



Article

A Novel UNet 3+ Change Detection Method Considering Scale Uncertainty in High-Resolution Imagery

Ting Bai ¹, Qing An ², Shiquan Deng ³, Pengfei Li ¹, Yepei Chen ^{1,*}, Kaimin Sun ^{4,5}, Huajian Zheng ^{6,7,8} and Zhina Song ¹

- ¹ School of Computer Science, Hubei University of Technology, Wuhan 430010, China; baiting@whu.edu.cn (T.B.); pengfei.li@hbut.edu.cn (P.L.); 20210077@hbut.edu.cn (Z.S.)
- ² School of Artificial Intelligence, Wuchang University of Technology, Wuhan 430223, China; 120160450@wut.edu.cn
- ³ Wuhan Academy of Water Science, Wuhan 430010, China; dengshiquan@whu.edu.cn
- ⁴ The State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China; sunkm@whu.edu.cn
- ⁵ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430010, China
- ⁶ Key Laboratory of Natural Resources Monitoring in Tropical and Subtropical Area of South China, Ministry of Natural Resources, Guangzhou 510663, China; zhenghuajian@mail.bnu.edu.cn
- ⁷ Surveying and Mapping Institute Lands and Resource Department of Guangdong Province, Guangzhou 510663, China
- ⁸ Guangdong Science and Technology Collaborative Innovation Center for Natural Resources, Guangzhou 510663, China
- * Correspondence: chenyp@whu.edu.cn

Abstract: The challenge of detecting changes in high-resolution remote sensing imagery often stems from the difficulties in effectively extracting features and constructing appropriate change detection models considering the scale characteristics of ground objects. To solve these issues, we propose a novel UNet 3+ change detection method that considers the scale characteristics inherent in various land-cover change types. Our method includes three key steps: a multi-scale segmentation method, a class-specific UNet 3+ method, and an object-oriented change detection method based on UNet 3+. To verify the effectiveness of this method, we select two datasets for experiments and compare our proposed method with the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, and the UNet 3+ multi-scale hierarchical sampling method. Our experimental results show that our proposed method has higher overall accuracy and F1, lower missed detection rate and false detection rate, and can detect more changes in ground features than other methods. To verify the scalability of this method, we compare this method with traditional change detection methods such as PCA-k-means, OCVA, a single-scale sampling method based on random forest, and a class-specific object-based method. Experimental results and accuracy indexes show that our proposed method better considers the scale characteristics of ground objects and achieves higher accuracy. Additionally, we compared our proposed method with other DLCD methods including LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff. Our results show that our proposed method effectively considers edge information and has an acceptable time consumption. Our approach not only considers the full-scale characteristics of the feature extraction but also the scale characteristics of the change detection model. In addition, it considers a more practical feature extraction unit (object), making it more accurate.

Keywords: change detection; UNet 3+; high-resolution; object-oriented



Citation: Bai, T.; An, Q.; Deng, S.; Li, P.; Chen, Y.; Sun, K.; Zheng, H.; Song, Z. A Novel UNet 3+ Change Detection Method Considering Scale Uncertainty in High-Resolution Imagery. *Remote Sens.* **2024**, *16*, 1846. <https://doi.org/10.3390/rs16111846>

Academic Editors: Kamil Krasuski and Damian Wierzbicki

Received: 3 April 2024
Revised: 6 May 2024
Accepted: 17 May 2024
Published: 22 May 2024



Copyright: © 2024 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

Change detection is an advanced monitoring method that utilizes multi-date high-resolution remote sensing images to identify changes on the Earth's surface. Consequently, it finds extensive utilization across various domains including land-use/land-cover

analysis [1], urbanization process [2], building damage assessment and disaster impact analysis [3], and numerous other fields [4,5]. The advancement of change detection holds the potential to significantly enhance the precision and timeliness of earth observation. However, the scale uncertainty issues in feature extraction and change detection pose limitations on the accuracy and applicability of change detection. For feature extraction, appropriate extraction units must be carefully selected because the type and size of the unit directly determine the accuracy of the expression of change information. For change detection, the selection of a specific change detection model will directly affect the learning ability of features, and, thus, influence the final change detection results. In addition, change detection is a more challenging task compared to single temporal remote sensing image applications such as land-use and land-cover classification, scene classification, and object detection. On one hand, change detection involves multiple temporal images and has a larger amount of data. On the other hand, change detection involves the extraction of a small amount of information, whereas other application problems involve the extraction of a large amount of information [6]. Therefore, the scale problem will have a more significant impact on change detection tasks.

To solve the scale problem in extracting change features, scholars have developed multi-scale feature expression methods, optimal scale selection methods, and deep feature expression methods. Multi-scale feature expression methods describe the characteristics of the target structure within a certain scale range (such as wavelet transform [7,8] and object-oriented). For example, Eklund et al. [9] extracted multi-scale features of ground objects based on wavelet transform and obtained change detection results layer-by-layer based on the multi-scale features. This method performs change detection based on pixels, so it cannot avoid salt and pepper noise, and does not consider the true scale of ground objects. Therefore, an object-oriented-based feature extraction method was developed. For example, the integration of spectral and spatial features in multi-scale objects is developed for change detection [10]. This study can perfectly unify the contradiction between fine-scale precision and coarse-scale separability [11,12]. However, this study is like the research on multi-scale feature expression methods based on wavelet transform. It integrates the multi-scale change detection results to obtain more accurate change detection results [13,14] or fuses spectral and spatial features of multi-scale objects for change detection [10]. However, fusing shallow multi-scale results or multi-scale features to improve the accuracy of change detection is limited, and it cannot be applied to complex scenes [9]. Optimal scale selection methods are used to choose the best scale from multiple options. The popular method is based on an object-oriented approach. For top-down segmentation, Zhou et al. [15] used the complexity of the segmented object and the prior knowledge of thematic maps to determine the optimal scale. The optimal scale parameter for the thematic map may not be the same as the optimal scale parameter for the selected classification framework [16]. Moreover, current optimal scale selection methods do not consider the differences in the characteristics of ground objects. Recently scholars developed deep feature expression methods. This method can extract low-dimensional and high-dimensional semantic features for change detection by utilizing the powerful image representation and understanding capabilities of deep learning models [17,18]. For example, Mou et al. [2] extracted the spectral, spatial, and temporal characteristics of change information, and combined RNN and CNN models to detect urban land expansion. This method is not a full-scale feature extraction network, so the extracted features are still incomplete. In addition, the characteristics of the ground objects are not effectively considered.

To address the scale issues in change detection models, scholars have developed machine learning models and deep learning models. Levien et al. [19] employed decision trees to detect changes in images. Lu et al. [20] utilized artificial neural network methods for change detection. Huang et al. [21] employed the dark object concept and SVM to automatically perform forest cover change analysis. Although these methods require fewer training samples, their feature learning capabilities are limited and they cannot learn the nonlinear relationship between the target's feature information and changes.

The emergence of deep learning methods has provided new insights. Currently, this method is widely used in image fusion, image registration, scene classification, object detection, land-use and land-cover classification, and object-based image analysis [22,23]. Initially, the majority of research on deep learning change detection (DLCD) was focused on differential feature representation, while the final classifier employed simple classifiers such as NNs [24,25]. For example, Xu et al. [26] employed auto encoders (AE) to extract deep features from two VHR images and computed the difference between these deep features for change detection. El Amin et al. [27] used the transfer learning method to learn the multi-scale spectral and spatial characteristics of ground objects through pre-trained CNN models for detecting changes in QuickBird-2 satellite images. Later, different deep learning neural networks were integrated to conduct change detection. Gong et al. [28] combined the CNN's ability to learn multi-scale features with the GAN's ability to automatically generate high-quality difference maps for change detection in high-resolution images. Zhang et al. [29] used CNN and transformer to detect binary changes in land use. Although the integration of the deep learning model has enhanced change detection accuracy, most methods are applied to specific sceneries, such as building change detection and landslide detection, etc. General problems such as land-cover and land-use change detection remain challenging [6]. Although some general DLCD methods have emerged, they do not address the scale characteristics of ground object changes and cannot produce end-to-end change detection results.

Recently, the emergence of the encoder-decoder end-to-end SegNet segmentation network [30] has sparked a boom in scene classification. The SegNet network utilizes up-pooling operations to achieve superior up-sampling, but it does not incorporate skip connections. Therefore, it cannot effectively account for the multi-scale deep characteristics of ground objects. To address this issue, the UNet network was introduced, extending SegNet by incorporating a plain skip connection between the encoder and decoder layers to enhance the acquisition of multi-scale deep features. This approach improved segmentation accuracy and has been successfully applied to medical images [31] and remote sensing images [32]. In 2018, Zhou et al. [33] proposed a novel medical image segmentation architecture called UNet++, which can be regarded as an extension of UNet. This network structure successfully reduces the gap between the feature maps from the encoder and decoder networks. In addition to the connection between the encoder and decoder networks, UNet++ utilizes a series of nested and dense skip connections, endowing the network architecture with the advantage of capturing details. Consequently, it can produce superior segmentation results compared to UNet. Peng et al. [34] proposed an end-to-end change detection method based on UNet++. However, the UNet++ network does not capture sufficient information from multi-scales. Therefore, Huang et al. [35] proposed a UNet 3+ model. The UNet 3+ incorporates the full-scale skip connection and deep supervision methods based on UNet++. The full-scale skip connection enables the combination of high-level and low-level semantic feature maps from different scales, while deep supervision learns hierarchical representations from full-scale aggregated feature maps [35]. Additionally, UNet 3+ reduces network parameters and improves computational efficiency. Mo et al. [36] utilized it to detect building changes. Hence, this network can be used to extract and learn full-scale features of images and holds great potential in detecting changes in land cover and land use.

However, these DLCD methods are based on pixel-based or region-based inputs, which do not consider the characteristics of the ground objects well and have poor edge consideration capabilities. Incorporating object-oriented methods into DLCD can effectively utilize the advantages of both feature extraction and change detection. Liu et al. [37] used CNN under the OBIA framework to achieve much higher accuracy than traditional classifiers such as random forests and support vector machines in drawing wet maps using unmanned aerial vehicle imagery. Liu et al. [38] used LSTM under the OBIA framework to perform change detection on aerial images, which has higher accuracy than pixel-based methods. Zheng et al. [39] used object-oriented and twin fully convolutional network

models for change detection and completed building hazard assessment work. These successful cases highlight the significant potential of integrating object-oriented methods with deep learning models. However, this method does not consider the scale characteristics of ground objects, leading to inaccuracies in the change detection model.

To solve the problem of incomplete feature extraction and inaccurate change detection models, we propose an object-oriented UNet 3+ change detection method that considers the scale characteristics. This method includes three steps: a multi-scale segmentation method; a class-specific UNet 3+ method; and an object-oriented change detection method based on UNet 3+. In the first step, multi-temporal images and historical land cover maps are overlaid and segmented to obtain multi-scale building, road, bare soil, vegetation, water, and concrete objects. In the second step, the multi-scale hierarchical sampling (MSHS) method is integrated with UNet 3+ for building, road, bare soil, vegetation, water, and concrete objects individually. By combining training samples at multiple scales, the optimal UNet 3+ model is selected for each type of ground feature change. The third step involves employing the optimal UNet 3+ deep learning model to detect changes in multi-scale objects such as buildings, roads, bare soil, vegetation, water, and concrete. This method considers the actual scale of change for each surface object. It not only enables the extraction and learning of full-scale features associated with surface object changes but also facilitates the selection of the optimal change detection model for each surface object. Consequently, it effectively addresses issues related to incomplete feature extraction and inaccuracies in change detection models.

2. Materials and Methods

2.1. Data

Two datasets are obtained by GF-2 satellite for experimental analysis, both from Liuzhou City, Guangxi Province, China. The R, G, and B bands are selected. The reference change/no change map and land-cover map are both from the China Surveying and Mapping Bureau. The reference change/no change maps and historical land-cover maps are produced based on data with a resolution better than 1 m, and the accuracy is controlled within 2 pixels. All images have a resolution of 0.8 m.

Image registration [40] and relative radiometric consistency methods [20] are used to preprocess each dataset. The images in Dataset 1 and Dataset 2 were automatically registered using the second-order affine polynomial and nearest-neighbor resampling method in ArcGIS [41]. This process results in a registration error of less than 0.5 pixels, which is considered acceptable for high-resolution imagery [41]. To ensure a consistent spectral response, histogram matching was applied to the image pair with the greatest spectral variance as reference images [42].

Dataset 1, shown in Figure 1a,b, consists of true color multi-spectral images of the GF-2 satellite in 2015 and 2016, respectively, with a size of 5640×2842 pixels. Figure 1c is a reference change/no change map, and Figure 1d is a land-cover map for 2015, which includes six land-cover types: buildings, roads, bare soil, water, vegetation, and concrete. Concrete refers to concrete surfaces other than buildings and roads. Therefore, land-cover change types include building change, road change, base soil change, water change, vegetation change, concrete change, and no change.

