Article

Uncertainty Analysis for Object-Based Change Detection in Very High-Resolution Satellite Images Using Deep Learning Network

Ahram Song 1, Yongil Kim 1 and Youkyung Han 2,*

1 Department of Civil and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea; aram200@snu.ac.kr (A.S.); yik@snu.ac.kr (Y.K.)
2 School of Convergence and Fusion System Engineering Kyungpook National University, Sangju 37224, Korea
* Correspondence: han602@knu.ac.kr

Received: 8 July 2020; Accepted: 20 July 2020; Published: 22 July 2020

Abstract: Object-based image analysis (OBIA) is better than pixel-based image analysis for change detection (CD) in very high-resolution (VHR) remote sensing images. Although the effectiveness of deep learning approaches has recently been proved, few studies have investigated OBIA and deep learning for CD. Previously proposed methods use the object information obtained from the preprocessing and postprocessing phase of deep learning. In general, they use the dominant or most frequently used label information with respect to all the pixels inside an object without considering any quantitative criteria to integrate the deep learning network and object information. In this study, we developed an object-based CD method for VHR satellite images using a deep learning network to denote the uncertainty associated with an object and effectively detect the changes in an area without the ground truth data. The proposed method defines the uncertainty associated with an object and mainly includes two phases. Initially, CD objects were generated by unsupervised CD methods, and the objects were used to train the CD network comprising three-dimensional convolutional layers and convolutional long short-term memory layers. The CD objects were updated according to the uncertainty level after the learning process was completed. Further, the updated CD objects were considered as the training data for the CD network. This process was repeated until the entire area was classified into two classes, i.e., change and no-change, with respect to the object units or defined epoch. The experiments conducted using two different VHR satellite images confirmed that the proposed method achieved the best performance when compared with the performances obtained using the traditional CD approaches. The method was less affected by salt and pepper noise and could effectively extract the region of change in object units without ground truth data. Furthermore, the proposed method can offer advantages associated with unsupervised CD methods and a CD network subjected to postprocessing by effectively utilizing the deep learning technique and object information.

Keywords: change detection; object-based image analysis; very high-resolution satellite images; deep learning network

1. Introduction

Object-based image analysis (OBIA) involves the segmentation of an image based on clusters of similar neighboring pixels exhibiting common properties such as spectral, textual, spatial, or topological properties [1]. The objective of OBIA in remote sensing is to provide adequate methods for analyzing very high-resolution (VHR) images within a spatial resolution of 1 m [2]. OBIA methods are often superior to pixel-based image analysis for classification and change detection (CD) in VHR remote
sensing images when a large amount of shadow and low spectral information or signal-to-noise ratio are observed [3–6]. Contrary to pixel-based analysis methods, which consider only spectral information, the OBIA methods classify pixels into homogenous image objects having useful features (e.g., shape, texture, and contexture relations with other objects); in addition, they can reduce the salt-and-pepper noise and spectral variation within a class in a VHR image [7,8]. For example, many materials may exhibit similar spectral reflectance in VHR images, such as bright desert soil vs. bright man-made features and cement roads vs. cement rooftops, even though they belong to completely different classes [4]. Therefore, distinguishing between such materials using only the spectral reflectance without considering the spatial information is difficult. However, the computational cost of the OBIA methods is higher than that of the pixel-based analysis methods, and the accuracy of the former is dependent on the segmentation results. For example, two types of errors, over- and under-segmentation, are observed in image segmentation; these errors cause the final objects to adversely express the real objects on the surface [7]. Recently, the development of high-performance computing systems and efficient software, such as eCognition [9], IMAGINE Objective [10], and ENVI’s feature extraction module [11], have enabled the implementation of OBIA [12].

CD is a major research area in remote sensing. Generally, object-based CD is performed by extracting the image objects exhibiting differences in spatial and/or spectral attributes of temporal images with time. The object-based CD methods can be classified as post-classification comparison and multitemporal image object analysis [10]. In the case of the post-classification comparison methods, the classification or segmentation results obtained from the temporal images are directly compared using geographic information system techniques, including polygon overlay and Boolean logic. The changed objects are the intersection of the multitemporal map objects involving different classes [13–16]. Thus, image objects having change information can be obtained; however, the CD accuracy is dependent on the quality of classification or segmentation results of both the temporal images. Multitemporal image object analysis involves the direct application of image segmentation and classification to the combined temporal images. In this case, the multitemporal composite or transformed images of the temporal images are considered as the input [17,18]. The transformed images can represent the same spectral, radiometric, or textural information with fewer bands to process as multitemporal composite datasets. However, multitemporal image object analysis is difficult to perform based on the “from-to” change information because the change information provides only the change magnitude and direction [13]. Thus, hybrid methods involving the combination of two or more pixel- or object-based CD methods have been developed. Pixel-based principal component analysis (PCA) and image differencing were applied to IKONOS images, and the changed areas were detected using object-based post-classification [19]. Thus, object-based post-classification can improve the results generated by pixel-based CD methods. Subsequently, multivariate alteration detection (MAD) transformation was applied to Quickbird images to detect changes using a pixel-based unsupervised approach, and the change objects were subclustered by performing fuzzy maximum likelihood estimation [20]. In addition, Han et al. [21] combined multiple pixel-based and object-based CD results using the weighted Dempster–Shafer fusion method. They detected changes in Worldview-3 images without ground truth data, and the pixel-based results were effectively extended to the object-based results.

Recently, deep learning has become an effective technique for application in remote sensing image analysis [22]. In general, representative deep learning architecture for image classification, such as convolutional neural network (CNN) and fully convolutional network (FCN), conducted pixel-based learning. Because the majority of the classification results of remote sensing images are expected to result in a dense class map as the output, the result map is a two-dimensional (2D) distribution of class labels with pixel correspondence in a “pixel-label” mode [23]. Recently, object-based methods involving pixel-based deep learning networks have been developed for performing remote sensing image analysis. The object information can be integrated during the preprocessing or postprocessing phases based on the image features extracted using a deep learning network [24]. Jin et al. [25] combined an object-based method with deep CNN for land use classification. Image segmentation regions with
high homogeneity were classified and extracted to obtain training data based on a typical rule set of feature objects. Then, the segmentation results were considered as the input of the CNN for improving the classification performance. Similarly, an object-based post-classification refinement method was proposed to obtain object-based thematic maps for land cover classification [26]. After obtaining a pixel-level classification map of the sentinel images using the CNN, the image object was labeled with the most frequent land cover class of its pixels. In addition, an object-based deep learning framework involving anisotropic diffusion data preprocessing and additional loss term was designed to reflect the object information on the FCN during the training process [24]. Thus, the pixel-based loss and object-based loss were combined, thereby obtaining the loss between the label and dominant label inside each object.

The aforementioned methods achieve promising results using a deep learning network based on the object information; however, these methods were developed for classification, and methods for CD have not yet been identified. In particular, when image features extracted by deep learning network were combined with objects, most of the studies reflect dominant or highest frequent label information of all the pixels inside an object without quantitative criteria. Although it is reasonable to consider the most frequent label of pixels in an object, it can cause errors if similar proportions are occupied by more than two labels. For example, let us assume that the pixels in an object are classified as two major land cover types: bare soil (49%) and water (51%). Then, the object is classified as water because it has the most frequent values. However, in this case, it is difficult to define the object as water because bare soil constitutes 49% of the pixels within the object.

Furthermore, training data are necessary to train the deep learning network because the loss is calculated between the predicted values and training data. Therefore, the areas in which changes occur should be identified. The samples randomly selected from the ground truth map were considered as the training data for estimating the CD accuracy with respect to the prediction map and ground truth map [27]. In this approach, the accuracy of the training data is 100% because training data are generated from the ground truth data. Although the aforementioned approach can be used to evaluate the performance of the proposed method and detect changes within the input sites, it is difficult to consider several real-world scenarios in which it is difficult to obtain prior change information about the site. For practical applications, it is necessary to ensure applicability to a broad area with minimal training data.

Herein, we propose a novel method that can be used to evaluate the uncertainty associated with object units in the case of object-based CD of the VHR satellite images. In particular, an initial change map was generated using several unsupervised CD methods to detect the changes in temporal images without considering any ground truth data. Therefore, the initial change map was integrated with objects and the selected objects included pixels having the same labels over a certain percentage to generate the training data of the CD network. After the learning of the CD network was completed, the pixels in the whole image were classified by the network as change or no-change. Subsequently, experiments were conducted to denote the effectiveness of the proposed method using two different multitemporal VHR satellite images, i.e., Worldview-3 and KOMPSAT-3. Then, the CD results were compared with the results obtained using various pixel-based CD methods, and the effects of the scale parameter and proportion value in objects were analyzed.

