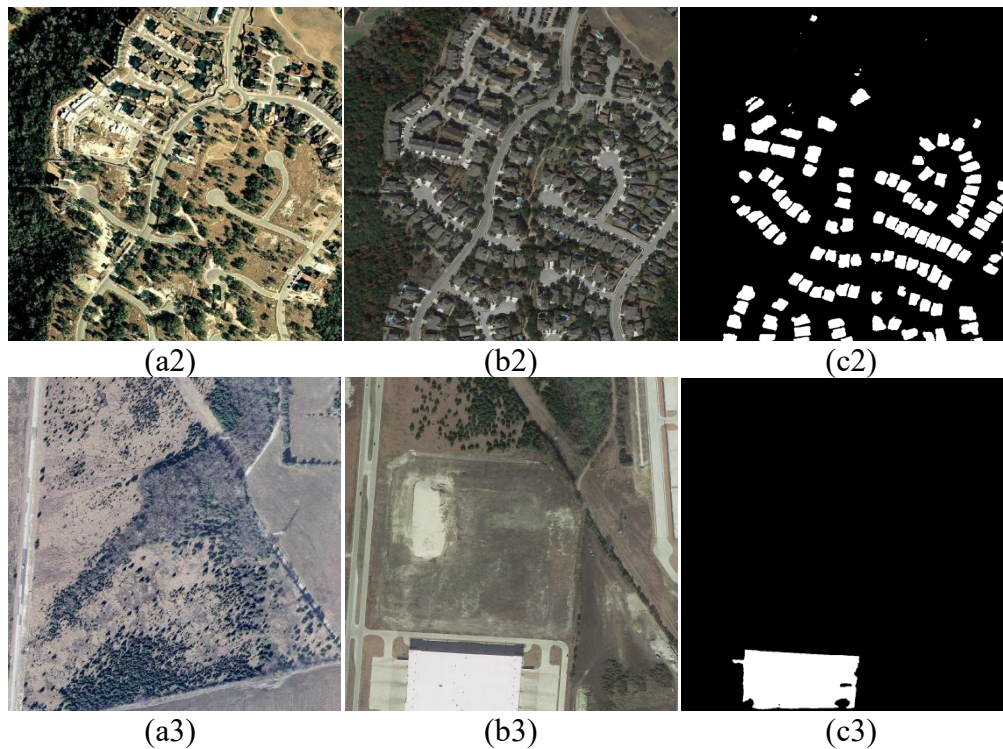


Figure 2 Experimental results

4.4 Results Analysis

Region-based change detection methods not only improve training efficiency but also enhance the overall accuracy of the results. U-net utilizes convolutional computations and feature map fusion in a layer-by-layer manner and is simpler than vectorization work. It performs better in detecting block-like areas of local information and can effectively identify areas with small band differences but significant boundary changes.

The U-Net model can effectively extract spatial positional information. However, when the model performs downsampling and upsampling operations, it may result in spatial information extraction with blurred boundaries. By using U-net++, edge noise is reduced, training efficiency is improved, and experimental results are optimized. In future research, further optimization of deep learning models and data processing methods can be pursued.

Currently, the U-Net model mainly focuses on pixel-level change detection. For improvements to the U-net model, the introduction of semantic information can be attempted to enhance the interpretability of the detection results. Furthermore, the integration of remote sensing data from different sources and resolutions can improve the accuracy and reliability of change detection. Additionally, integrating data preprocessing, feature extraction, and classification/segmentation into an end-to-end training framework can simplify the model training process and improve efficiency.

5. Conclusion

In the vast amount of remote sensing data, rapid and accurate change detection analysis is becoming increasingly important. Based on the U-net network model, this study utilized a regional change detection approach to train and construct a remote sensing change detection model. At the same time, U-net++ was attempted in the experiment, the dataset was optimized, the loss function was adjusted, and large-scale remote sensing dynamic change detection was accomplished using the time series of remote sensing images.

The experimental results fully demonstrate the effectiveness of the U-net network in high-resolution remote sensing change detection. Faced with the complexity of current land cover types, the interwoven distribution of land features, and blurred boundaries, the U-net model can effectively

acquire land cover type information within the study area, thus providing assistance for subsequent precise land resource planning and efficient ecological restoration. Additionally, it also provides an effective method for urban planning, environmental protection, disaster risk assessment, and other fields.

References

- [1] Yang B, Mao Y, Chen J, Liu J Q, Chen J, Yan K. A review of remote sensing change detection based on deep learning: a bibliometric and analysis. *Journal of Remote Sensing*, 2023, 27(09): 1988-2005.
- [2] Wang Z Q, Li C, Chen Y L, Liu Z N, Ding H D, Yu Y, Li T P, Guo Z X. Progress in interpretation and evaluation monitoring of geological disasters based on remote sensing technology. *Journal of Nanhua University (Natural Sciences)*, 2023, 37(02): 31-36.
- [3] Wu Y K. Research on remote sensing image change detection technology. National University of Defense Technology, 2013.
- [4] Ma Y F, Li H. A review of remote sensing change detection technology methods. *Geomatics and Spatial Information Technology*, 2014, 37(01): 132-134.
- [5] Zhang Z X, Jiang H W, Pang S Y, Hu X Y. Review and prospect in change detection of multi-temporal remote sensing iamges[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(07): 1091-1107.
- [6] Evan S ,Jonathan L ,Trevor D .Fully Convolutional Networks for Semantic Segmentation.[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(4): 640-651.
- [7] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. *Proceedings of International Conference on Medical Image Computer-assisted Intervention Munich, Germany Springer*, 2015: 234-241.
- [8] Peng D, Zhang Y, Guan H. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++[J]. *Remote Sensing*, 2019, 11(11): 1382.
- [9] Shen Y, Xie F, Wang Y L, Tan B, Fan J X. Fully automatic high-throughput colony counter based on Unet++ segmentation model[J]. *Food and Machinery*, 2023, 39(11): 53-57+142.
- [10] Wu X L, Qi S M, Meng J, et al. MSE-UNet: Cell segmentation model based on multi-scale features and attention mechanism[J]. *Journal of Qufu Normal University (Natural Science Edition)*, 2023, 49(02): 71-76.
- [11] Han X. Building Change Detection in High-Resolution Remote Sensing Images Based on Deep Learning[D]. Changan University, 2022. DOI: 10.26976/d.cnki.gch au.2022.001082.
- [12] Xie N, Chen L, Pei Z Q, He Z Q, Chen T. Research on White Body Welding Spot Positioning Based on Improved U-Net[J]. *Journal of Testing Technology*, 2024, 38(01): 12-18.