Dataset 2, shown in Figure 2a,b, consists of true color multi-spectral images of the GF-2 satellite in 2015 and 2016, respectively, with a size of 4401×3417 pixels. Figure 2c is a reference change/no change map, and Figure 2d is a land-cover map for 2015, which includes six land-cover types: buildings, roads, bare soil, water, vegetation, and concrete. Concrete refers to concrete surfaces other than buildings and roads. Therefore, land-cover change types include building change, road change, base soil change, water change, vegetation change, concrete change, and no change.

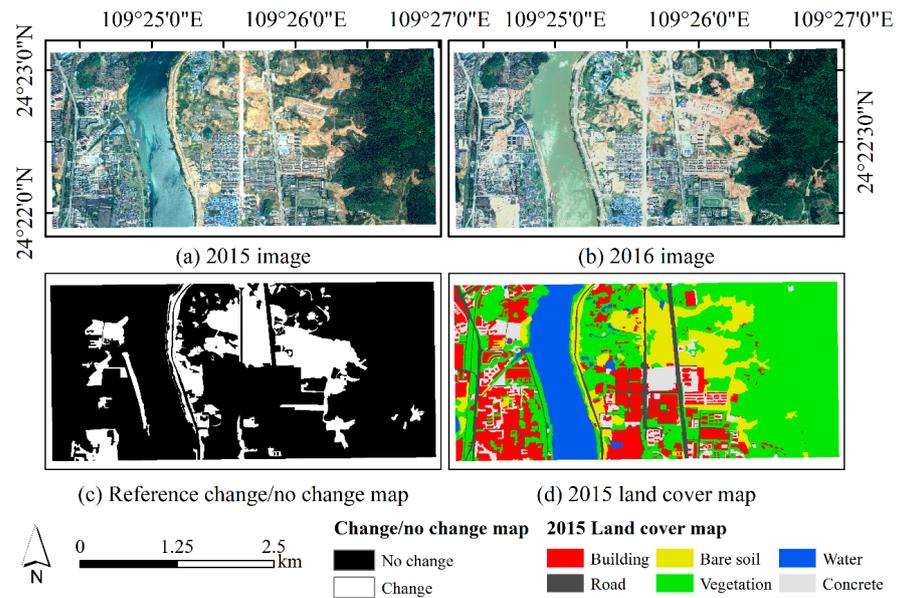


Figure 1. Dataset 1: (a,b) are true color multi-spectral images of the GF-2 satellite in 2015 and 2016, respectively; (c) is a reference change/no change map; and (d) is a land-cover map for 2015.

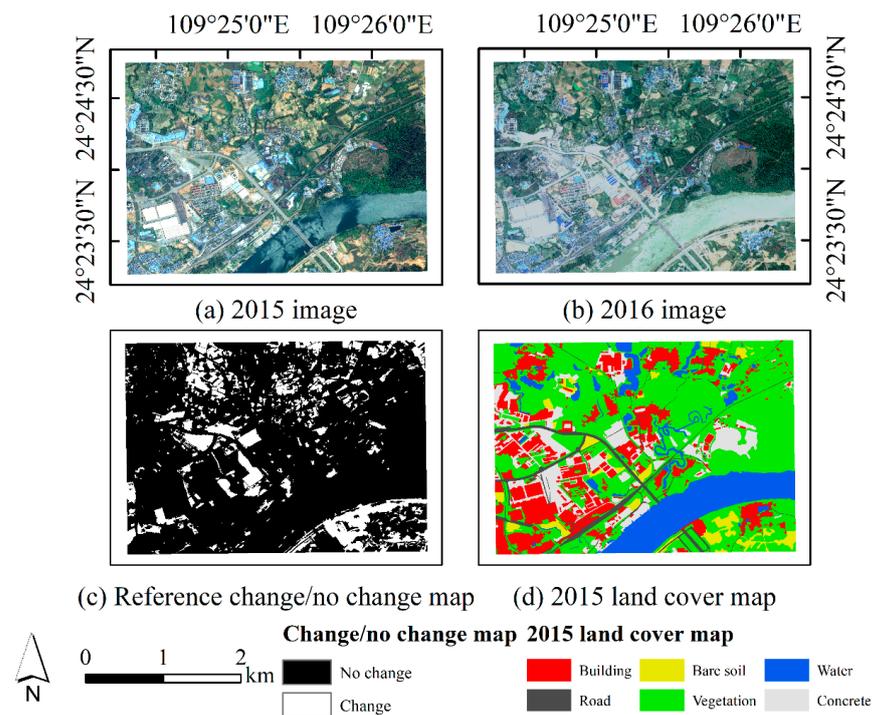


Figure 2. Dataset 2: (a,b) are true color multi-spectral images of the GF-2 satellite in 2015 and 2016, respectively; (c) is a reference change/no change map; and (d) is a land-cover map for 2015.

2.2. Method

To address the scale issues related to incomplete feature extraction and inaccuracies in change detection models, we propose an object-oriented UNet 3+ change detection method that considers the scale characteristics inherent in various land-cover change types. This method fully utilizes the multi-scale feature expression ability of object-oriented methods, as well as the powerful generalization ability and full-scale feature learning ability of UNet 3+. The flowchart of this method is shown in Figure 3.

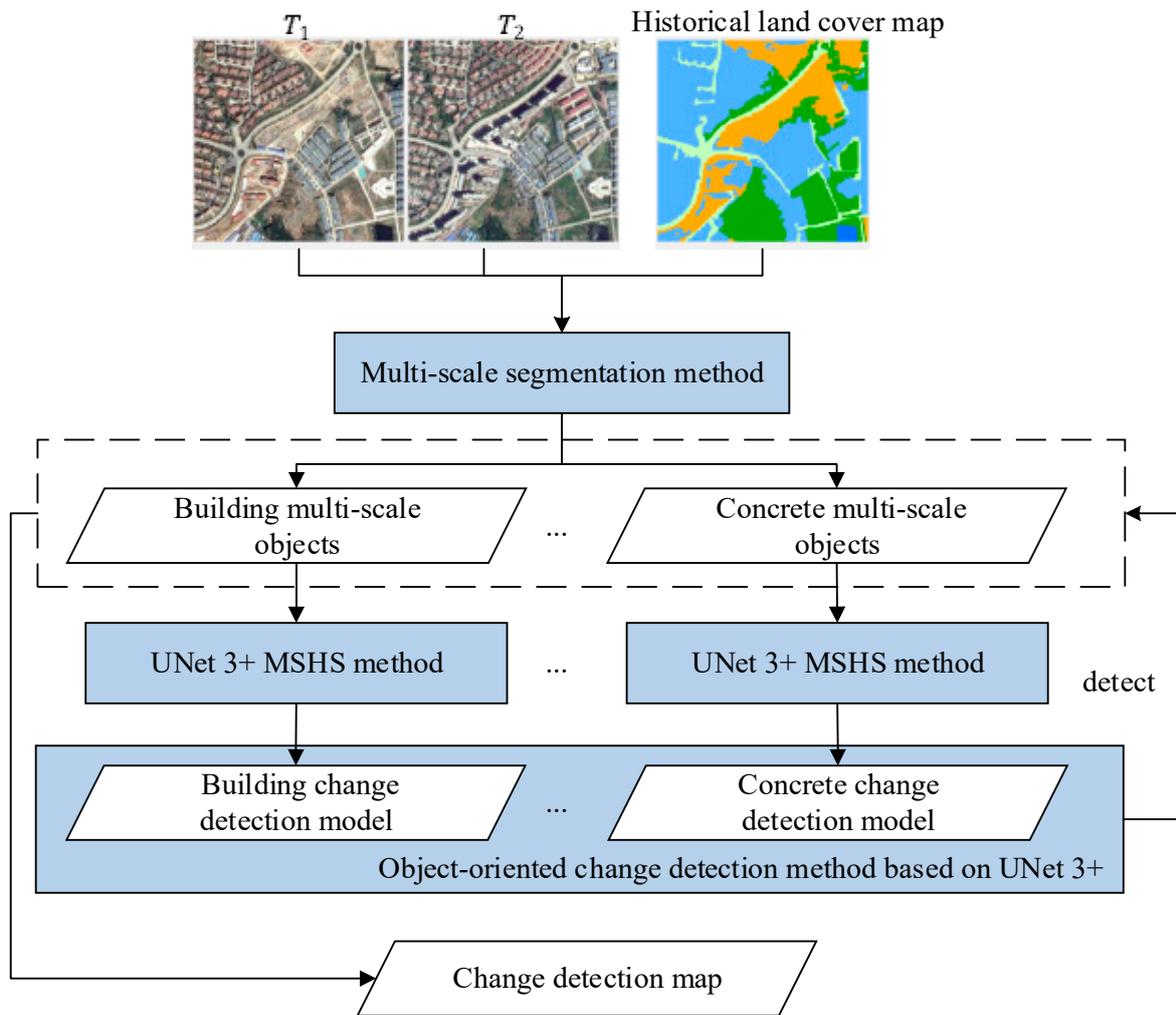


Figure 3. The flowchart of our proposed method.

Based on Figure 3, the method consists of three main steps. The first step is a multi-scale segmentation method that considers the scale characteristics inherent in various land-cover change types. The second step involves a class-specific UNet 3+ MSHS method that also considers the scale characteristics of different land-cover change types. Finally, the third step is an object-oriented change detection method based on UNet 3+. The following subsections will provide a detailed explanation of each step.

2.2.1. A Multi-Scale Segmentation Method

The multi-scale segmentation method aggregates pixels and groups them together based on their shape and compactness through the use of segmentation scale parameter, to create image objects [43]. Assuming that we have two remote sensing images, image S_1 in T_1 and image S_2 in T_2 , we can combine them using a simple band stacking technique to create a single image pair S . We then use multi-scale segmentation methods along with historical land-cover maps to obtain objects from the image pair S . The image pair S can be split from top to bottom to create sub-scale object layers $\{L_1, L_2, \dots, L_N\}$. After the segmentation process, each object corresponds to a specific land-cover type such as buildings, roads, bare soil, vegetation, water, and concrete, based on the attribute information extracted from the historical land-cover map. To use these objects in a deep learning network, we resize them to a size of 256×256 using bilinear interpolation, as the network cannot handle irregular objects. The flowchart for this process is shown in Figure 4.

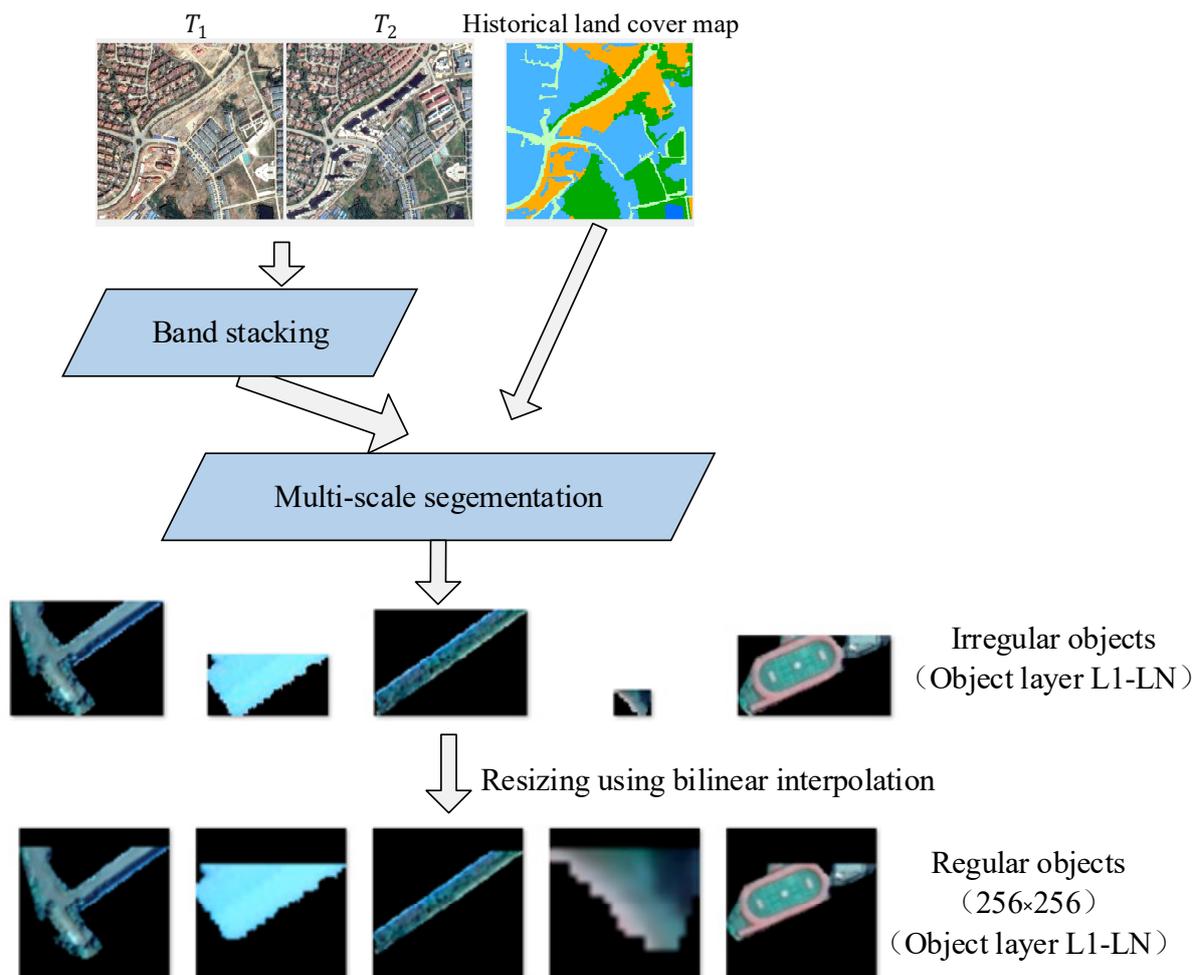


Figure 4. The flowchart of a multi-scale segmentation method.