2. Methods

The proposed method involved two phases, and the objective was to detect changes occurring in the case of object-level temporal VHR satellite images in the absence of ground truth or prior knowledge based on the unsupervised CD method and deep learning network. The first step was to generate CD objects, which represent change or no-change classes over a specific percentage of objects, as label data for the CD network; the next step was to update the CD objects for obtaining the CD map of the entire region.
Figure 1 depicts the architecture of the proposed method. Two images $I_{t1}$ and $I_{t2}$ were acquired from the same region at time $t1$ and $t2$, respectively. First, various unsupervised CD methods were applied to $I_{t1}$ and $I_{t2}$ for obtaining the initial CD map. Then, objects were generated from the two images as the segmentation results of the principle components of these images. The initial CD map comprised three classes: “change”, “no-change”, and “no-value”. The map was reconstructed with respect to the units of objects (CD objects) according to the percentage of each class. The detailed method for obtaining the initial CD map and CD objects is explained in Section 2.1.

Figure 1. Framework of the proposed method. The objects were obtained from principal components (PC) of two temporal images, and the CD objects are used as label data for the deep learning network. The result map of the deep learning network is integrated with the object boundary to update the CD objects, which involve change or no-change classes over a specific percentage for the entire region. This process is repeatedly performed until all the objects of the entire image are classified into change or no-change classes.

The CD objects were fed into the CD network, and the objects were classified as change, no-change, and no-value. Only the change and no-change classes were used as label data. The pixels in the no-value class were not used for training the network because they were masked during training. After the training phase, the network generated a binary map containing change and no-change classes for the entire image. The CD result map was reconstructed using objects. The reliable objects, in which the percentage of change and no-change classes was greater than a specific value, were selected and added to the initial CD objects. The detailed method of updating CD objects is explained in Section 2.2. The initial CD objects included no-values. When learning with the initial training data, it was impossible to train the locations of pixels with no-values. The updated CD objects were used for the training data of the CD network to train a CD network for the entire regions of images. This process was iteratively performed until the percentage of change and no-change classes in all objects exceeded a specific value.

2.1. Generating CD Objects

Training data were necessary to learn the change and no-change classes in temporal images because the CD network in this work conducted supervised learning. The quality of the training data considerably influenced the accuracy of the CD result. Figure 2 presents the generation of CD objects as the label data. Although the unsupervised CD methods are easily affected by spot noise, they can be applied to images and without prior knowledge for performing quantitative analysis. In this study, five pixel-based unsupervised CD methods were used, which can extract the changed pixels by measuring the spectral difference between two images to obtain an initial CD map. Various unsupervised CD methods can extract different pixels as changed/unchanged areas because the calculation method and criteria for judging the changes are different. Therefore, we selected the pixels identified as belonging to the same class in more than four methods as the final initial CD map to extract the pixels determined in the majority of the algorithms. In addition, the thresholds percentage related to the level of uncertainty were determined to reconstruct the initial CD map in units of objects. If the pixels within an object had the same class for more than the determined threshold percentage (in this study, 50%, 60%, and 70% were used as the threshold percentages), the object was classified as belonging to the class of most pixels. Thus, all the pixels within
the object have the value of class occupied by the majority of the pixels when the pixels in an object are classified as the same item based on a threshold percentage. Furthermore, the objects were defined as no-value when the pixels within an object did not have the same class over a certain percentage. Finally, the CD objects comprised three classes, i.e., change, no-change, and no-value, having the same value in the case of object units.

Figure 2. Generation of CD objects. The pixels were classified as belonging to the same class in more than four methods among the five unsupervised CD methods and were selected to constitute the initial CD map, which was reconstructed in object units with threshold percentages, i.e., specific percentages of pixels in an object; in this study, 50%, 60%, and 70% were used to generate CD objects.

2.1.1. Initial CD Map Generated from Unsupervised CD Methods

The unsupervised pixel-based CD methods employ a pixel as the basic unit of analysis and can extract changes based on the spectral information [4]. The traditional pixel-based CD methods achieve remarkable performance for low- and moderate-resolution satellite images. However, they are often unsuitable for application to VHR images. They often result in salt-and-pepper noise because of the detection of several changes owing to spot noise [28]. Despite these limitations, they are widely used in case of VHR images because they can be easily applied without the requirement of prior information on the study site [29,30]. Generally, difference images (DIs), which highlight the spectral difference between two images acquired based on the same region at different time points, were generated. In addition, decision functions, such as the application of a threshold value and clustering algorithms, were used to classify the change from no-change in most of the unsupervised CD methods. Clustering algorithms, such as k-means clustering, have been widely used because selecting appropriate threshold values is difficult, especially when ground truth data are unavailable. In the case of binary CD, the clustering algorithms divide all the pixels of the DI into two classes.

Herein, five traditional methods, namely image differencing, image regression, change vector analysis (CVA), iterative reweighted multivariate alteration detection (IR-MAD), and PCA, were used to generate the DI, and k-means clustering was exploited to differentiate the change class from the no-change class.

Image differencing is an easy method for interpreting CD results. The temporal images are directly subtracted from \( I_{11} \) to \( I_{12} \) in a pixel-by-pixel manner [31].

\[
\text{ImageDiff} = |(I_{11}(x, y) - I_{12}(x, y))| \tag{1}
\]

where ImageDiff is the difference image generated by image differencing and \((x, y)\) are the coordinates. The difference image has the same number of bands as that in the input images. For example, the difference image will have four bands if the input images have four spectral bands. Based on image differencing, the absolute values of the difference between the corresponding pixels in temporal images can be calculated; large values in the difference image represent changed pixels. Because the output value is an absolute value, the same value may exhibit different meanings; thus, it requires a preprocessing step of atmospheric calibration.

Based on image regression, the relation between \( I_{11} \) and \( I_{12} \) can be established and the pixel values of \( I_{12} \) can be estimated using a regression function such as the least squares regression [32]. In image regression, the pixels from \( I_{11} \) are assumed to be a linear function of time with respect to the pixels
from \( I_2 \) [31]. In other words, \( I_{11} \) is the reference image and \( I_{12} \) is the subject image. \( I_{12} \) is adjusted to match the radiometric conditions of \( I_{11} \). If \( \hat{I}_{12} \) is the predicted value obtained from the regression line, the difference image can be defined as

\[
\text{ImageRegr} = |(I_{11}(x, y) - \hat{I}_{12}(x, y))|
\]

where ImageRegr represents the difference image generated by image regression. This can reduce the effect of atmospheric and environmental conditions but requires an accurate regression function [32].

The magnitude of change between \( I_{11} \) and \( I_{12} \) is calculated based on CVA. The pixel values considered as a vector of the spectral bands and the change vector are calculated by subtracting the vectors for all pixels in the case of different data [33]. The magnitude represents the degree of change; thus, it can be used to distinguish between the change and no-change classes [34]. To apply CVA, preprocessing steps such as data transformation are required. For example, principal component (PC), tasseled cap, and spectral index can be utilized to generate spectral features from each image pair. CVA can deal with any number of spectral bands and produce detailed CD information; however, it is difficult to identify land cover change trajectories [32]. The change magnitude of CVA was calculated as follows:

\[
CVA = \sqrt{\sum_{k=1}^{N} (I_{11,k}(x, y) - I_{12,k}(x, y))^2}
\]

where \( I_{11,k} \) and \( I_{12,k} \) are the \( k \)th band of \( I_{11} \) and \( I_{12} \), respectively, and \( N \) is the number of spectral bands of both images.

Multivariate data are transformed via PCA to obtain a new set of components for reducing data redundancy. PCA produces new components based on eigenvector analysis of the covariance matrix, and most of the variance with respect to all the original variables can be observed in the case of the first few components. For achieving the CD of the remote sensing data, \( I_{11} \) and \( I_{12} \) with \( \alpha \) and \( \beta \) bands are combined into one image with \((\alpha + \beta)\) bands; then, the stacked image is transformed into \((\alpha + \beta)\) PCs. Among the temporal images, high and low correlations can be observed with respect to the unchanged and changed areas, respectively. In addition, generally, the first four components exhibit change information [35]. PCA can effectively reduce the redundancy data and denote different change information; however, obtaining a suitable interpretation of different datasets is difficult because it is scene dependent.

IR-MAD is a regularized, iteratively reweighted MAD method based on canonical correlation analysis. In this method, the coupling vector exhibiting the highest correlation to a set of multivariate variables is estimated [36]. MAD finds the differences between linear combinations of spectral bands from \( I_{11} \) and \( I_{12} \). Thus, a set of \( N \) change maps is obtained, where \( N \) represents the maximum number of bands and each change map is orthogonal to the remaining change maps. Further, the uncorrelated difference images can be sequentially extracted, where each new image shows maximum change under the constraint of being uncorrelated with the previous images [37]. The intensity image of IR-MAD was calculated as follows:

\[
\text{IRMAD} = \sum_{k=1}^{N} \left( \frac{U_k - V_k}{\sigma_k} \right)^2
\]

where \( \sigma_k \) is the standard deviation of the \( k \)th band. \( U_k \) and \( V_k \) can be derived as follows:

\[
U_k = a^T I_{11,k} \quad V_k = b^T I_{12,k}
\]

where \( a \) and \( b \) are the transformation vectors obtained via canonical correlation analysis. In this method, high weights are assigned to the unchanged pixels during iterations to reduce the negative effect of changed pixels during convergence [21].
All pixels in DI were clustered using the k-means algorithm to produce a binary CD map. Two classes with pixels belonging to the change and no-change classes were used. In this case, the pixel value "0" denoted the no-change class, whereas the pixel value “1” denoted the change class. K-means clustering can be used to iteratively partition the pixels into two groups, where each pixel belongs to only one class. Further, the pixels are assigned to a cluster based on the sum of the squared distance between the cluster’s centroid and pixels. The initial centroids of k classes were randomly selected, and this process was repeated until there was no change in the centroids.

The traditional methods have unique advantages and disadvantages because each algorithm is based on different principles. The performance of the methods may vary depending on the input images because the quality of the difference images is dependent on the input image characteristics and environmental conditions during image acquisition. Hence, it is difficult to decide the best method for all the cases. Therefore, we integrated the CD results obtained using various methods and defined the changed pixel, which is classified as changed pixels by most of the algorithms. CM<sub>sum</sub>(x, y) is the sum of classes classified as change at (x, y) position of the change maps extracted by k-means clustering of various methods. Because there are five methods, the range was 0 ≤ CM<sub>sum</sub> ≤ 5. The initial change map can be defined using the decision function (Equation (7)). The pixels, which were not included in the change and no-change classes, exhibit no-value in the initial CD map.

\[
\text{Initial CD map } (x, y) = \begin{cases} 
\text{no-change} & (\text{if } CM_{\text{sum}}(x, y) \leq 1) \\
\text{no-value} & (\text{if } 2 \leq CM_{\text{sum}}(x, y) \leq 3) \\
\text{change} & (\text{if } CM_{\text{sum}}(x, y) \geq 4) 
\end{cases}
\]

(7)

The pixels at (x, y) with CM<sub>sum</sub> ≥ 4 were classified as the change class, whereas those with CM<sub>sum</sub> ≤ 1 were classified as the no-change class. In addition, the pixels with 2 ≤ CM<sub>sum</sub> ≤ 3 were classified as the no-value class, indicating that two or more algorithms classified the pixel as belonging to the change class. Thus, the changes at these pixels are difficult to distinguish.

2.1.2. Segmentation of Temporal Images

Image segmentation is the process of partitioning an image into multiple segments. In other words, image segmentation is to assign labels to all pixels in an image such that the pixel has the same label to share certain characteristics. The image objects are treated as several superpixels, which can be defined as a group of pixels exhibiting common characteristics. A graph-based segmentation algorithm can be used to effectively generate superpixels [38]. Felzenszwalb and Huttenlocher [39] developed an efficient segmentation algorithm based on graph theory. This algorithm calculates the gradient between two adjacent pixels, which is weighted according to the pixel properties [40]. Further, it minimizes the difference between the gradients within the segment but maximizes the differences between adjacent segments. There are three parameters associated with this algorithm: k, m, and σ. k sets the observation scale, and larger segments are obtained with increasing k. m is the minimum object size of the components. A small component can be observed when there is a major difference between the neighboring components. σ is the diameter of a Gaussian kernel for slightly smoothing the image prior to segmentation.

In this study, objects were obtained from the PC images generated after PCA using the open-source image-processing toolkit, scikit-image, in Python. The newly obtained PC images were used as the input of the segmentation algorithm to effectively reflect the changed information between l<sub>11</sub> and l<sub>12</sub> simultaneously. σ was always set to 0.8, which is the default value and does not visually change the image but helps to eliminate image artifacts [39]. The optimal scale size of the input images can differ with spatial resolution and material type. To analyze the effect of k value, the experiments were conducted at k values of 30, 50, 100, 200, and 300. m is also related to the size of materials in the input images. The m values are determined in an empirical manner. For example, m was 60 for the study site,
where changes could be observed with respect to small objects, such as buildings, and \( m \) was set to 100 for changes in land cover without any building material.

2.1.3. Reflection of the Uncertainty in an Object Unit

The CD object can be generated from the initial CD map and segmentation results. The initial CD map has three pixel-level classes: change, no-change, and no-value. Several thresholds, which are the percentages of the pixel’s label in an object, were determined to reconstruct the initial CD map with the object unit. If the pixels included in an object have three classes in similar proportions, the uncertainty is associated with defining an object as one class increases. Therefore, only objects containing at least more than 50% pixels with the same label were selected to generate CD objects. In addition, when a specific class of the pixels within an object does not occupy more than the threshold percentage, the object is defined to have a no-value. If the threshold percentage is large, a reliable object is selected as the CD object because most of the pixels in an object have the same class; however, the number of objects satisfying this condition can be reduced. To obtain the optimal threshold, the CD objects were generated using different percentage from 50% to 70% at an interval of 10%, and each condition was categorized into uncertainty Levels 3, 2, and 1, respectively. The higher is the level of uncertainty, the lower is the percentage of same classes represented by the pixels in the object. However, the number of samples and quality of data are equally important because the CD objects are used as training labels for the CD network. Therefore, we do not consider a percentage of more than 80% because insufficient CD objects are available when the threshold is set to 80% or more.

2.2. Updating CD Object

CD objects were used as label data for training the CD network. Figure 3 shows the architecture of the CD network, which comprises 3D and 2D convolutional layers and convolutional long short-term memory (LSTM) layers. 3D convolutional layers can extract spatial and spectral feature maps from hyper or multispectral images, and convolutional LSTM can analyze the temporal relation between two images [27,41]. Convolutional LSTM is a modification of the conventional LSTM in which the matrix multiplication operators are replaced with convolution operators [42]. Convolutional LSTM is suitable for application to remote sensing images when compared with conventional LSTM because the size of the weight matrix increases the computational cost; further, spatial connectivity is ignored when applying the conventional LSTM [41].
Two temporal images and CD object were used as the input data. After the training step, the network produces a binary CD map. $w$, $h$, and $\lambda$ represent the width, height, and number of spectral bands, respectively. $\Omega_c$ and $\Omega_u$ are the change and no-change classes, respectively. $\omega_c$, $\omega_u$, and $\omega_n$ are the change, no-change, and no-value classes of the initial CD map, respectively.

The training samples were randomly extracted from the temporal VHR images and the CD objects. Each training sample was a 3D patch with dimensions of $w \times h \times \lambda$, where $w$ and $h$ are the lengths of the column and row, respectively, and $\lambda$ is the number of spectral bands. The size of the 3D patches was empirically set to $10 \times 10 \times \lambda$ in this study. The central points of the 3D patches were randomly extracted only at the locations of the change and no-change classes of the CD objects. We selected 40,000 pixels as training data, 20,000 pixels as validation data, and 30,000 pixels as testing data. Because convolutional layers exploited information from neighboring pixels and the training and validation patches were extracted from the CD objects, their features were likely to overlap owing to the shared source of information. Overlap between training and validation data can result in intrinsic positive bias in the CD result. However, in this study, because the images of the study areas consist of relatively few pixels (e.g., 1200 $\times$ 1200 pixels), the number of training patches was reduced when extracting without overlap. Therefore, data for network training were randomly extracted to increase the amount of training data.

After the training samples were extracted, two patches captured from the same location of two temporal images were separately fed into the 3D convolutional layers in parallel. The filter size of these convolutional layers was $(3 \times 3 \times 3)$, which is the optimal alternative for 3D convolution in spatiotemporal feature learning. Next, the spatial–spectral feature maps were fed into the convolutional LSTM layers to denote the temporal information and recode the change rules. The outputs from the convolutional LSTM layers were passed through 2D convolutional layers with $(3 \times 3)$ filters to generate a score map. The final number of feature maps was equal to the number of classes. The binary cross entropy $L$ is used as loss function of the network and it can be defined as follows:

$$L = -\frac{1}{N} \sum_{i=0}^{n} (y_i \log(\hat{y}_i)) + (1 - y_i) \log(1 - \hat{y}_i))$$ (8)
where \( n \) is the number of samples, \( y_i \) is ground truth value, and \( \hat{y}_i \) is the predicted value. Finally, the pixels were classified into change or no-change classes according to the score map.