2.2.2. A Class-Specific UNet 3+ Method

The MSHS method can automatically add training samples of changing and no changing regions without increasing manual workload [44]. By combining the object-oriented method, MSHS method, and UNet 3+, we propose a class-specific UNet 3+ method. This method considers the full-scale characteristics inherent in various land-cover change types, establishing a more robust change detection model. It also enhances the UNet 3+ model by extracting feature information from objects, accurately capturing the scale of geographical entities, and considering the edge information. In this section, we introduce the UNet 3+ network, which includes full-scale skip connections and full-scale deep supervision. Subsequently, we discuss a class-specific UNet 3+ method.

1. UNet 3+

The UNet 3+ model has introduced a new full-scale skip connection that improves the interconnection between encoders and decoders [35]. This feature also optimizes the internal connection between decoder subnets, which results in an overall improvement in the model's performance [35]. Unlike UNet and UNet++, which employ a plain skip connection nested and dense skip connections similar to DenseNet [45], neither effectively captures sufficient feature information across all scales. To address this issue, UNet 3+ combines smaller and same-scale feature maps from the encoder with larger-scale feature maps from the decoder at each decoder layer. Consequently, the UNet 3+ model can capture full-scale information, including both small-scale detail and large-scale semantic information [35], which is illustrated in Figure 5.

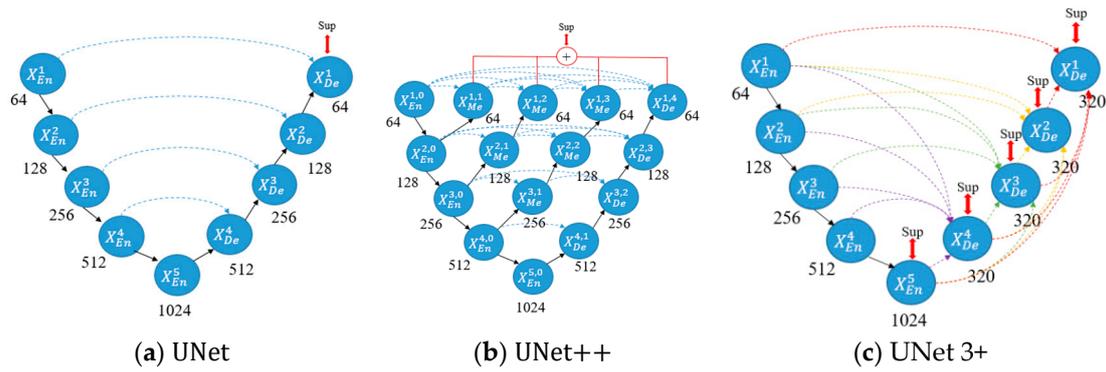


Figure 5. Comparison of UNet 3+ with UNet and UNet++. Compared to UNet and UNet++, UNet 3+ utilizes full-scale skip connections to capture small-scale detail and large-scale semantic information.

(1) Full-scale skip connections

Full-scale skip connections use skip connections to combine the smaller and same-scale feature maps from the encoder and the high-scale semantic feature maps from the decoder. An example of full-scale skip connections for UNet 3+ [35] is shown in Figure 6. A feature map X^3_{De} is established, like UNet, where the feature map comes from the encoding layer X^3_{En} at the same scale. In addition, a set of skip connections between encoders and decoders can provide low dimensional information from encoding layers X^1_{En} and X^2_{En} at a smaller scale using non-overlapping maximum pooling operations. In addition, a series of internal decoder skip connections transmit high-dimensional semantic information X^4_{De} and X^5_{De} from large-scale decoding layers through bilinear interpolation.

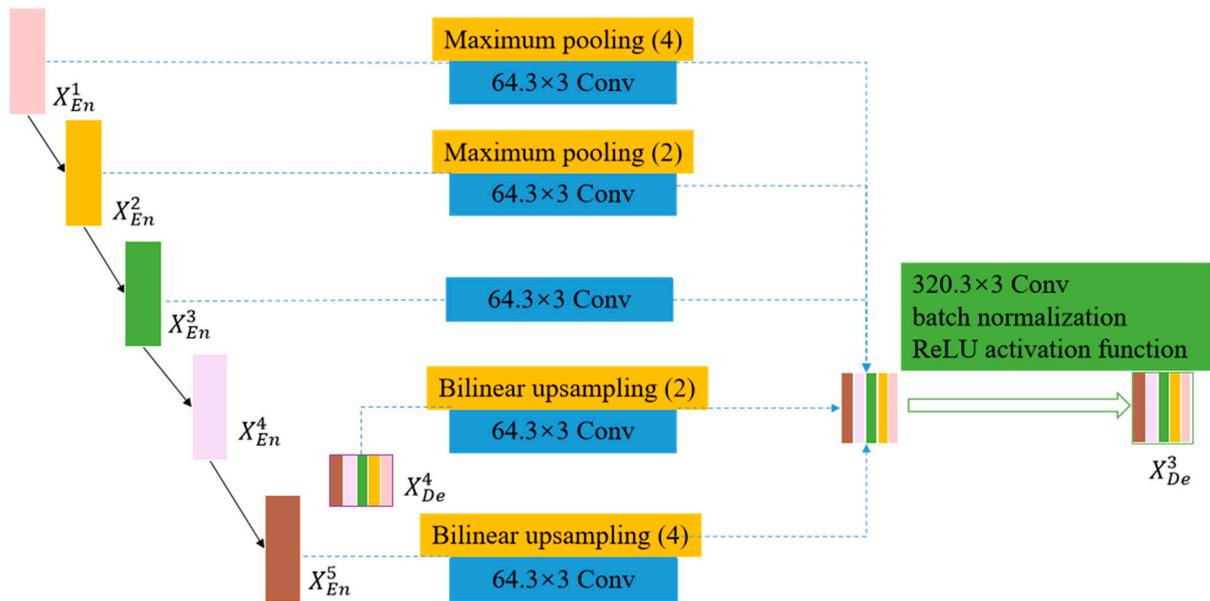


Figure 6. The establishment of a full-scale feature graph in the third decoding layer X^3_{De} .

There are five feature maps with the same resolution, and it is necessary to further unify the number of channels and reduce excess information. We chose a 3×3 kernel size with 64 filters for convolution. To integrate shallow information with deep semantic information seamlessly, a feature aggregation mechanism is further implemented on feature maps at five scales, including a 3×3 kernel size with 320 filters, batch normalization, and ReLU activation functions. The skip connection is represented by the formula as follows. i

refers to the down-sampling layer in the encoding layer and N is the number of encoding layers. The feature map X_{De}^i [35] is as follows:

$$X_{De}^i = \left\{ H \left(\left[\underbrace{C \left(D \left(X_{En}^k \right) \right)_{k=1}^{i-1}}_{Scales: 1^{th} - i^{th}}, c \left(X_{En}^i \right), \underbrace{C \left(U \left(X_{De}^k \right) \right)_{k=i+1}^N}_{Scales: (i+1)^{th} - N^{th}} \right] \right), i = 1, \dots, N-1 \quad (1)$$

$C(\cdot)$ is a convolutional operator, and $H(\cdot)$ uses convolution, batch normalization, and ReLU activation functions to implement feature aggregation mechanisms. $D(\cdot)$ and $U(\cdot)$ refer to downsampling and upsampling operators, respectively. $[\cdot]$ represents concatenation.

(2) Full-scale deep supervision

To learn multi-scale hierarchical feature maps from the full-scale feature map set, full-scale deep supervision was adopted in UNet 3+ [35]. Compared to the deep supervision implemented in the generated full-scale feature maps in UNet++, UNet 3+ generates a side output from each decoder level and is supervised by the ground truth map.

To achieve deep supervision, the last layer stage of each decoder is sent to regular 3×3 convolutional layers and then is subjected to bilinear upsampling and S-type functions. To highlight the boundaries of segmentation, a multi-scale structural similarity index loss function was introduced in UNet 3+ to assign greater weights to fuzzy boundaries. Therefore, a UNet 3+ network pays more attention to fuzzy boundaries because the larger the regional distribution difference, the higher the MS-SSIM value.

Cropping two corresponding $N \times N$ blocks from the segmentation results P and the ground truth map G , denoted as $P = \{p_j : j = 1, \dots, N^2\}$ and $G = \{g_j : j = 1, \dots, N^2\}$, and the MS-SSIM loss function [35] for P and G is defined as follows:

$$l_{ms-ssim} = 1 - \prod_{m=1}^M \left(\frac{2u_p u_g + c_1}{u_p^2 + u_g^2 + c_1} \right)^{\beta_m} \left(\frac{2\sigma_{pg} + c_2}{\sigma_p^2 + \sigma_g^2 + c_2} \right)^{\gamma_m} \quad (2)$$

M refers to the number of scales. $u_p, u_g, \sigma_p, \sigma_g$ are the mean and variance of p and g . σ_{pg} refers to the deviation of p and g . The importance of two components for each scale is defined by β_m and γ_m . Add two constants $c_1 = 0.01$ and $c_2 = 0.03$ to avoid unstable situations of being divided by zero. In our experiment, the scale was set to 5, according to [35].

By combining focus loss (l_{fl}), MS-SSIM loss ($l_{ms-ssim}$), and IoU (l_{iou}), we develop a hybrid loss function for segmentation at three levels, including pixel level, region level, and map level. This loss function can capture large-scale and fine structures with clear boundaries. The hybrid loss function [35] is as follows.

$$l = l_{fl} + l_{ms-ssim} + l_{iou} \quad (3)$$

2. A Class-specific UNet 3+ method

MSSH can learn multi-scale features from training samples and have higher accuracy compared to single-scale sampling methods [43,44]. In this paper, we apply the MSSH method to the UNet 3+ method to expand the multi-scale sample data and establish a more robust deep learning model. The flowchart of a class-specific UNet 3+ MSSH method is shown in Figure 7.

As shown in Figure 7, the proposed method involves incorporating MSSH into generating various sample combinations. These combinations are then fed into multiple UNet 3+ networks to detect changes in different types of land cover such as building, road, bare soil, vegetation, water, and concrete. The error of cross-validation of training samples is used to evaluate the performance of these networks. In the case of MSSH and RF, the minimum

out-of-pocket error is used as a constraint condition [43]. For GBDT, Adaboost, and SVM, the error of cross-validation of training samples is used instead of out-of-pocket error since these methods do not have out-of-pocket errors [44]. The optimal classifiers for GBDT, Adaboost, and SVM are selected based on the minimal cross-validation error. Similarly, in the case of UNet 3+, the error of cross-validation of training samples is used since this method also does not have an out-of-pocket error. When the error of cross-validation is minimal, the corresponding UNet 3+ model is considered optimal for detecting changes in buildings, roads, bare soil, vegetation, water, and concrete.