The binary CD result map obtained using the CD network was integrated with the segmentation results. The CD objects representing meaningful classes, such as change or no-change, were used to train the CD network samples, and the objects retained their existing properties. On the contrary, the objects having no-values must be updated. The process of updating CD objects is described in Figure 4. The binary CD result map was integrated with the object units and the candidate area to be updated, and the CD objects with no-values were selected. The CD result map was reconstructed according to the class percentage of the pixels within an object to consider the uncertainty of an object when updating CD objects. The process of reconstructing the object class follows the same rule and threshold as those associated with the previously generated CD object. For example, if the previous CD objects were generated with a threshold of 60%, only the objects classified as change or no-change pixels with more than 60% are selected. The objects that do not follow the aforementioned conditions are assigned no-values. The CD objects were updated by combining the previous CD objects and the newly generated objects. The updated CD objects were also fed into the CD network, and the network was trained using the randomly extracted samples of CD objects. This process was repeatedly performed until all the objects in the area were classified as change or no-change classes.

![Figure 4.](image)

**Figure 4.** The process of updating CD objects. CD objects are fed into the CD network, and the network produces a binary CD map, which can be divided into “noncandidate objects for updating” and “candidate objects for updating”. After uncertainty analysis, the selected objects were added to the previous CD object. \( \Omega_f \) and \( \Omega_n \) are the change and no-change classes, respectively. \( \omega_f \), \( \omega_c \), and \( \omega_u \) are the change, no-change, and no-value classes of the initial CD map, respectively.

### 2.3. Performance Evaluation

The classification performance can be assessed using the confusion matrix, also known as the error matrix. Binary CD can be interpreted as a classification task involving two classes. The confusion matrix is a 2 \( \times \) 2 table that contains four outcomes produced by the binary classifier. To evaluate the performance of the CD methods, various performance measures showing how well the classifier accurately identifies the objects, such as the overall accuracy (OA), precision, recall, and F1 score, can be calculated from the confusion matrix. Test data were used to calculate the accuracy of the CD methods. OA represents the proportion of accurately classified prediction with respect to the observation, and it can be described in terms of true positive (TP), true negative (TN), false negative (FN), and false positive (FP).

\[
\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}
\]

OA is a simple methodology for evaluating the classification accuracy and works well when FP and FN exhibit similar costs. However, when the class distribution is dissimilar, FP and FN are considerably
different, and OA is inappropriate for showing the effectiveness of the result. In this case, the F1 score is a better way to evaluate the results. F1 score is the harmonic mean of precision and recall (Equation (10)). Precision can be obtained by dividing the total number of accurately classified positive pixels with the total number of predicted positive pixels, whereas recall is the ratio of the total number of accurately classified positive pixels divided by the total number of positive pixels (Equations (11) and (12)). In addition, negative predicted value (NPV) represents how many of negative pixels are true negatives (Equation (13)).

$$\text{F1 score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}}$$  \hspace{1cm} (10)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$  \hspace{1cm} (11)

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$  \hspace{1cm} (12)

$$\text{NPV} = \frac{\text{TN}}{\text{FN} + \text{TN}}$$  \hspace{1cm} (13)

3. Dataset

Multispectral VHR images of two sites were used for CD (Figure 5) [21]. The temporal images of Site 1 were acquired from WorldView-3 with a spatial resolution of 1.24 m and 8 bands. The images were acquired from Gwangju city in South Korea. This city includes industrial areas, residences, agricultural lands, rivers, and changed regions because of large-scale urban development. The temporal multispectral images ($I_{t1}$ and $I_{t2}$) were acquired on 26 May 2017, and 4 May 2018, respectively. The Site 2 images were acquired from KOMPSAT-3 multispectral sensor images having a spatial resolution of 2.8 m and 4 bands. $I_{t1}$ and $I_{t2}$ were acquired on 16 November 2013, and 26 February 2019, respectively, from the area located over Sejong city in South Korea. This area is an administrative city in South Korea and has been developing since 2007; a central administrative agency has also relocated to this area. Large-scale high-rise buildings and complexes have been constructed in a short period, considerably changing the image pair.

To perform effective and reliable CD, accurate geometrical preprocessing such as orthorectification should be performed with respect to the multitemporal VHR images to minimize geometric misalignment [45,46]. Image $I_{t2}$ was co-registered to the coordinates related to image $I_{t1}$ by applying the phase-based correlation method [47] with an improved piecewise linear transformation warping [48]. The co-registration was applied to warp image $I_{t2}$ related to the coordinates of image $I_{t1}$. We did not perform pansharpening because the spatial resolution of the image was sufficient to describe the scene in detail.

The ground truth data were manually obtained based on various VHR web maps, the value of the normalized difference vegetation index, and field survey. We defined the changes when the classes of land cover had changed such as vegetation to bare soil. The classes of land cover were defined as vegetation, bare soil, buildings, water, and roads. Vegetation was defined as crop land and trees with high vegetation vitality. Buildings having height were defined as “buildings”. “Bare soil” represented ground without buildings and vegetation (or areas with very low vegetation vitality), and “roads” encompassed asphalt roadways. Changes owing to relief displacement and shadows were not considered as changes in the ground truth data. In particular, the greenhouse of Site 1 appears in different colors depending on the influence of light and internal materials in temporal images. Such differences do not denote the changed area; the area in which a greenhouse was newly constructed instead of bare soil was selected as the changed area in the ground truth map. Moreover, the slight differences in vegetation vitality owing to seasonal differences were not considered as changes.
After the CD objects were generated, the CD networks were trained using the CD objects. In this study, (Equation (7)). The pixels equally determined to denote changed or unchanged area in more than four various unsupervised CD methods, and Table 1 shows the accuracies of the results. Image regression denotes the lowest accuracies at both the sites. PCA and CVA have the highest accuracies methods were extracted as the initial CD maps. Figures 6 and 7 show the results of CD generated using the proposed method. Further, we compared the CD results for various threshold percentages, including 50%, 60%, and 70%. The Site 2 images were acquired from KOMPSAT-3 over Sejong city, South Korea on: (d) 16 November 2013; and (e) 26 February 2019. (c,f) The ground truth maps for Sites 1 and 2, respectively. \( \Omega_c \) and \( \Omega_d \) are the change and no-change classes, respectively.

4. Results

The experiments were conducted on Sites 1 and 2 having different VHR satellite images. Initially, the CD objects were generated using various unsupervised CD methods and segmented objects. After the CD objects were generated, the CD networks were trained using the CD objects. In this study, the final epoch was set to 200 in the case of the Adam optimizer with a learning rate of \( 10^{-4} \) and a batch size of 256. The CD objects were iteratively updated after every 50 epochs. The pixel-based CD results generated from the CD network were compared to show the effectiveness of the proposed method.

4.1. Generation of CD Objects

The CD objects were generated by integrating the initial CD map with the segmented objects. The initial CD maps were obtained from unsupervised CD results using the decision function (Equation (7)). The pixels equally determined to denote changed or unchanged area in more than four methods were extracted as the initial CD maps. Figures 6 and 7 show the results of CD generated using various unsupervised CD methods, and Table 1 shows the accuracies of the results. Image regression denotes the lowest accuracies at both the sites. PCA and CVA have the highest accuracies at Sites 1 and 2, respectively. The materials and properties of the changes at the study sites can affect the CD results. PCA can detect the changes from vegetation to bare soil, and image differencing can detect newly constructed buildings. Because the unsupervised CD methods are pixel-based methods, they only consider the spectral difference between two images. Therefore, the salt-and-pepper noise and the difference caused by shadows and lights were detected as the changed class. For example, in the case of Site 1, the greenhouse where the color looks different by light and inner materials were also
extracted as the changed area. Furthermore, in the case of Site 2, most CD methods classified the shadows apart from buildings as the changed class.

Figure 6. CD result maps of Site 1 generated using various pixel-based unsupervised CD methods: (a) ImageDiff; (b) ImageRegr; (c) CVA; (d) IR-MAD; (e) PCA; and (f) initial CD map. $\Omega_c$ and $\Omega_u$ are the change and no-change classes, respectively. $\omega_c$, $\omega_u$, and $\omega_n$ are the change, no-change, and no-value classes of the initial CD map, respectively.