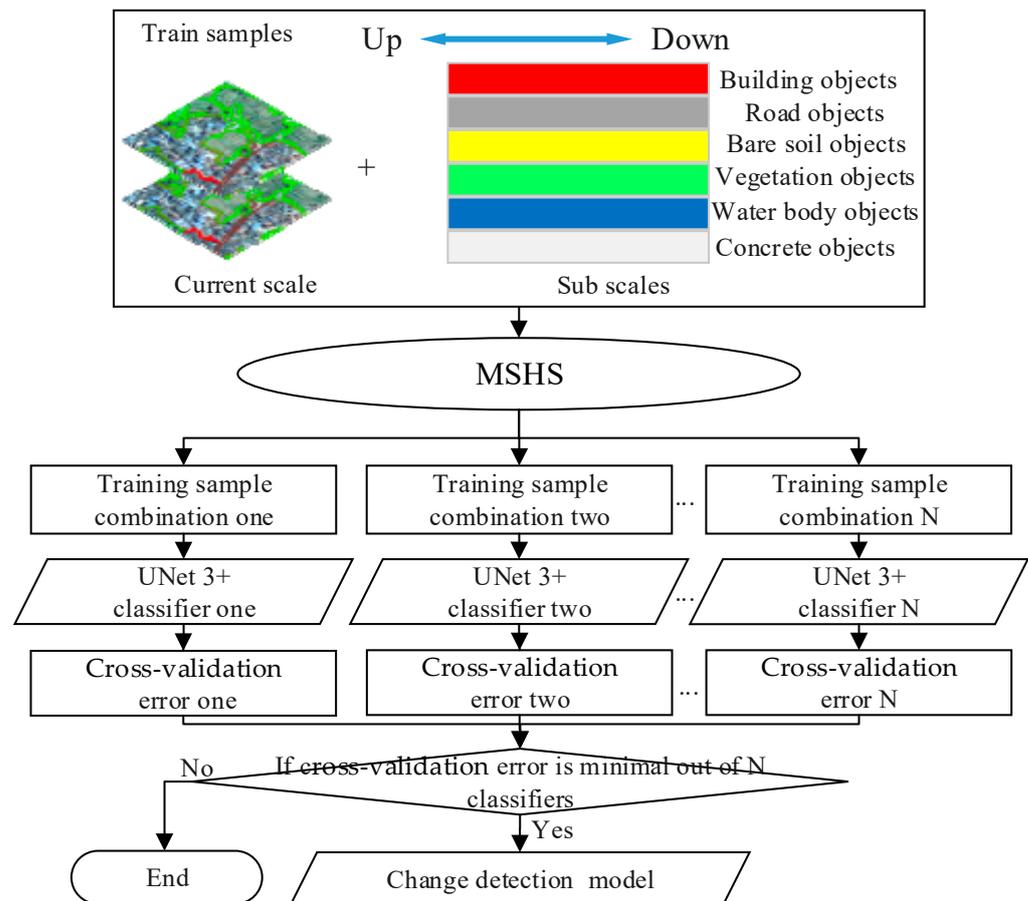


Figure 7. The flow chart of the class-specific UNet 3+ method.

Comparing the model structure of MSHS method using UNet 3+ with the model structure of MSHS methods using RF, SVM, GBDT, and Adaboost classifiers, we found that the former is more concise and can directly extract full-scale features without the need for traditional object-oriented feature extraction steps.

2.2.3. An Object-Oriented Change Detection Method Based on UNet 3+

Step 1: Overlay bi-temporal images after registration and relative radiometric correction to form a new image, and then use a multi-scale segmentation method and a historical land-cover map to segment and obtain the current scale object layers and sub-scale object layers of the bi-temporal images. According to the land-cover map, we obtain the corresponding land-cover types for each object and then use bilinear interpolation to adjust the size of irregular objects to a regular rectangle of 256×256 pixels.

Step 2: For building, road, bare soil, vegetation, water, and concrete objects, we perform MSHS separately. We distribute the training samples evenly across different scales and assign labels to each object based on a reference change/no change pixel map. Finally, we combine current-scale samples with sub-scale samples for each type of object.

Step 3: We feed the combinations of training samples and their corresponding change types into UNet 3+. Furthermore, we train several UNet 3+ classifiers and choose the most suitable change detection model for buildings, roads, bare soil, vegetation, water, and concrete. Our selection is based on the constraint condition of minimizing the cross-validation error of the verification data.

Step 4: The optimal change detection model is utilized to identify changes in the multi-scale object layers $\{L1, L2, \dots, LN\}$ of the images captured during two different periods, obtaining pixel-to-pixel change detection results of each building, road, bare soil, vegetation, water, and concrete object at multiple scales.

To obtain the change detection result of the object, it is necessary to convert the pixel-to-pixel result into an object result. Therefore, the change rate parameter is proposed to determine whether the object has changed. The formula for the change rate and the object-oriented change detection result is as follows:

$$\begin{cases} OC = 1, & \frac{\text{Change}}{\text{Nochange}} \geq CR \\ OC = 0, & \frac{\text{Change}}{\text{Nochange}} < CR \end{cases} \quad (4)$$

CR is the change rate parameter and is a given parameter value, ranging from 0 to 1, and its size directly affects the final object change detection result. Change is the number of pixels that have changed in the object, and Nochange is the number of pixels that have not changed in the object. OC refers to the object's change detection result, where 1 indicates that the object has changed, and 0 indicates that the object has not changed.

2.2.4. Accuracy Verification

In this paper, we utilize the confusion matrix to calculate the missed detection rates (MDR), false alarm rates (FAR), overall accuracy (OA), and F1 score (F1) [46] to evaluate the accuracy of change detection results. The confusion matrix is shown in Table 1.

Table 1. A confusion matrix for binary change detection based on objects.

Number of Objects		Ground Truth	
		Change	No Change
Change detection result	Change	True positives (<i>TP</i>)	False positive (<i>FP</i>)
	No change	False negatives (<i>FN</i>)	True negatives (<i>TN</i>)

TP represents the objects that change in the ground truth map, and the classification method correctly classifies them as the number of changed objects. *TN* represents the objects that do not change in the ground truth map, and the classification method correctly classifies them as unchanged objects. *FP* represents the objects that do not change in the ground truth map but are incorrectly classified as changed objects by the classification method. *FN* represents the objects that change in the ground truth map, but the classification method incorrectly classifies them as unchanged objects.

$$MDR = FN / (TP + FN) \quad (5)$$

$$FAR = FP / (TP + FP) \quad (6)$$

$$OA = (TP + TN) / (TP + FP + FN + TN) \quad (7)$$

$$P = TP / (TP + FP) \quad (8)$$

$$R = TP / (TP + FN) \quad (9)$$

$$F1 = 2P * R / (P + R) \quad (10)$$

The *FAR* indicates the ratio of unchanged objects detected as changed objects in the ground truth map. The *MDR* indicates the ratio of changed objects not correctly detected as

changed objects in the ground truth map. The P refers to the precision. The R refers to the recall. The $F1$ refers to the harmonic mean of precision and recall [47]. The OA indicates the ratio of correctly detected changed and unchanged samples to the total number of changed and unchanged samples, reflecting the overall detection accuracy. The values of the MDR , FAR , $F1$, and OA vary between 0 and 1. The closer the MDR and FAR are to 0, the closer the OA and $F1$ are to 1, indicating a high accuracy of change detection methods.

3. Results

To validate the effectiveness of our proposed method, we conducted experiments using two datasets for change detection. We compared our proposed method with the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, and the UNet 3+ MSHS method. The UNet 3+ single-scale sampling method and the UNet 3+ MSHS method do not consider the scale characteristics of change types of different land cover. Our proposed method and the class-specific UNet 3+ single-scale sampling method both consider scale characteristics of change types of different land cover. This comparison experiment is performed on a laptop equipped with an Intel i9-13980HX CPU and an NVIDIA RTX 4090 Tensor Core GPU card using the PyTorch framework.

3.1. Sampling

In this experiment, 80% of the images in Dataset 1 were used for training, and the remaining 20% of the images in Dataset 1 and Dataset 2 were used as the testing data to calculate the accuracy of the final change detection results. The training and testing processes are completely independent. In the experiment of the optimal DLCD model selection, 75% of the training samples were selected as the training data, and the remaining part of the training samples were used as the verification data to calculate the error of cross-validation.

In the multi-scale segmentation experiment, for Dataset 1 and Dataset 2, we selected five segmentation scales $\{L1, L2, \dots, L5\}$ with a segmentation interval of 40, ranging from scale 40 to 200. The shape parameter was set to 0.3, and the compactness parameter was set to 0.5. During sampling, the training objects at five scales need to be kept in the same position. The random polygon samples and the object layer at five scales are stacked to obtain the training objects. We selected the object with the most pixels superimposed on the random polygon sample as the training object. After determining the training objects at level 5, the training objects at levels 1–4 are determined through the hierarchical relationship between multi-scale layers. The label of each training object is determined based on the pixel ground truth map.

3.2. Model Training and Testing

We performed a comprehensive analysis of the selection of training samples for single-scale sampling methods at scales 40, 80, 120, 160, and 200. Furthermore, we calculated the maximum number of training samples using MSHS for buildings, roads, bare soil, vegetation, water, and concrete. To analyze the accuracy of the trained model, we evaluated the loss value for training samples at iteration 0, 5, and 10. This analysis is presented in Table 2.

Table 2. The number of training samples for the single-scale sampling method at scales 40, 80, 120, 160, and 200, the maximum number of training samples using MSHS, and the loss value of our trained model at iterations 0, 5, and 10 for buildings, roads, bare soil, vegetation, water, and concrete.

Land Cover	Single-Scale Sampling					MSHS	Loss at Iteration 0	Loss at Iteration 5	Loss at Iteration 10
	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200				
Building	5475	1523	734	427	311	8470	0.1512	0.1200	0.1103
Road	1212	410	213	144	78	2057	0.2375	0.1899	0.1719
Bare soil	3166	875	440	255	182	4918	0.2464	0.2099	0.1940
Vegetation	6828	1656	787	462	313	10,046	0.1477	0.1161	0.1067
Water	882	256	136	84	48	1406	0.1879	0.1379	0.1165
Concrete	1680	508	268	169	111	2736	0.1665	0.1307	0.1197

From Table 2, the proposed MSHS has more training samples compared to the single-scale sampling method. In addition, after iterating the UNet 3+ change detection model from 5 to 10 times, the loss values of the change detection models for buildings, vegetation, water, and concrete are similar and are both approximately 0.1, which are considered suitable for the next step of change detection. For roads and bare soil, the loss values of the change detection models from iteration 5 to 10 are similar and both are less than 0.2. Therefore, they are also used for the next step of change detection.

The proposed DLCD model trained on 80% of the pixels in Dataset 1 was used to detect the remaining 20% of pixels in Dataset 1. Comparative experiments were conducted to compare our proposed method with the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, and the UNet 3+ MSHS method. The change detection results of the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, the UNet 3+ MSHS method, and our proposed method at the scale 200 for Dataset 1 are shown in Figure 8.

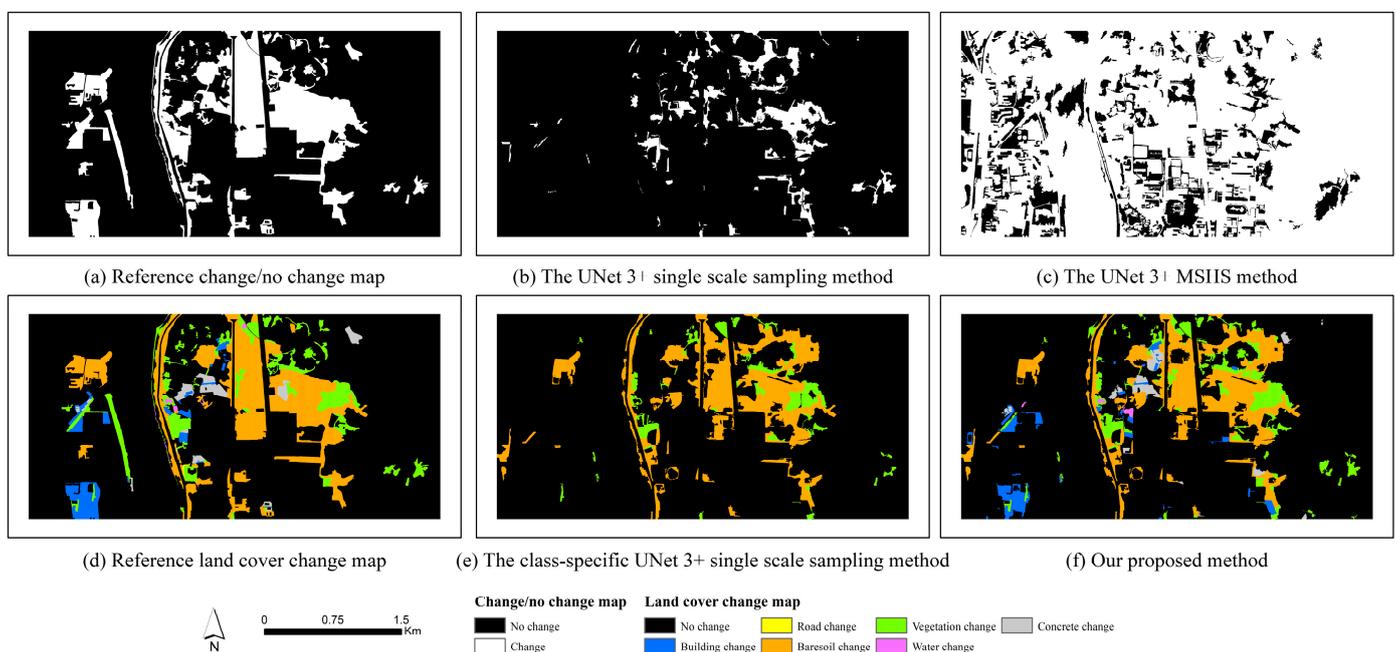


Figure 8. Reference change/no change map, reference land-cover change map, and experimental results of the UNet 3+ single sampling method, the class-specific UNet 3+ single-scale sampling method, the UNet 3+ MSHS method, and our proposed method at the scale 200 for Dataset 1.

From Figure 8, the change detection results of the class-specific UNet 3+ single-scale sampling method and our proposed method have significantly improved visual accuracy compared to the results of the UNet 3+ single sampling method and the UNet 3+ MSHS method. This illustrates the importance of considering the scale characteristics of different types of land-cover change. In addition, by comparing our proposed method with the class-specific UNet 3+ single-scale sampling method, our proposed method could detect more change in types of land cover. This illustrates the effectiveness of the MSHS method.

The FAR, MDR, OA, and F1 values of change detection results using the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, the UNet 3+ MSHS method, and our proposed method for Dataset 1 are calculated as shown in Table 3.

From Table 3, the OA and F1 values of the change detection results for the class-specific UNet 3+ single-scale sampling method and our proposed method have been significantly improved compared to the UNet 3+ single-scale sampling method and the UNet 3+ MSHS method. This illustrates the importance of considering the scale characteristics of different types of land-cover change. In addition, comparing our proposed method with the class-

specific UNet 3+ single-scale sampling method, our proposed method has higher OA and F1, and a lower MDR, indicating the effectiveness of the proposed MSHS.

Table 3. Comparative analysis results of accuracy indexes of the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, the UNet 3+ MSHS method, and our proposed method for Dataset 1.

Accuracy	The UNet 3+ Single-Scale Sampling Method	The Class-Specific UNet 3+ Single-Scale Sampling Method	The UNet 3+ MSHS Method	Our Proposed Method
FAR	24.75%	19.29%	79.59%	22.69%
MDR	76.34%	44.41%	19.08%	29.83%
OA	82.85%	88.24%	31.80%	89.72%
F1	36.00%	65.83%	32.60%	73.56%

To conduct an accurate analysis of change detection results in buildings, roads, bare soil, vegetation, water, and concrete using our proposed method and the class-specific UNet 3+ single-scale sampling method, the FAR, MDR, and F1 values of change detection results using these two methods for Dataset 1 were calculated, as shown in Table 4.

Table 4. Comparative analysis results of accuracy indexes of change detection results using our proposed method and the class-specific UNet 3+ single-scale sampling method for Dataset 1.

Accuracy Indexes	FAR		MDR		F1	
	Our Proposed Method	The Class-Specific UNet 3+ Single-Scale Sampling Method	Our Proposed Method	The Class-Specific UNet 3+ Single-Scale Sampling Method	Our Proposed Method	The Class-Specific UNet 3+ Single-Scale Sampling Method
Building	0.29	1	0.35	1	0.68	0
Road	1	1	1	1	0	0
Bare soil	0.21	0.21	0.28	0.28	0.75	0.75
Vegetation	0.18	0.11	0.26	0.51	0.78	0.63
Water	0.74	1	0.67	1	0.29	0
Concrete	0.45	1	0.58	1	0.48	0
No changes	0.07	0.11	0.05	0.03	0.94	0.93

From Table 4, the class-specific UNet 3+ single-scale sampling method only detects changes in bare soil, vegetation, and no changes, while our proposed method can detect changes in buildings, bare soil, vegetation, water, and concrete, which illustrates the fact that our proposed method can detect more change types of land cover, proving the effectiveness of the combination of MSHS and the UNet 3+ method, which considers the scale characteristics inherent in various land-cover change types.

3.3. Experimental Results

To clarify the generalization ability of our proposed method, we used the object-oriented change detection model based on UNet 3+ trained from Dataset 1 to detect changes in buildings, roads, bare soil, vegetation, water, and concrete at scales 40, 80, 120, 160, and 200 for Dataset 2. The change detection results of bare soil and vegetation using our proposed method at scales 40, 80, 120, 160, and 200 for Dataset 2 are shown in Figures 9 and 10.

From Figures 9 and 10, the change detection results of bare soil and vegetation at different scales are visually different. The change detection results at small scales have more false detections, and the large-scale change detection results have more missed detection, indicating the scale's effect on change detection results.

We further analyzed the relationship between the object-oriented change detection results at scales 40, 80, 120, 160, and 200 and the end-to-end pixel change detection results for bare soil and vegetation. The details of change detection results using our proposed

method and the ground truth maps at scales 40, 80, 120, 160, and 200 for bare soil and vegetation objects for Dataset 2 are shown in Figures 11 and 12.

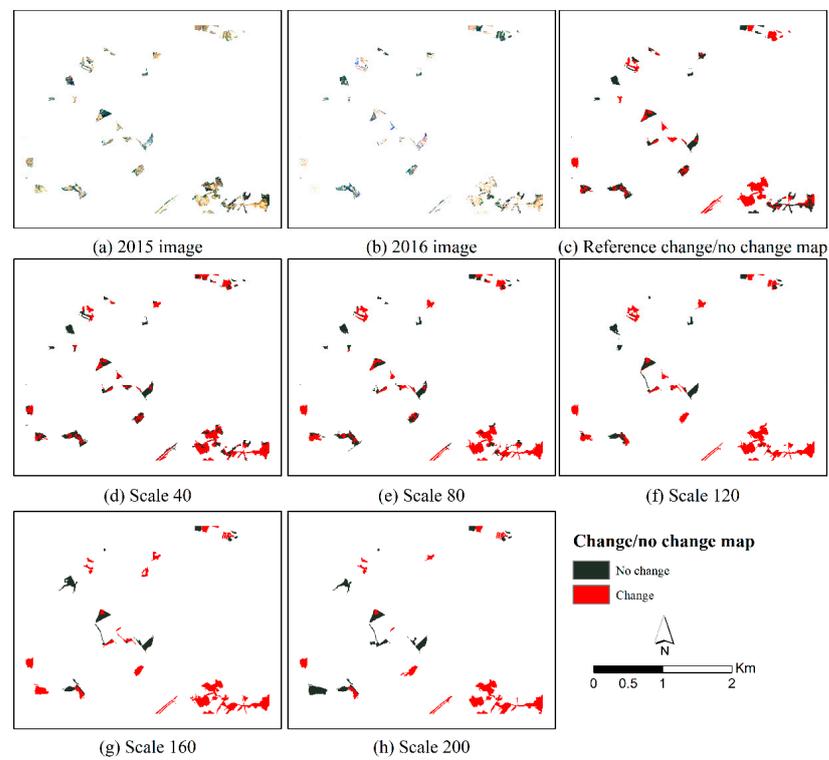


Figure 9. Change detection results of bare soil using our proposed method at scales 40, 80, 120, 160, and 200 for Dataset 2.

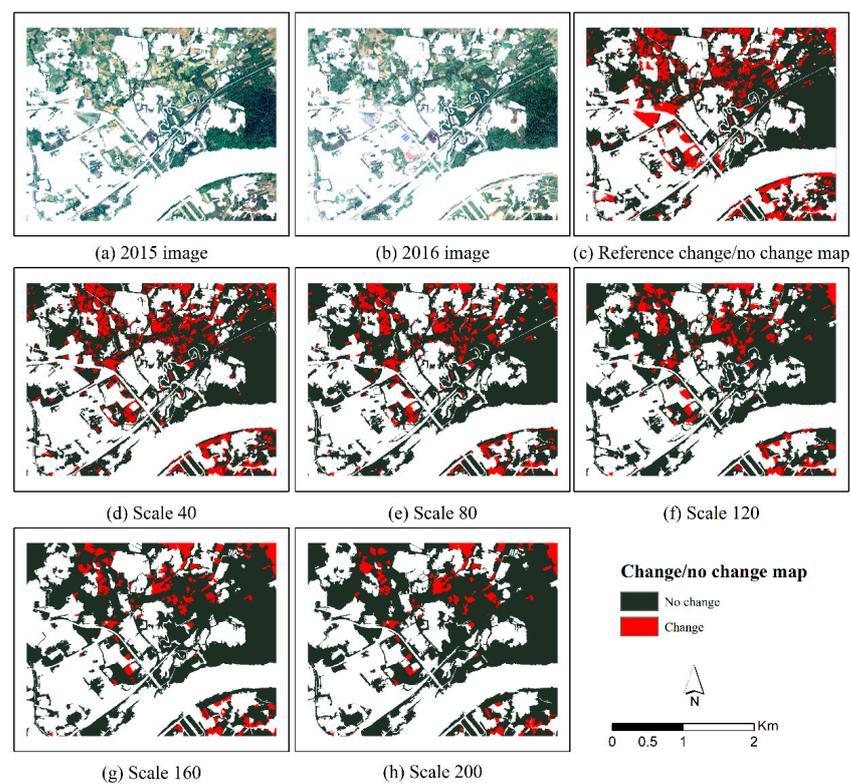


Figure 10. Change detection results of vegetation using our proposed method at scales 40, 80, 120, 160, and 200 for Dataset 2.

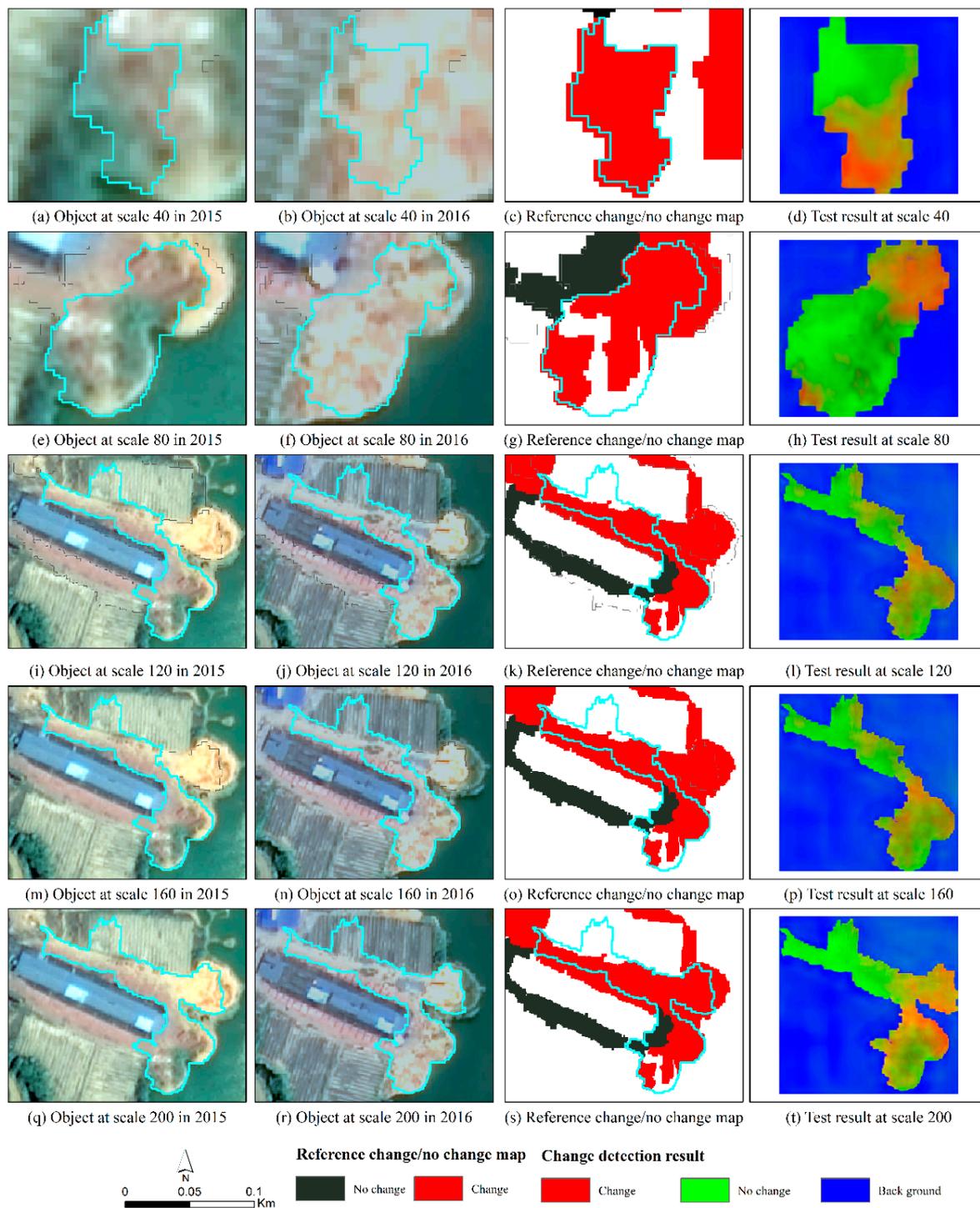


Figure 11. Details of the change detection result of bare soil objects at scales 40, 80, 120, 160, and 200 for Dataset 2. The change detection results of bare soil are displayed in RGB. The red band represents change. The green band represents unchanged, and the blue band represents the background.

From Figures 11 and 12, the object-oriented change detection model based on UNet 3+ can specifically determine whether each pixel in the object has changed, which can effectively consider the edges of the object. Compared to non-end-to-end methods, this method is more accurate. These illustrate that the proposed method fully combines the advantages of pixels and object-oriented methods, which can obtain multi-scale object-based change detection results and pixel-based change maps inside the object.