Figure 7. CD result maps of Site 2 generated using various pixel-based unsupervised CD methods: (a) ImageDiff; (b) ImageRegr; (c) CVA; (d) IR-MAD; (e) PCA; and (f) initial CD map. $\Omega_c$ and $\Omega_u$ are the change and no-change classes in CD result map, respectively. $\omega_c$, $\omega_u$, and $\omega_n$ are the change, no-change, and no-value classes of the initial CD map, respectively.
Parameters were $k$, $m$, and $\sigma$. Figure 9 shows CD objects with different uncertainty levels in objects. Tables 2 and 3 represent the accuracy of the CD maps generated using various pixel-based unsupervised CD methods.

| Study Site | Methods  | OA    | Precision | Recall   | F1-Score  |
|------------|----------|-------|-----------|----------|-----------|
| Site 1     | ImageDiff| 0.8316| 0.6147    | 0.7760   | 0.6860    |
|            | ImageRegr| 0.7778| 0.5107    | 0.6685   | 0.5790    |
|            | CVA      | 0.8298| 0.6189    | 0.7676   | 0.6853    |
|            | IR-MAD   | 0.8200| 0.5946    | 0.7520   | 0.6641    |
|            | PCA      | 0.8471| 0.8177    | 0.7135   | 0.7620    |
| Site 2     | ImageDiff| 0.8649| 0.7797    | 0.6676   | 0.7193    |
|            | ImageRegr| 0.7705| 0.5417    | 0.4848   | 0.5117    |
|            | CVA      | 0.8716| 0.7762    | 0.6863   | 0.7285    |
|            | IR-MAD   | 0.8033| 0.7496    | 0.5410   | 0.6285    |
|            | PCA      | 0.8108| 0.6693    | 0.5622   | 0.6110    |

Figure 8 shows input images with the boundary of segmented objects. The segmentation parameters were $k = 200$, $m = 100$, and $\sigma = 0.8$ for Site 1 and $k = 50$, $m = 60$, and $\sigma = 0.8$ for Site 2. $k$ values of 30–300 were applied to the input image, and Figure 8 shows the optimal parameter value. Because Site 2 involves changes in buildings, a low value was more effective in distinguishing between building objects. The segmented results of $I_{t1}$ (Figure 8a,e) cannot reflect the object information at the time of change. For example, newly built buildings and bare soil areas were not reflected in the early obtained images. On the contrary, the segmented results of $I_{t2}$ (Figure 8b,f) contain current information, such as newly constructed materials but do not reflect the previous state of the area. Figure 8c,g shows the colored composite of PCs with the segmented results for Sites 1 and 2, respectively. The advantage of using the PCs extracted from the stacked images is that they can consider the temporal image information. Therefore, we subjected the combination of PCs to segmentation and used the segmented results to generate CD objects.

![Figure 8](image-url)  
*Figure 8. Colored infrared images with the boundary of segmentation for $I_{t1}$ at (a) Site 1 and (d) Site 2 and $I_{t2}$ at (b) Site 1 and (e) Site 2. Colored composites of PCs with the segmentation results for (c) Site 1 and (f) Site 2.*

We reconstructed the initial CD map using the segmented results to effectively generate CD objects. Figure 9 shows CD objects with different uncertainty levels in objects. Tables 2 and 3 represent...
the number of pixels in three classes with the uncertainty levels and the accuracy of the CD objects. The uncertainty Levels 1, 2, and 3 indicate the threshold percentages of 70%, 60%, and 50%, respectively. With increasing uncertainty levels, the number of CD objects representing the $\omega_c$ or $\omega_u$ classes increase. On the contrary, objects classified as $\omega_n$ increase with decreasing uncertainty levels. In addition, the accuracies of $\omega_c$ and $\omega_u$ were improved with decreasing uncertainty levels.

Figure 9. CD objects overlapped with the boundaries of segments at different uncertainty levels: Level 3 at (a) Site 1 and (d) Site 2; Level 2 at (b) Site 1 and (e) Site 2; and Level 1 at (c) Site 1 and (f) Site 2. $\omega_c$, $\omega_u$, and $\omega_n$ are the change, no-change, and no-value classes of the initial CD object, respectively.

Table 2. The number of pixels with different classes and uncertainty levels.

| Study Site | Uncertainty Level | The Number of Pixels in |       |       |       |
|-----------|-------------------|-------------------------|-------|-------|-------|
|           |                   | $\omega_c$ | $\omega_u$ | $\omega_n$ |       |
| Site 1    | Level 1           | 185,626    | 938,376    | 315,998    |       |
|           | Level 2           | 215,381    | 997,600    | 227,019    |       |
|           | Level 3           | 263,464    | 1,043,117  | 133,419    |       |
| Site 2    | Level 1           | 19,215     | 96,322     | 44,463     |       |
|           | Level 2           | 22,066     | 108,230    | 29,704     |       |
|           | Level 3           | 24,280     | 117,225    | 18,495     |       |

Table 3. Accuracy of the CD objects with different uncertainty levels.

| Study Site | Uncertainty Level | Precision | NPV   |
|-----------|-------------------|-----------|-------|
| Site 1    | Level 1           | 0.9420    | 0.9432|
|           | Level 2           | 0.9224    | 0.9331|
|           | Level 3           | 0.9183    | 0.9077|
| Site 2    | Level 1           | 0.9420    | 0.9432|
|           | Level 2           | 0.9153    | 0.9399|
|           | Level 3           | 0.8912    | 0.9339|
4.2. CD Result of Traditional Approaches Using the CD Network

Generally, object information can be obtained from the preprocessing and postprocessing phases in which the input images are modified to obtain object information or refine the output with object units. We compared the CD results of the traditional approaches to confirm the effectiveness of the proposed CD method.

We defined four different cases using the CD network; the detailed experimental conditions are described as follows:

- **Case 1**: The original multitemporal images were used as the input data, and the initial pixel-level CD map was used as the label data in the case of the CD network. After training, the network produced a pixel-level CD map. In this case, object information could not be obtained.
- **Case 2**: The original multitemporal images were used as the input data, and the CD objects generated in Section 4.1 were used as the label data to train the CD network. In other words, the object information was reflected in the preprocessing phase, and the network produced a pixel-level CD map.
- **Case 3**: The segmentation image, in which each image has a unique value, was added to the original images. Thus, a band containing object information was stacked onto the existing bands. The new images adding one more band were used as the input data, and the initial pixel-level CD map was used as the label data for the CD network.
- **Case 4**: The result of Case 1 was subjected to postprocessing. In this case, the object was reclassified as the most dominant class of pixels within the object.

The experimental setting of the CD network and input materials was similar to those in the proposed method. Figure 10 shows the CD results, and Table 4 describes the accuracies. By comparing Cases 1 and 2, the CD accuracies at both the sites can be improved by adding object information via initial CD map reconstruction. In addition, when stacking original images onto the segmentation band as one image (Case 3), the accuracies decreased because the salt-and-pepper noise increased. However, the pixel changes can be attributed to shadows and light difference. The CD accuracies were the highest in Case 4. The noise within the image was eliminated. However, at Site 2, most of the building objects were also removed. This is because the shape of the buildings was underestimated in Case 1; therefore, in the postprocessing steps, the objects were classified as unchanged objects. Therefore, the OAs of Sites 1 and 2 are similar, but the F1 score of Site 2 is lower than that of Site 1.

Table 4. Accuracy of CD maps generated via traditional approaches using a CD network.

| Study Site | Methods | OA   | Precision | Recall | F1-Score |
|------------|---------|------|-----------|--------|---------|
| Site 1     | Case 1  | 0.8623 | 0.6535    | 0.8726 | 0.7473  |
|            | Case 2  | 0.8727 | 0.6335    | 0.9378 | 0.7562  |
|            | Case 3  | 0.8436 | 0.5618    | 0.8982 | 0.6913  |
|            | Case 4  | 0.8874 | 0.6878    | 0.9336 | 0.7921  |
| Site 2     | Case 1  | 0.8787 | 0.5127    | 0.8883 | 0.6501  |
|            | Case 2  | 0.8821 | 0.4933    | 0.9438 | 0.6479  |
|            | Case 3  | 0.8772 | 0.5114    | 0.8801 | 0.6469  |
|            | Case 4  | 0.8831 | 0.5068    | 0.9296 | 0.6560  |
Figure 10. CD results of traditional approaches using a CD network for Sites 1 and 2, respectively: (a,e) Case 1 (original images + initial CD map); (b,f) Case 2 (original images + CD objects); (c,g) Case 3 (adding object band images + initial CD map); and (d,h) Case 4 (postprocessing after Case 1). Ω_c and Ω_n are the change and no-change classes, respectively.

4.3. CD Result of the Proposed Method

In the proposed method, CD objects were used to train the CD network; further, the binary CD map generated from the network was used to update CD objects. In this step, the results of CD objects were obtained depending on the uncertainty level. To obtain the optimal uncertainty level, we compared the CD results obtained under different conditions. The CD network was trained until epoch 200, and the CD objects were updated after every 50 epochs. e_0, e_1, e_2, e_3, and e_4 are the updating points and represent epochs 0, 50, 100, 150, and 200, respectively. When all the CD objects were assigned changed or unchanged classes before e_4, the learning process was completed. Moreover, we set the learning process as finished even when the CD objects still contained a no-value class after conducting the learning process until e_4.

To analyze the effects of the uncertainty level in various cases, we present three conditions: (1) maintain the same level during every update phase; (2) increase the level as the update progresses; and (3) decrease the level during the update phase. Figures 11–16 show the CD results obtained at Sites 1 and 2, and Table 5 shows the accuracies of all the cases. When maintaining the same level during the updating process (Figures 11 and 12), the CD results with an uncertainty Level 3, in which the objects are updated when the pixels have the same class more than 50% in the object, were fully generated within e_1. On the contrary, the CD result could not be used to obtain values for the whole area when the proposed method was used in case of uncertainty Levels 2 and 1. This is because, as the level increases, all the pixels in the object must have the same class to update the CD objects. Thus, it is difficult to update the objects. In addition, because the CD network was trained with the updated CD objects, the CD result maps generated from the CD network would not have been considerably different if there were slight differences between the updated and previous objects. However, the accuracies of the CD results with Levels 2 and 1 were higher than that of the CD result with Level 3 (Table 5) for both sites. For example, the CD results with Level 2 resulted in an OA of 0.9174 and an F1 score of 0.8542 for Site 1 and an OA of 0.9006 and an F1 score of 0.7353 for Site 2.
Remote Sens. 2020, 12, x FOR PEER REVIEW 17 of 26

Figure 11. The CD results of the proposed method for each epoch (e₀ and epoch 0 and e₄ is epoch 200) with different uncertainties (the uncertainty level maintains one value in the training phase) at Site 1: (a) Level 3; (b) Level 2; and (c) Level 1. Grey (Ωₓ), white (Ωₒ), and black (Ωᵦ) colors represent the change, no-change, and no-value classes, respectively.

Figure 12. The CD results of the proposed methods for each epoch with different uncertainties (the uncertainty level maintains one value in the training phase) at Site 2: (a) Level 3; (b) Level 2; and (c) Level 1. Grey (Ωₓ), white (Ωₒ), and black (Ωᵦ) colors represent the change, no-change, and no-value classes, respectively.
Although the CD results obtained with uncertainty Level 3 can produce CD objects via one updating phase, the accuracies of the CD results were lower than those of the remaining cases. On the contrary, CD results with Levels 1 and 2 were not valid for the CD objects in the entire area; however, the generated CD objects could achieve increased CD accuracies. Therefore, the uncertainty level was changed during the update process. Figures 13 and 14 show the CD results obtained using an increased uncertainty level during the update. Thus, the proportion of pixels having the same class that can changed during the update process. Figures 13 and 14 show the CD results obtained using an increased uncertainty level during the update. Thus, the proportion of pixels having the same class that can changed during the update process. Figures 13 and 14 show the CD results obtained using an increased uncertainty level during the update. Thus, the proportion of pixels having the same class to define CD objects can be gradually increased. In particular, several CD objects could not be decreased when the uncertainty changed from Level 2 to Level 3. Grey ($\Omega_c$), white ($\Omega_u$), and black ($\Omega_v$) colors represent the change, no-change, and no-value classes, respectively.

Although the CD results obtained with uncertainty Level 3 can produce CD objects via one updating phase, the accuracies of the CD results were lower than those of the remaining cases. On the contrary, CD results with Levels 1 and 2 were not valid for the CD objects in the entire area; however, the generated CD objects could achieve increased CD accuracies. Therefore, the uncertainty level was changed during the update process. Figures 13 and 14 show the CD results obtained using an increased uncertainty level during the update. Thus, the proportion of pixels having the same class that can define CD objects is gradually reduced. In this case, the uncertainty of the objects increases; however, all the objects in the input image can be classified as change or no-change classes. The CD results with an increased uncertainty from Level 1 to Level 3 show OA = 0.8611 and F1 score = 0.7218 at Site 1 and OA = 0.8859 and F1 score = 0.6770 at Site 2. Further, when the uncertainty changed from Level 2 to Level 3, the accuracies increased to become the highest, e.g., OA = 0.9299 and F1 score = 0.8745 at Site 1 and OA = 0.9012 and F1 score = 0.7347 at Site 2 (Table 5).
were similar, indicating that only some CD objects were newly added during the updating process.

The CD results with the uncertainty decreasing from Level 3 to Level 1 resulted in OA = 0.8957 and F1 score = 0.9006 at Site 1 and OA = 0.9174 and F1 score = 0.9373 at Site 2 (Table 5). In addition, the CD objects generated at e<sub>3</sub> to e<sub>4</sub> gradually increased. In particular, several CD objects could not be decreased when the uncertainty level was decreased to Level 1 (Figures 15b and 16). In addition, the CD objects generated at e<sub>1</sub>–e<sub>4</sub> were similar, indicating that only some CD objects were newly added during the updating process. The CD results with the uncertainty decreasing from Level 3 to Level 1 resulted in OA = 0.8769 and F1 score = 0.9206 at Site 1 and OA = 0.9174 and F1 score = 0.9373 at Site 2. In addition, when the uncertainty changed from Level 2 to Level 1, OA = 0.9299 and F1 score = 0.9422 at Site 1 and OA = 0.9185 and F1 score = 0.9411 at Site 2. Grey, white, and black colors represent the change, no-change, and no-value classes, respectively.

Figures 15 and 16 show the CD results obtained using a decreased uncertainty level during the updating process. Thus, the proportion of pixels having the same class to define CD objects can be gradually increased. In particular, several CD objects could not be decreased when the uncertainty level was decreased to Level 1 (Figures 15b and 16). In addition, the CD objects generated at e<sub>1</sub>–e<sub>4</sub> were similar, indicating that only some CD objects were newly added during the updating process. The CD results with the uncertainty decreasing from Level 3 to Level 1 resulted in OA = 0.8735 and F1 score = 0.9193 at Site 1 and OA = 0.9299 and F1 score = 0.9422 at Site 2. In addition, when the uncertainty changed from Level 2 to Level 1, OA = 0.8611 and F1 score = 0.9608 at Site 1 and OA = 0.8180 and F1 score = 0.9446 at Site 2 (Table 5).
Table 5. Accuracy of the CD maps generated using the proposed method with different uncertainty levels.

| Study Site | Uncertainty Level | OA    | Precision | Recall | F1-Score |
|------------|------------------|-------|-----------|--------|----------|
| Site 1     | Level 1          | 0.8180| 0.9446    | 0.4262 | 0.5874   |
|            | Level 2          | 0.9174| 0.9373    | 0.7847 | 0.8542   |
|            | Level 3          | 0.8769| 0.6444    | 0.9206 | 0.7581   |
|            | Level 3 to Level 1| 0.8735| 0.6330    | 0.9193 | 0.7498   |
|            | Level 2 to Level 1| 0.9185| 0.9411    | 0.7846 | 0.8558   |
|            | Level 1 to Level 3| 0.8611| 0.9608    | 0.5780 | 0.7218   |
|            | Level 2 to Level 3| 0.9299| 0.8158    | 0.9422 | 0.8745   |
| Site 2     | Level 1          | 0.8998| 0.6252    | 0.8908 | 0.7348   |
|            | Level 2          | 0.9006| 0.6220    | 0.8991 | 0.7353   |
|            | Level 3          | 0.8802| 0.9359    | 0.5159 | 0.6651   |
|            | Level 3 to Level 1| 0.8987| 0.8908    | 0.6252 | 0.7348   |
|            | Level 2 to Level 1| 0.8957| 0.9146    | 0.6766 | 0.7776   |
|            | Level 1 to Level 3| 0.8859| 0.5223    | 0.9341 | 0.6700   |
|            | Level 2 to Level 3| 0.9012| 0.6164    | 0.9092 | 0.7347   |

5. Discussion

5.1. Comparison with Traditional CD Approaches

The changed pixels can be extracted using the unsupervised CD methods without the requirement of any training data based on the spectral difference between temporal images. These methods tend to perceive shadows and changes in the color of a substance caused by atmospheric effects as the real change. Therefore, spot and salt-and-pepper noise could be observed in the CD result maps. Depending on the objective of the study, unsupervised CD methods can be used appropriately when finding changes; however, they may be unsuitable to extract only the changes of the land cover materials. At Site 1, PCA showed the highest accuracies (OA = 0.8471 and F1 score = 0.7620), whereas CVA had the highest accuracies at Site 2 (OA = 0.8716 and F1 score = 0.7285).