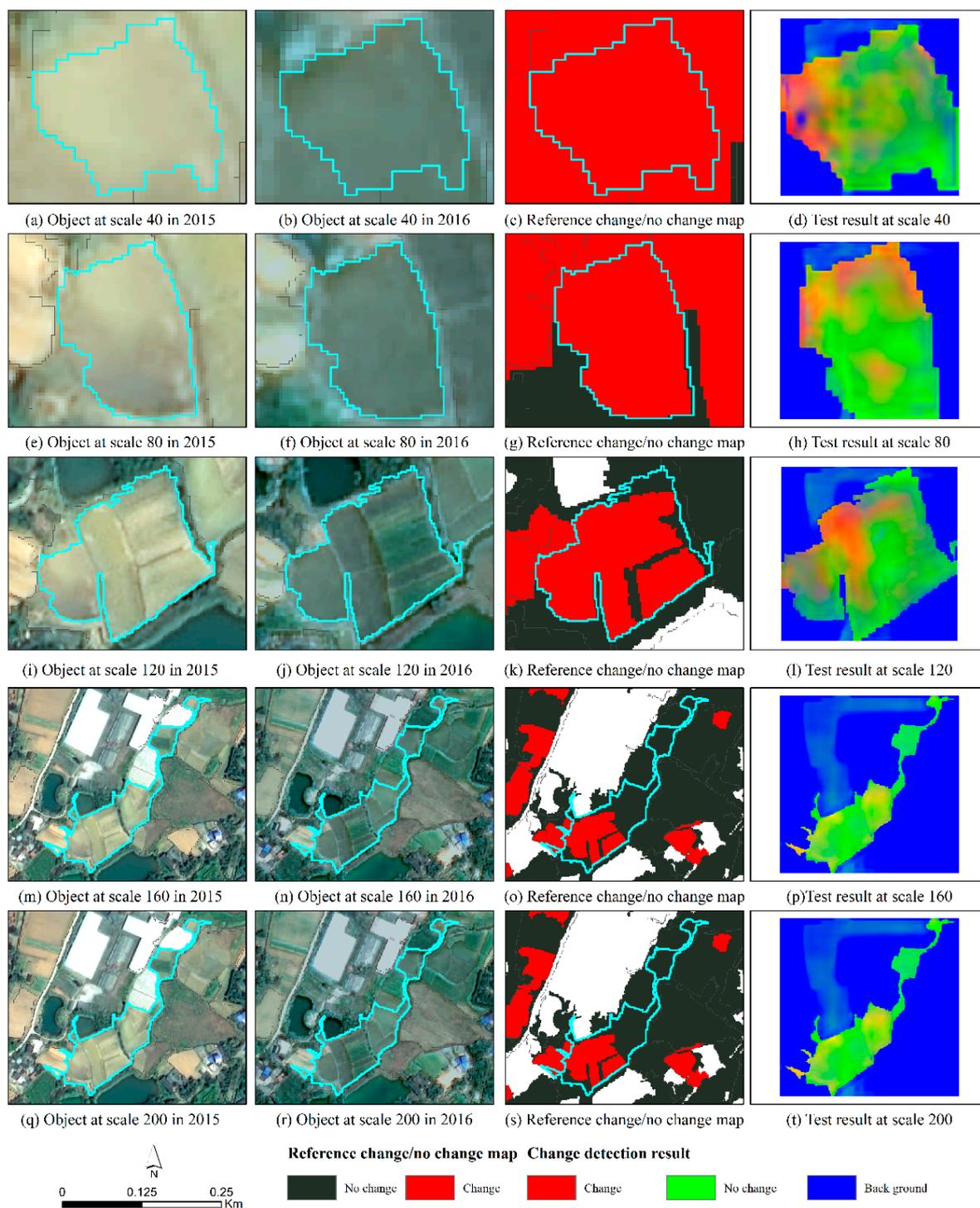


Figure 12. Details of the change detection result of vegetation objects at scales 40, 80, 120, 160, and 200 for Dataset 2. The change detection results of vegetation are displayed in RGB. The red band represents change. The green band represents unchanged, and the blue band represents the background.

3.4. Accuracy Analysis of the Algorithm

To assess the model's ability to generalize, the MDR, FAR, OA, and F1 values of the change detection results of bare soil, vegetation, buildings, concrete, roads, and water using our proposed method at scale 40, 80, 120, 160, and 200 for Dataset 2 were then calculated, and are presented in Tables 5–10.

Table 5. Accuracy indicators for change detection results of bare soil at different scales for Dataset 2.

Accuracy	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200
FAR	19.09%	25.34%	26.41%	30.11%	27.69%
MDR	19.69%	14.54%	11.63%	11.24%	17.70%
OA	77.41%	73.84%	72.94%	70.39%	70.27%
F1	80.61%	79.70%	80.31%	78.20%	76.98%

Table 6. Accuracy indicators for change detection results of vegetation at different scales for Dataset 2.

Accuracy	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200
FAR	0.17%	12.65%	17.69%	20.97%	20.24%
MDR	16.09%	41.30%	49.92%	56.34%	63.17%
OA	96.73%	89.87%	87.54%	85.98%	85.21%
F1	91.18%	70.21%	62.27%	56.24%	50.39%

Table 7. Accuracy indicators for change detection results of buildings at different scales for Dataset 2.

Accuracy	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200
FAR	37.38%	47.02%	41.70%	34.66%	56.35%
MDR	56.53%	74.84%	78.54%	86.98%	95.97%
OA	91.21%	89.17%	89.48%	89.37%	88.46%
F1	51.32%	34.12%	31.37%	21.71%	7.38%

Table 8. Accuracy indicators for change detection results of concrete at different scales for Dataset 2.

Accuracy	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200
FAR	3.00%	21.02%	32.71%	35.83%	39.67%
MDR	44.06%	59.38%	60.90%	64.97%	70.09%
OA	92.75%	87.96%	85.78%	85.42%	83.46%
F1	70.96%	53.65%	49.46%	45.32%	39.99%

Table 9. Accuracy indicators for change detection results of roads at different scales for Dataset 2.

Accuracy	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200
FAR	100.00%	100.00%	100.00%	100.00%	100.00%
MDR	100.00%	100.00%	100.00%	100.00%	100.00%
OA	99.78%	99.78%	99.78%	99.78%	99.78%
F1	0.00%	0.00%	0.00%	0.00%	0.00%

Table 10. Accuracy indicators for change detection results of water at different scales for Dataset 2.

Accuracy	Scale 40	Scale 80	Scale 120	Scale 160	Scale 200
FAR	61.56%	62.12%	78.49%	93.73%	90.48%
MDR	46.28%	54.18%	86.40%	91.14%	90.61%
OA	96.42%	96.61%	96.55%	94.99%	96.00%
F1	44.81%	41.48%	16.66%	7.34%	9.45%

Based on the findings from Tables 5–10, it is evident that the model can provide highly precise change detection outcomes for bare soil and vegetation for Dataset 2. This is because there are more training samples for bare soil and vegetation. The accuracy values of buildings and concrete are not high due to poor projection and shadows. However, the accuracy values of roads and water are comparatively lower. This is because these two change detection models used fewer training samples to train the model, and their changes

in the image were limited. The roads did not change, and the water changes were minimal in the image.

4. Discussion

4.1. Comparison with Other Traditional Change Detection Methods

To evaluate the effectiveness of our proposed method, the binary change detection results are compared with the PCA-k-means [48], the object-oriented change vector analysis method [49], the single-scale sampling method based on random forest, and the class-specific object-based method [44] for Dataset 2. The object-oriented change vector analysis method, the single-scale sampling method based on random forest, and our proposed method all perform change detection at scale 40. The comparison results are shown in Figure 13.

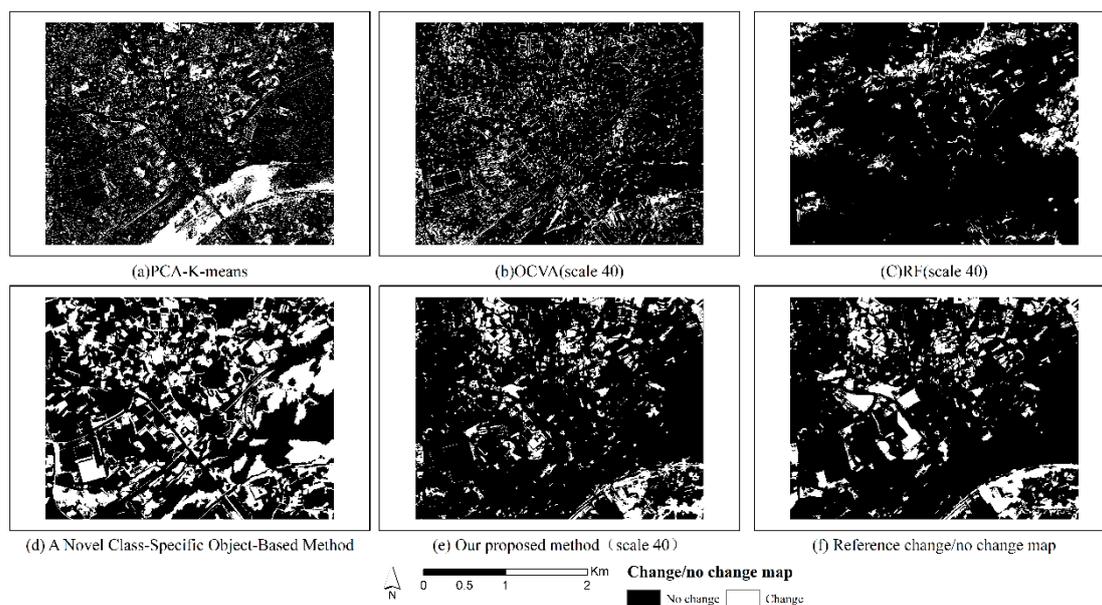


Figure 13. Comparison of change detection results and reference change/no change map between the PCA-k-means, the object-oriented change vector analysis method, the single-scale sampling method based on random forest, the class-specific object-based method, and our proposed method for Dataset 2.

As shown in Figure 13, visually, the proposed method has fewer false positives and false negatives, so it has higher visual accuracy compared to the PCA-k-means, the object-oriented change vector analysis, the single-scale sampling method based on random forest, and the class-specific object-based method for Dataset 2.

For Dataset 2, we calculated the FAR, MDR, OA, and F1 values of our proposed method, the PCA-k-means, the object-oriented change vector analysis, the single-scale sampling method based on random forest, and the class-specific object-based method, as shown in Table 11.

As shown in Table 11, for Dataset 2, compared to the pixel-based method (PCA-k-means), our proposed method has higher OA and F1 while having lower false detection, demonstrating the effectiveness of the combination of an object-oriented approach and UNet 3+. Compared to the single-scale method (OCVA and single-scale sampling method based on random forest) and the multi-scale method considering the scale uncertainty (the class-specific object-based method and our proposed method), the latter has higher accuracy. This is because the latter can consider the scale characteristics of different land-cover change types, which is not only close to the true expression of ground objects but also more targeted. Compared to traditional machine learning methods (single-scale sampling method based on random forest and the class-specific object-based method), our proposed method has higher OA and F1 while having lower false detection, demonstrating the effectiveness of the UNet 3+ model. Compared to the class-specific object-based method,

our proposed method has higher OA and F1, and lower false detection, indicating the effectiveness of the combination of the object-oriented method, MSHS, and UNet 3+ model.

Table 11. Accuracy comparison analysis results of the PCA-k-means, the object-oriented change vector analysis method, the single-scale sampling method based on random forest, the class-specific object-based method, and our proposed method for Dataset 2.

Dataset 2	FAR (%)	MDR (%)	OA (%)	F1 (%)
PCA-K-means	85.99	64.77	77.88	20.05
OCVA (scale 40)	76.92	82.63	76.98	19.82
RF (scale 40)	86.10	92.53	77.26	9.72
A class-specific object-based method	45.71	50.94	93.40	51.54
Our proposed method (scale 40)	9.48	25.00	94.62	82.04

4.2. Comparison with Other DLCD Methods

To evaluate the effectiveness of our proposed method, the binary change detection results are compared with LamboiseNet (light UNet++) [50], BIT [51], CDNet [52], FCSiamConc [53], and FCSiamDiff [53]. We use the default parameters from the literature for the LamboiseNet (light UNet++), BIT, CDNet, FCSiamConc, and FCSiamDiff. LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff are region-based network algorithms (256×256). Therefore, when performing change detection, edges less than 256 are discarded. The corresponding results are shown in Figure 14.