CD was achieved using the CD network in four different cases. When compared with unsupervised CD methods, the CD result maps produced by the CD network involved only some salt-and-pepper noise, and the shadows around the building were not extracted as the changed area because the CD network used the initial CD map or the generated CD objects in which shadows were classified as no-class as label data. However, small objects such as buildings were underestimated. In particular, postprocessing can slightly change the no-change object. The accuracies of postprocessing were the highest at both the sites (OA = 0.8874 and F1 score = 0.7921 at Site 1 and OA = 0.8831 and F1 score = 0.6560 at Site 2).

The proposed method denotes the advantages of unsupervised CD methods and CD network with postprocessing. This method can generate training data for a deep learning network by generating label data even in areas in which prior information about changes is not available. Unlike the unsupervised CD methods, the proposed method does not produce salt-and-pepper noise and the shadows around trees and buildings were not extracted as the changed class. Furthermore, compared with the CD network using postprocessing, the proposed method can appropriately extract the shape of a building. The accuracies of the proposed method were OA = 0.9299 and F1 score = 0.8745 at Site 1 and OA = 0.9012 and F1 score = 0.7347 at Site 2.

5.2. The Effect of Uncertainty Level

When generating CD objects, a high uncertainty level indicates that the objects have a low percentage of the same class of pixels in an object. Thus, the higher is the level, the more easily is the objects can be assigned as change or no-change class even if the pixels within the object have
different classes. Therefore, the number of CD objects required to train the CD network increases, and many objects are updated at the defined epoch. Further, the percentage of the same pixels in the object increases if the uncertainty level is low. Therefore, the reliability of the object representing change or no-change can increase. However, in this case, the number of objects to be updated is small; therefore, the CD network results were severely changed. Therefore, not all the objects in the image were classified as change or no-change classes by increasing the epoch when the uncertainty levels were 1 and 2.

The experimental results showed that uncertainty Level 2 was appropriate for the two sites and gradually decreasing the level was considerably effective. In this case, the accuracies were observed to be the highest. When the CD network is trained, a large number of CD objects can be used as training data because the objects would not have been updated significantly if the training data did not change significantly.

5.3. The Effect of Segmentation Scale

The scale parameter is the most important factor during the segmentation process. In this study, different scale factors were applied to each experimental site because the optimal values may vary depending on the shape and size of the material in the image. The proposed method with different scales \( k \) was applied to analyze the effect of the scale factors. In the experiments, the uncertainty level was set based on the highest accuracies among the results in Section 4.3. Figure 17 shows the CD result obtained using different scales overlapping with the segment boundaries, and Table 6 gives the accuracies of the CD result maps. According to the results, the scale value can affect the accuracies of the proposed CD method. Optimal scale values can be observed at both the sites (Figure 18). \( k = 200 \) was the most effective value at Site 1, where bare soil changes were dominant, and \( k = 50 \) was the most effective at Site 2, where building changes were dominant. The optimal value \( k \) was related to the minimum size of the object of change to be extracted from the study site. If \( k \) was set to small regardless of the minimum size of changes, the CD objects would be inaccurate because the pixels in a small object can easily have the same class. Furthermore, if \( k \) becomes too large, the places at which change and no-change occurred can be considered as one object. Therefore, it is important to select an appropriate \( k \) based on the size of the object to be changed.

![Figure 17](image_url)

*Figure 17.* CD results using different segmentation scales overlapped with the segment boundaries at Sites 1 and 2: (a,f) scale 30; (b,g) scale 50; (c,h) scale 100; (d,i) scale 200; and (e,j) scale 300.
Table 6. Accuracy of CD maps generated using the proposed methods with different scale factors.

| Study Site | Scale Factors | OA   | Precision | Recall | F1-Score |
|------------|---------------|------|-----------|--------|----------|
| Site 1     | 30            | 0.8795| 0.6495    | 0.9256 | 0.7634   |
|            | 50            | 0.8969| 0.7194    | 0.9185 | 0.8068   |
|            | 100           | 0.8965| 0.7102    | 0.9267 | 0.8041   |
|            | 200           | 0.9299| 0.8158    | 0.9422 | 0.8745   |
|            | 300           | 0.9097| 0.7621    | 0.9228 | 0.8348   |
| Site 2     | 30            | 0.8920| 0.5627    | 0.9197 | 0.6982   |
|            | 50            | 0.9012| 0.6164    | 0.9092 | 0.7347   |
|            | 100           | 0.8915| 0.6063    | 0.8644 | 0.7127   |
|            | 200           | 0.8935| 0.5621    | 0.9306 | 0.7009   |
|            | 300           | 0.8931| 0.5632    | 0.9261 | 0.7004   |

Figure 18. F1 scores and OAs with different scale factors at: (a) Site 1; and (b) Site 2.

5.4. Limitations and Future Work

In the proposed method, the initial CD map generated various unsupervised CD methods. Although we used only the pixels that exhibited the same class in more than four methods among five when reducing the effect of noise and shadow, the initial CD map was observed to be dependent on the nature of unsupervised CD methods. For example, the region that changed from low vegetation to bare soil could not be extracted as a change class in the initial CD map because there was a minor difference between the two materials. In addition, buildings with dark roof similar to cement ground surface were not classified as change class in the initial map because the differences in spectral characteristics from the ground surface were not significant. Since the accuracy of the CD objects can affect the final CD results, it is important to produce qualified CD objects. To solve this issue, it might be helpful to add improved unsupervised CD methods, which can deal with fine spectral differences, to construct the initial CD map.

Furthermore, since the initial CD object can have a no-value, the deep learning network did not provide the best performance compared to that when training took place using the entire region. Therefore, we plan to develop a method that can provide appropriate initial values when learning with limited data based on transfer learning that can use information learned from similar tasks. Thus, the CD objects constructed at uncertainty Level 1 will become available. Finally, the method of automatically finding the optimal scale \( k \) considerably affects the performance of the proposed method and can reduce the difficulty associated with the empirical determination of the optimal value. Furthermore, the proposed method can be applied to a wide range of study sites.

6. Conclusions

A novel object-based CD method is proposed to detect the changes in VHR satellite images using deep learning networks without the requirement of ground truth data. The proposed method generated a pixel-based initial CD map using various unsupervised CD methods; further, the map
was reconstructed to produce CD objects, which have three classes: change, no-change, and no-value. To update the no-value objects, only two classes (change and no-change) of the CD objects were used to label the data for training the CD network. Further, objects were defined and updated according to the uncertainty level. The updated CD objects were used as the training data of the CD network. This process was iteratively conducted until the entire area was classified into two classes in object units or defined area. The experiments on Worldview-3 and KOMPSAT-3 datasets confirmed that the proposed method achieved the best performance when compared with the traditional CD approaches. In particular, uncertainty Level 2 was appropriate, and the changes at both the sites could be detected by decreasing the uncertainty level during the updating process. However, the performance of the proposed method can depend on the scale size; therefore, the optimal value should be established by considering the minimum size of the object of change to be extracted from the study site. Future work to automatically detect the optimal scale size and develop transfer learning is being conducted to overcome the limitation of insufficient training data caused by no-values in the CD objects.

Author Contributions: Conceptualization, Methodology, Software, Formal analysis, and Investigation, A.S., Y.K., and Y.H.; Resources, Validation, Data curation, Writing (original draft preparation), Funding acquisition, and Visualization, A.S.; and Writing (review and editing), Supervision, and Project administration, Y.K. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2019R1A6A3A0109230211) and by the Satellite Information Utilization Center Establishment Program of the Ministry of Land, Infrastructure, and Transport of Korean government, grant number 20SIUE-B148326-03.

Acknowledgments: The authors would like to thank the anonymous reviewers for their very competent comments and helpful suggestions.

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

References

1. Chen, G.; Weng, Q.; Hay, G.J.; He, Y. Geographic Object-based Image Analysis (GEOBIA): Emerging trends and future opportunities. GISci. Remote Sens. 2018, 55, 159–182. [CrossRef]
2. Lang, S. Object-based image analysis for remote sensing applications: Modeling reality-dealing with complexity. In Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications; Blaschke, T., Lang, S., Hay, G.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 3–27.
3. Laliberte, A.S.; Rango, A. Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery. IEEE Trans. Geosci. Remote Sens. 2009, 47, 761–770. [CrossRef]
4. Myint, S.W.; Gober, P.; Brazel, A.; Grossman-Clarke, S.; Weng, Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens. Environ. 2011, 115, 1145–1161. [CrossRef]
5. Pu, R.; Landry, S.; Yu, Q. Object-based urban detailed land cover classification with high spatial resolution ikonos imagery. Int. J. Remote Sens. 2011, 32, 3285–3308. [CrossRef]
6. Ye, S.; Pontius, R.G.; Rakshit, R. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. ISPRS J. Photogramm. 2018, 141, 137–147. [CrossRef]
7. Liu, D.; Xia, F. Assessing object-based classification: Advantages and limitations. Remote Sens. Lett. 2010, 1, 187–194. [CrossRef]
8. Gao, Y.; Mas, J.F. A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions. J. Earth Sci. 2008, 2, 27–35.
9. Definients Image. eCognition User’s Guide 4; Definients Image: Bernhard, Germany, 2004.
10. Hexagon Geospatial. ERDAS Imagine; Erdas Inc.: Madison, AL, USA, 2016.
11. Feature Extraction Module Version 4.6. In ENVI Feature Extraction Module User’s Guide; ITT Corporation: Boulder, CO, USA, 2008.
12. Chen, G.; Hay, G.J.; Carvalho, L.M.T.; Wulder, M.A. Object-based change detection. Int. J. Remote Sens. 2012, 33, 4434–4457. [CrossRef]
13. Stow, D. Geographic object-based image change analysis. In Handbook of Applied Spatial Statistics; Fischer, M.M., Getis, A., Eds.; Springer: Berlin, Germany, 2010; Volume 4, pp. 565–582.
14. Walter, V. Object-based classification of remote sensing data for change detection. *ISPRS J. Photogramm. Remote Sens.* 2004, 58, 225–238. [CrossRef]

15. Im, J.; Jensen, J.; Tullis, J. Object-based CD using correlation image analysis and image segmentation. *Int. J. Remote Sens.* 2008, 29, 399–423. [CrossRef]

16. Ma, L.; Li, M.; Blaschke, T.; Ma, X.; Tiede, D.; Cheng, L.; Chen, Z.; Chen, D. Object-based CD in urban areas: The effects of segmentation strategy, scale, and feature space on unsupervised methods. *Remote Sens.* 2016, 8, 761. [CrossRef]

17. Deselée, B.; Bogaert, P.; Defourny, P. Forest CD by statistical object-based method. *Remote Sens. Environ.* 2006, 102, 1–11. [CrossRef]

18. Chehata, N.; Orny, C.; Boukir, S.; Guyon, D.; Wigner, J.P. Object-based CD in wind storm-damaged forest using high-resolution multispectral images. *Int. J. Remote Sens.* 2014, 35, 4758–4777. [CrossRef]

19. Al-Khudhairy, D.; Caravaggi, I.; Giada, S. Structural damage assessments from Ikonos data using change detection, object-oriented segmentation, and classification techniques. *Photogramm. Eng. Remote Sens.* 2005, 71, 825–837. [CrossRef]

20. Niemeyer, I.; Marpu, P.; Nussbaum, S. CD using object features. In *Object-Based Image Analysis*; Blaschke, T., Lang, S., Hay, G., Eds.; Lecture Notes in Geoinformation and Cartography; Springer: Berlin/Heidelberg, Germany, 2008; pp. 185–201.

21. Han, Y.; Javed, A.; Jung, S.; Liu, S. Object-Based CD of Very High Resolution Images by Fusing Pixel-Based CD Results Using Weighted Dempster–Shafer Theory. *Remote Sens.* 2020, 12, 983. [CrossRef]

22. 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]

23. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* 2017, 9, 498. [CrossRef]

24. Papadomanolaki, M.; Vakalopoulou, M.; Karantzalos, K. A novel object-based deep learning framework for semantic segmentation of very high-resolution remote sensing data: Comparison with convolutional and fully convolutional networks. *Remote Sens.* 2019, 11, 684. [CrossRef]

25. Jin, B.; Ye, P.; Zhang, X.; Song, W.; Li, S. Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images. *J. Indian Soc. Remote Sens.* 2019, 47, 951–965. [CrossRef]

26. Liu, S.; Qi, Z.; Li, X.; Yeh, G.A. Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and sar data. *Remote Sens.* 2019, 11, 690. [CrossRef]

27. Song, A.; Kim, Y. Transfer Change Rules from Recurrent Fully Convolutional Networks for Hyperspectral Unmanned Aerial Vehicle Images without Ground Truth Data. *Remote Sens.* 2020, 12, 1099. [CrossRef]

28. Hussain, M.; Chen, D.M.; Cheng, A.; Wei, H.; Stanley, D. CD from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogram Remote Sens.* 2013, 80, 91–106. [CrossRef]

29. Dalla Mura, M.; Benediktsson, J.A.; Bovolo, F.; Bruzzone, L. An unsupervised technique based on morphological filters for CD in very high resolution images. *IEEE Geosci. Remote Sens. Lett.* 2008, 5, 433–437. [CrossRef]

30. Bruzzone, L.; Bovolo, F. A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. *Proc. IEEE* 2013, 101, 609–630. [CrossRef]

31. Singh, A. Review Article Digital CD techniques using remotely-sensed data. *Int. J. Remote Sens.* 1989, 10, 989–1003. [CrossRef]

32. Lu, D.; Mausel, P.; Brondizio, E.; Moran, E. CD techniques. *Int. J. Remote Sens.* 2004, 25, 2365–2401. [CrossRef]

33. Malila, W.A. Change vector analysis: An approach for detecting forest changes with Landsat. In *LARS Symposium*; Laboratory for Applications of Remote Sensing: West Lafayette, IN, USA, 1980.

34. Ilsever, M.; Unsalan, C. Two-Dimensional Change Detection Methods; Springer: London, UK, 2012.

35. Deng, J.; Wang, K.; Deng, Y.; Qi, G. Pca-based land-use CD and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* 2008, 29, 4823–4838. [CrossRef]
37. Nielsen, A.A. Multi-channel remote sensing data and orthogonal transformations for change detection. In *Machine Vision and Advanced Image Processing in Remote Sensing*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 37–48.

38. Zhang, G.; Jia, X.; Kwok, N.M. Super pixel based remote sensing image classification with histogram descriptors on spectral and spatial data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 4335–4338.

39. Felzenszwalb, P.; Huttenlocher, D. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* 2004, 59, 167–181. [CrossRef]

40. Schallner, L.; Rabold, J.; Scholz, O.; Schmid, U. Effect of Superpixel Aggregation on Explanations in LIME–A Case Study with Biological Data. *arXiv* 2019, arXiv:1910.07856. Available online: https://link.springer.com/chapter/10.1007/978-3-030-43823-4_13 (accessed on 22 July 2020).

41. Song, A.; Choi, J.; Han, Y.; Kim, Y. Change detection in hyperspectral images using recurrent 3D fully convolutional networks. *Remote Sens.* 2018, 10, 1827. [CrossRef]

42. Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 2015, 1, 802–810.

43. Acquarelli, J.; Marchiori, E.; Buydens, L.M.C.; Tran, T.; van Laarhoven, T. Spectral-spatial classification of hyperspectral images: Three tricks and a new learning setting. *Remote Sens.* 2018, 10, 1156. [CrossRef]

44. Li, Y.; Zhang, H.; Shen, Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* 2017, 9, 67. [CrossRef]

45. Aguilar, M.A.; del Mar Saldana, M.; Aguilar, F.J. Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images. *Int. J. Appl. Earth Obs. Geoinf.* 2013, 21, 427–435. [CrossRef]

46. Gašparović, M.; Dobrinić, D.; Medak, D. Geometric accuracy improvement of WorldView-2 imagery using freely available DEM data. *Photogramm. Rec.* 2019, 34, 266–281. [CrossRef]

47. Han, Y.; Choi, J.; Jung, J.; Chang, A.; Oh, S.; Yeom, J. Automated co-registration of multi-sensor orthophotos generated from unmanned aerial vehicle platforms. *J. Sens.* 2019, 2019. [CrossRef]

48. Han, Y.; Kim, T.; Yeom, J. Improved piecewise linear transformation for precise warping of very-high-resolution remote sensing images. *Remote Sens.* 2019, 11, 2235. [CrossRef]