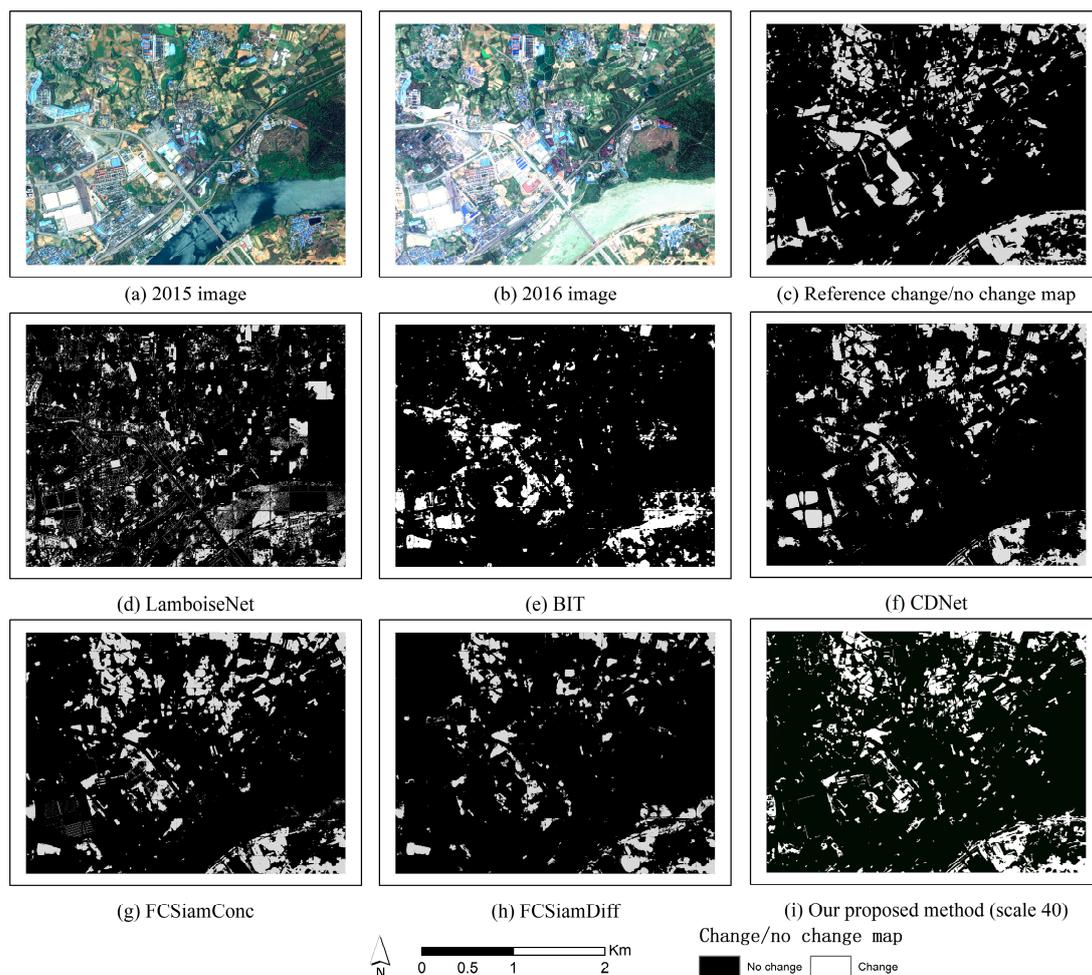


Figure 14. Comparison of change detection results between the LamboiseNet, BIT, CDNet, FCSiamConc, FCSiamDiff, and our proposed method for Dataset 2.

As shown in Figure 14, visually, the proposed method has higher accuracy compared to the LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff methods. The LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff methods have a high false negative rate and poor object edges, indicating that this method considers the advantages of the object-oriented method for detecting edge information of land-cover change.

For Dataset 2, we calculated the FAR, MDR, OA, and F1 values of our proposed method, LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff, as shown in Table 12.

Table 12. Accuracy comparison analysis results of LamboiseNet, BIT, CDNet, FCSiamConc, FCSiamDiff, and our proposed method for Dataset 2.

Dataset 2	FAR (%)	MDR (%)	OA (%)	F1 (%)
LamboiseNet	75.82	78.19	75.94	22.93
BIT	42.71	63.33	85.12	44.72
CDNet	38.21	50.59	86.68	54.91
FCSiamConc	34.97	41.72	88.01	61.47
FCSiamDiff	27.50	58.46	87.82	52.82
Our proposed method (scale 40)	9.48	25.00	94.62	82.04

As shown in Table 12, for Dataset 2, compared to the LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff, our proposed method has higher OA and F1, and lower false detection, indicating the effectiveness of DLCD method considering the scale characteristics of different land-cover change types.

We ran LamboiseNet, BIT, CDNet, FCSiamConc, FCSiamDiff, and our proposed method on a laptop equipped with an Intel i9-13980HX CPU and an NVIDIA RTX 4090 Tensor Core GPU card using the PyTorch framework for comparison experiments. Training time for each training sample of these five DLCD methods is shown in Table 13.

Table 13. Training time for each training sample of LamboiseNet, BIT, CDNet, FCSiamConc, FCSiamDiff, and our proposed method for Dataset 2.

Methods	Training Time for Each Training Sample (S)	Average Time for Each Training Sample (S)
LamboiseNet	4.2	4.2
BIT	0.96	0.96
CDNet	1.21	1.21
FCSiamConc	0.88	0.88
FCSiamDiff	0.95	0.95
Our proposed method (building)	1.23	
Our proposed method (road)	1.22	
Our proposed method (bare soil)	1.14	
Our proposed method (vegetation)	1.24	1.21
Our proposed method (water)	1.18	
Our proposed method (concrete)	1.22	

From Table 13, compared to other DLCD methods, the average training time for each training sample of our proposed method for buildings, roads, base soil, vegetation, water, and concrete is about 1.21 s, which is an acceptable time consumption.

4.3. Sensitivity of the Algorithm

To test the impact of scale on the proposed method, the FAR, MDR, OA, and F1 values of the change detection results of buildings, roads, bare soil, vegetation, water, and concrete at scales 40–200 (with an interval of 40) for Dataset 2 were calculated for a change rate of 0.05, as shown in Figure 15. To test the impact of change rate on the proposed method, the FAR, MDR, OA, and F1 values of the change detection results of buildings, roads, bare soil,

vegetation, water, and concrete at scale 40 with a change rate of 0.05–0.4 (interval of 0.05) for Dataset 2 were calculated, as shown in Figure 16.

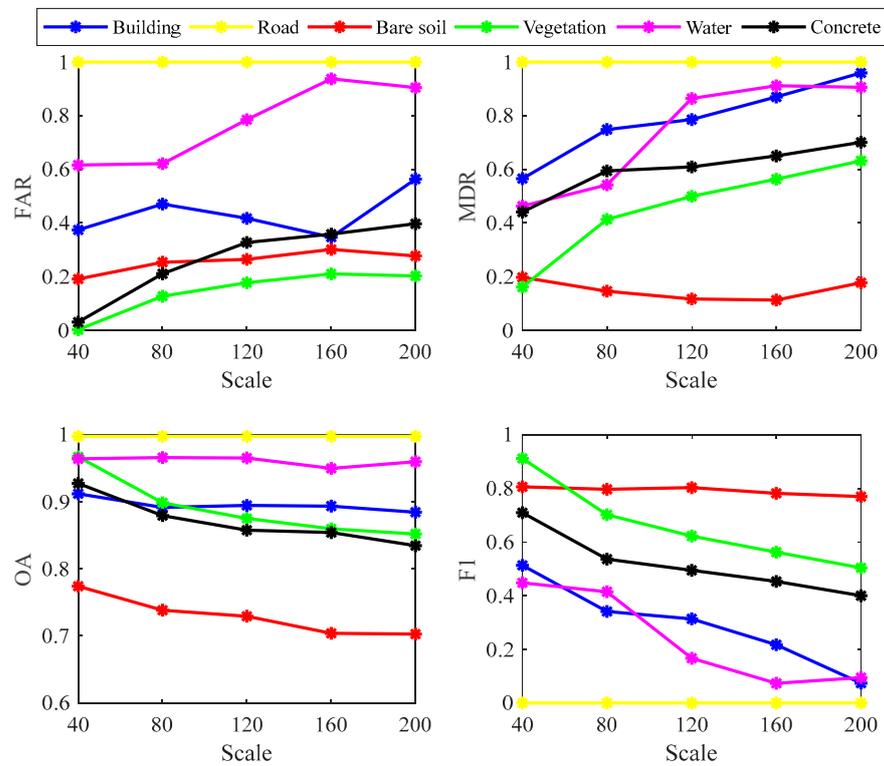


Figure 15. The FAR, MDR, OA, and F1 values of the change detection results of buildings, roads, bare soil, vegetation, water, and concrete at scales 40–200 (with an interval of 40) for Dataset 2.

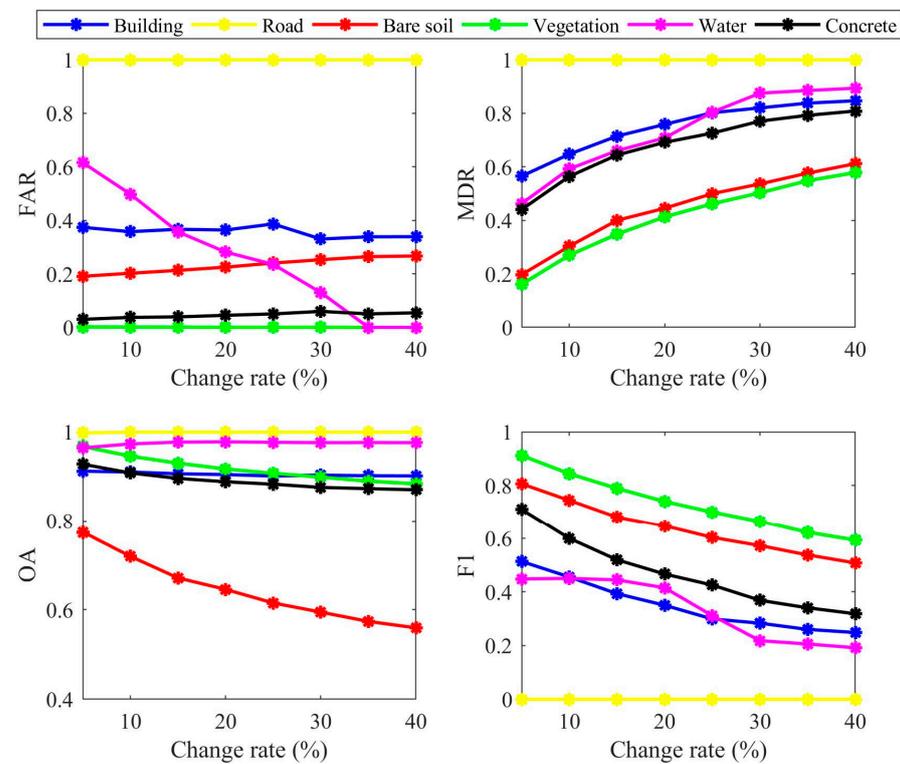


Figure 16. The FAR, MDR, OA, and F1 of the change detection results of buildings, roads, bare soil, vegetation, water, and concrete at scale 40 with a change rate of 0.05–0.4 (interval of 0.05) for Dataset 2.

From Figure 15, the scale has a significant impact on the MDR, FAR, OA, and F1 values of buildings, bare soil, vegetation, water, and concrete. For vegetation and concrete, the MDR shows an upward trend with the increase in scale, and the OA and F1 show a downward trend, indicating that a small scale is suitable for the change detection of vegetation and concrete. For buildings, the MDR shows an upward trend, and the OA shows a downward and upward trend first, and then downward, while F1 shows a downward trend. This indicates that a small scale is suitable for the change detection of buildings. In the future, smaller scales can be considered for change detection of these land-cover types. For bare soil, the MDR shows a downward and upward trend, reaching the lowest at scale 120, and the OA shows a downward trend, while F1 shows a downward and upward trend first, and then downward. For water, the MDR shows an upward and downward trend, and the OA shows an upward trend and downward first, and then upward, while F1 shows a downward and upward trend. For buildings, the FAR shows an upward and downward trend first, and then upward. For bare soil and vegetation, the FAR shows an upward trend and a downward trend. For water, the FAR shows an upward and downward trend. For concrete, the FAR shows an upward trend. These FAR trends indicate that there is no uniform rule for the impact of scale on the FAR of various ground objects. It needs to be determined according to the specific situation in practical application. However, for roads, due to too few changes, it is impossible to effectively draw laws on the MDR, FAR, OA, and F1. It can be inferred from the rules of other ground objects that when there are as many training samples of roads as other features, the change detection accuracy of the road is easily affected by the scale.

As shown in Figure 16, the change rate has a significant impact on the MDR and F1 of buildings, bare soil, vegetation, water, and concrete. Except for roads and water, the MDR of other ground objects gradually increases with the increase in the change rate, while the OA and F1 gradually decrease with the increase in the change rate, indicating that it is appropriate to choose a small change rate when using this method for change detection. For the FAR, the change rate has little impact. For roads and water, due to too little change, it cannot be effectively detected. Based on the rules of other ground objects, it seems that when there are as many training samples of roads as other features, the accuracy of change detection on roads can be significantly impacted by change rates.

In this paper, we focus on comparing the effects of different scales and change rates on the accuracy of surface objects, without obtaining the optimal scale and change rate for each land-cover change type. In the future, we plan to combine the multi-scale feature expression, optimal scale selection, and optimal change rate selection method for efficient change detection. The model's accuracy may suffer when there are fewer training samples or fewer changes for ground objects on remote sensing images during change detection for objects such as roads and water. In the future, we will add prior information for roads and water bodies, such as water body index, road shape index, etc., so that the model can detect a small number of changes.

In addition, we employed the bilinear interpolation method, a commonly used interpolation method, to transform irregular objects into regular objects, which retains the edge information of the object and meets the input requirements of the deep learning model. However, this method has an impact on the quality of training and testing objects, which, in turn, affects the performance of the deep learning model. To reduce the impact of the bilinear interpolation method on the deep learning model, we used an MSHS method for fitting and used the same interpolation method during both the training and testing phases. The study of interpolation methods on the impact of objects is an important research direction for future improvements in our proposed method. As a result, we plan to explore the impact of various interpolation methods on the performance of DLCD models in the future. Our proposed method relies on RGB bands, potentially resulting in the loss of valuable information from other spectral bands. In future work, we plan to explore the integration of additional band information to fully exploit the spectral characteristics of ground objects for more effective change detection. Moreover, errors in classification maps

generated by standard post-classification procedures may accumulate. Hence, we chose a direct classification method for change detection in this study. Nonetheless, we plan to explore the possibility of employing post-processing classification methods to leverage prior information and further enhance the efficacy of our approach in future research.

5. Conclusions

In this paper, for the scale issues in feature description and change detection models, we propose a novel UNet 3+ change detection method considering scale uncertainty for land-cover change detection. Our proposed method comprises three main steps: a multi-scale segmentation method, the class-specific UNet 3+ method, and the object-oriented change detection method based on UNet 3+. To validate the effectiveness of our proposed method, we employed the model trained on Dataset 1 to validate both the remaining parts of Dataset 1 and Dataset 2. Our proposed method was compared against the UNet 3+ single-scale sampling method, the class-specific UNet 3+ single-scale sampling method, and the UNet 3+ MSHS method. The results indicate that our approach achieves higher OA and F1 while exhibiting lower MDR and FAR. Furthermore, our proposed method demonstrates the ability to detect more changes in land cover. To assess the scalability of our proposed method, we compared it with traditional change detection methods such as PCA-k-means, OCVA, the single-scale sampling method based on random forest, and the class-specific object-based method. Our experimental results and accuracy demonstrate that our proposed method considers the scale characteristics of different land cover change types using the UNet 3+ model, resulting in higher accuracy. Additionally, we compared our proposed method with other DLCD methods including LamboiseNet, BIT, CDNet, FCSiamConc, and FCSiamDiff. The results show that our proposed method effectively considers edge information and has an acceptable time consumption. In summary, our proposed method considers the scale characteristics inherent in various land-cover change types and builds specific change detection models for each land-cover type, which better considers edge information. However, the accuracy of our proposed method may be influenced by scale, change rate, and interpolation method. In addition, our proposed method shows remarkable performance on bare soil and vegetation, but it shows suboptimal performance for roads and water, which undergo fewer changes, and for buildings and concrete due to poor projection and shadows. Therefore, in the future, we will introduce more geoscience knowledge to improve our proposed method, and establish a knowledge- and data-driven DLCD model.

Author Contributions: Conceptualization, T.B.; methodology, T.B.; data, S.D. and P.L.; validation, Q.A., S.D., T.B. and H.Z.; investigation, T.B., S.D. and Y.C.; writing—original draft preparation, T.B.; review and editing, T.B. and Y.C.; visualization, T.B., S.D., P.L., Z.S. and K.S.; supervision, Y.C. and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (no. 42301457, no. 42192583, and no. 42301434).

Data Availability Statement: The data used in this study can be accessed by contacting the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The authors would like to thank the China Surveying and Mapping Department for providing the historical land-cover maps, and change/no change maps.

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

References

1. Ding, L.; Zhang, J.; Guo, H.; Zhang, K.; Liu, B.; Bruzzone, L. Joint spatio-temporal modeling for semantic change detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [[CrossRef](#)]
2. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 924–935. [[CrossRef](#)]
3. Rao, A.; Jung, J.; Silva, V.; Molinaro, G.; Yun, S.-H. Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning. *Nat. Hazards Earth Syst. Sci.* **2023**, *23*, 789–807. [[CrossRef](#)]

4. Gao, S.; Sun, K.; Li, W.; Li, D.; Tan, Y.; Wei, J.; Li, W. A building change detection framework with patch-pairing single-temporal supervised learning and metric guided attention mechanism. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *129*, 103785. [[CrossRef](#)]
5. Wei, J.; Sun, K.; Li, W.; Li, W.; Gao, S.; Miao, S.; Zhou, Q.; Liu, J. Robust change detection for remote sensing images based on temporospatial interactive attention module. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103767. [[CrossRef](#)]
6. Bai, T.; Wang, L.; Yin, D.; Sun, K.; Chen, Y.; Li, W.; Li, D. Deep learning for change detection in remote sensing: A review. *Geo-Spat. Inf. Sci.* **2023**, *26*, 262–288. [[CrossRef](#)]
7. Ouma, Y.O.; Tateishi, R. A fast environmental change detection approach based on unsupervised multiscale texture clustering. *Int. J. Environ. Stud.* **2005**, *62*, 79–93. [[CrossRef](#)]
8. Bovolo, F.; Bruzzone, L. A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2963–2972. [[CrossRef](#)]
9. Eklund, P.W.; You, J.; Deer, P. Mining remote sensing image data: An integration of fuzzy set theory and image understanding techniques for environmental change detection. In Proceedings of the Data Mining and Knowledge Discovery: Theory, Tools, and Technology II, Orlando, FL, USA, 6 April 2000; pp. 265–272.
10. Wang, X.; Liu, S.; Du, P.; Liang, H.; Xia, J.; Li, Y. Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. *Remote Sens.* **2018**, *10*, 276. [[CrossRef](#)]
11. Wang, P. Research on image segmentation method based on multi-scale theory. Ph.D. Thesis, University of Science and Technology of China, Hefei, China, 2007.
12. Huang, Z. Research on Multiscale Methods in Object-Based Image Analysis. Ph.D. Thesis, National University of Defense Technology, Changsha, China, 2014.
13. Feng, W.; Sui, H.; Tu, J.; Huang, W.; Xu, C.; Sun, K. A novel change detection approach for multi-temporal high-resolution remote sensing images based on rotation forest and coarse-to-fine uncertainty analyses. *Remote Sens.* **2018**, *10*, 1015. [[CrossRef](#)]
14. Zheng, Z.; Cao, J.; Lv, Z.; Benediktsson, J.A. Spatial-Spectral Feature Fusion Coupled with Multi-Scale Segmentation Voting Decision for Detecting Land Cover Change with VHR Remote Sensing Images. *Remote Sens.* **2019**, *11*, 1903. [[CrossRef](#)]
15. Zhou, Y.N.; Li, J.; Feng, L.; Zhang, X.; Hu, X. Adaptive Scale Selection for Multiscale Segmentation of Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3641–3651. [[CrossRef](#)]
16. Grybas, H.; Melendy, L.; Congalton, R.G. A comparison of unsupervised segmentation parameter optimization approaches using moderate-and high-resolution imagery. *GIScience Remote Sens.* **2017**, *54*, 515–533. [[CrossRef](#)]
17. Gong, M.; Zhan, T.; Zhang, P.; Miao, Q. Superpixel-based difference representation learning for change detection in multispectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2658–2673. [[CrossRef](#)]
18. Bengio, Y. Deep learning of representations for unsupervised and transfer learning. In Proceedings of the ICML Workshop on Unsupervised and Transfer Learning, Washington, DC, USA, 2 July 2011; pp. 17–36.
19. Levien, L.M.; Fischer, C.; Roffers, P.; Maurizi, B.; Suero, J.; Fischer, C.; Huang, X. A machine-learning approach to change detection using multi-scale imagery. In Proceedings of the ASPRS Annual Conference, Portland, OR, USA, 20 May 1999; p. 22.
20. Lu, D.; Mausel, P.; Brondizio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2401. [[CrossRef](#)]
21. Huang, C.; Song, K.; Kim, S.; Townshend, J.R.; Davis, P.; Masek, J.G.; Goward, S.N. Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sens. Environ.* **2008**, *112*, 970–985. [[CrossRef](#)]
22. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195. [[CrossRef](#)]
23. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
24. Lyu, H.; Lu, H.; Mou, L. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection. *Remote Sens.* **2016**, *8*, 506. [[CrossRef](#)]
25. Wang, Q.; Zhang, X.; Chen, G.; Dai, F.; Gong, Y.; Zhu, K. Change detection based on Faster R-CNN for high-resolution remote sensing images. *Remote Sens. Lett.* **2018**, *9*, 923–932. [[CrossRef](#)]
26. Xu, Y.; Xiang, S.M.; Huo, C.L.; Pan, C.H. Change Detection Based on Auto-encoder Model for VHR Images. In Proceedings of the Mippr 2013: Pattern Recognition and Computer Vision, Wuhan, China, 27 October 2013; p. 891902.
27. El Amin, A.M.; Liu, Q.; Wang, Y. Convolutional neural network features based change detection in satellite images. In Proceedings of the First International Workshop on Pattern Recognition, Tokyo, Japan, 11 July 2016; p. 100110W.
28. Gong, M.G.; Niu, X.D.; Zhang, P.Z.; Li, Z.T. Generative Adversarial Networks for Change Detection in Multispectral Imagery. *IEEE Geosci Remote S* **2017**, *14*, 2310–2314. [[CrossRef](#)]
29. Zhang, X.; Cheng, S.; Wang, L.; Li, H. Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [[CrossRef](#)]
30. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
31. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
32. Kim, J.H.; Lee, H.; Hong, S.J.; Kim, S.; Park, J.; Hwang, J.Y.; Choi, J.P. Objects segmentation from high-resolution aerial images using U-Net with pyramid pooling layers. *IEEE Geosci. Remote S* **2018**, *16*, 115–119. [[CrossRef](#)]

33. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Granada, Spain, 20 September 2018; pp. 3–11.
34. Peng, D.; Zhang, Y.; Guan, H. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [[CrossRef](#)]
35. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4 May 2020; pp. 1055–1059.
36. Mo, J.; Seong, S.; Oh, J.; Choi, J.J.I.A. SAUNet3+ CD: A Siamese-attentive UNet3+ for change detection in remote sensing images. *IEEE Access* **2022**, *10*, 101434–101444. [[CrossRef](#)]
37. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GIScience Remote Sens.* **2018**, *55*, 243–264. [[CrossRef](#)]
38. Liu, T.; Yang, L.; Lunga, D. Change detection using deep learning approach with object-based image analysis. *Remote Sens. Environ.* **2021**, *256*, 112308. [[CrossRef](#)]
39. Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A.; Zhang, L. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. *Remote Sens. Environ.* **2021**, *265*, 112636. [[CrossRef](#)]
40. Townshend, J.R.; Justice, C.O.; Gurney, C.; McManus, J. The impact of misregistration on change detection. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 1054–1060. [[CrossRef](#)]
41. Pu, R.; Landry, S. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sens. Environ.* **2012**, *124*, 516–533. [[CrossRef](#)]
42. Ma, L.; Li, M.; Blaschke, T.; Ma, X.; Tiede, D.; Cheng, L.; Chen, Z.; Chen, D. Object-based change detection in urban areas: The effects of segmentation strategy, scale, and feature space on unsupervised methods. *Remote Sens.* **2016**, *8*, 761. [[CrossRef](#)]
43. Bai, T.; Sun, K.; Deng, S.; Li, D.; Li, W.; Chen, Y. Multi-scale hierarchical sampling change detection using Random Forest for high-resolution satellite imagery. *Int. J. Remote Sens.* **2018**, *39*, 7523–7546. [[CrossRef](#)]
44. Bai, T.; Sun, K.; Li, W.; Li, D.; Chen, Y.; Sui, H. A novel class-specific object-based method for urban change detection using high-resolution remote sensing imagery. *Photogramm. Eng. Remote Sens.* **2021**, *87*, 249–262. [[CrossRef](#)]
45. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
46. Lunetta, R.S.; Johnson, D.M.; Lyon, J.G.; Crotwell, J. Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote Sens. Environ.* **2004**, *89*, 444–454. [[CrossRef](#)]
47. Beitzel, S.M. *On Understanding and Classifying Web Queries*; Illinois Institute of Technology: Chicago, IL, USA, 2006.
48. Celik, T. Unsupervised change detection in satellite images using principal component analysis and k -means clustering. *IEEE Geosci. Remote S* **2009**, *6*, 772–776. [[CrossRef](#)]
49. Sun, K.; Chen, Y. The Application of Objects Change Vector Analysis in Object-level Change Detection. In Proceedings of the International Conference on Computational Intelligence and Industrial Application (PACIIA), Wuhan, China, 6–7 November 2010; pp. 383–389.
50. Baudhuin, H.; Lambot, A. Change Detection in Satellite Imagery Using Deep Learning. Master’s Thesis, Université Catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium, 2020.
51. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [[CrossRef](#)]
52. Alcantarilla, P.F.; Stent, S.; Ros, G.; Arroyo, R.; Gherardi, R.J.A.R. Street-view change detection with deconvolutional networks. *Auton. Robot.* **2018**, *42*, 1301–1322. [[CrossRef](#)]
53. Daudt, R.C.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067.

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.