*remote sensing*

MDPI

*Article*

# An Unsupervised Remote Sensing Image Change Detection Method Based on RVMamba and Posterior Probability Space Change Vector

**Jiaxin Song [1,2,3], Shuwen Yang [1,2,3,*], Yikun Li [1,2,3] and Xiaojun Li [1,2,3]**

[1] Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China; 13240120@stu.lzjtu.edu.cn (J.S.); liyikun@mail.lzjtu.cn (Y.L.); xjli@mail.lzjtu.cn (X.L.)
[2] National-Local Joint Engineering Research Center of Technologies, Lanzhou 730070, China
[3] Applications for National Geographic State Monitoring, Lanzhou 730700, China
* Correspondence: yangshuwen@mail.lzjtu.cn

**Abstract:** Change vector analysis in posterior probability space (CVAPS) is an effective change detection (CD) framework that does not require sound radiometric correction and is robust against accumulated classification errors. Based on training samples within target images, CVAPS can generate a uniformly scaled change-magnitude map that is suitable for a global threshold. However, vigorous user intervention is required to achieve optimal performance. Therefore, to eliminate user intervention and retain the merit of CVAPS, an unsupervised CVAPS (UCVAPS) CD method, RFCC, which does not require rigorous user training, is proposed in this study. In the RFCC, we propose an unsupervised remote sensing image segmentation algorithm based on the Mamba model, i.e., RVMamba differentiable feature clustering, which introduces two loss functions as constraints to ensure that RVMamba achieves accurate segmentation results and to supply the CSBN module with high-quality training samples. In the CD module, the fuzzy C-means clustering (FCM) algorithm decomposes mixed pixels into multiple signal classes, thereby alleviating cumulative clustering errors. Then, a context-sensitive Bayesian network (CSBN) model is introduced to incorporate spatial information at the pixel level to estimate the corresponding posterior probability vector. Thus, it is suitable for high-resolution remote sensing (HRRS) imagery. Finally, the UCVAPS framework can generate a uniformly scaled change-magnitude map that is suitable for the global threshold and can produce accurate CD results. The experimental results on seven change detection datasets confirmed that the proposed method outperforms five state-of-the-art competitive CD methods.

**Keywords:** unsupervised change vector analysis in posterior probability space; Mamba; context-sensitive Bayesian network (CSBN); change detection



**Citation:** Song, J.; Yang, S.; Li, Y.; Li, X. An Unsupervised Remote Sensing Image Change Detection Method Based on RVMamba and Posterior Probability Space Change Vector. *Remote Sens.* **2024**, *16*, 4656. https://doi.org/10.3390/rs16244656

Academic Editor: Andrea Garzelli

Received: 27 October 2024
Revised: 29 November 2024
Accepted: 9 December 2024
Published: 12 December 2024



## 1. Introduction

The application of remote sensing technology has gained widespread attention [1–5]. In particular, remote sensing-based CD has attracted considerable attention within and outside the remote sensing (RS) research community. The image difference method is one of the earliest techniques for change detection, and it identifies changed regions by calculating pixel-level differences between images taken at different time points. While simple and intuitive, this method is highly susceptible to false detections, particularly in the presence of significant noise interference. Principal component analysis (PCA) is a statistically based change detection technique that effectively reduces redundant information in data and enhances the extraction of key change features [6], particularly for complex change types. However, PCA primarily relies on global image features and may struggle to capture small-scale, local changes. Additionally, since PCA is based on linear transformations, it is less effective in detecting nonlinear changes. The post-classification comparison (PCC) method is commonly used in remote sensing change detection [7], where changes are identified by

comparing classification results before and after an event. A significant limitation of this method is the accumulation of classification errors, where inaccuracies in the classification of pre- and post-change images directly affect the accuracy of the change detection results.

Traditional change detection methods, while addressing the problem of remote sensing image change detection to some extent, are limited by the complex spectral, spatial, or structural information of images. These methods typically require manual parameter selection and preprocessing. Furthermore, they often fail to fully leverage the deep features and spatial information present in images, which hampers their application in large-scale real-time CD tasks. Alternatively, recently developed deep learning CD methods can perform supervised tasks more effectively and efficiently [8,9]. Based on a substantial volume of training data and high computational power, deep learning methods can automatically learn the parameters for feature extractors and CD models and relieve the requirement of manually determining critical parameters [10,11]. Therefore, deep learning-based methods are becoming more mainstream for change detection.

For instance, convolutional neural networks (CNNs) can extract rich spatial information from RS imagery [12,13]. Therefore, CNN-based CD methods have become popular in the RS community. For example, Ou proposes a CNN framework involving slow–fast band selection (SFBS) and feature fusion grouping (SFBS-FFGNET) for hyperspectral image change detection. Based on SFA, SFBS is proposed to select the slow characteristic band and the fast characteristic band to extract changed and unchanged features during change detection [14]. Zhang presented an ensemble deep learning framework for change detection in VHR images, where discriminative deep metric learning based on the dissimilarity degree is used to quantitatively adjust the discriminative distance metric of two CNN output layers [15].

To maximize the use of contextual and semantic information in RS images at the multiscale level, researchers are focused on developing a variety of CNNs with specific network structures. Autoencoders (AEs) [16], particularly UNet, have been widely used in RS CD tasks. UNet-based CD methods concatenate bitemporal images into multiband inputs, extract the change features using a Siamese encoder [17], and convert the CD task into a semantic classification task [18]. However, CNN models inherently suffer from a limited receptive field, and they cannot capture long-term dependencies between pixels. This issue primarily arises from the CNN design, which is focused on local receptive fields [19]. In addition, Siamese neural network models rely on the similarity and difference in features between images and have certain limitations for effectively distinguishing slight or cumulative long-term changes.

The emergence of transformers offers an effective solution to this issue. The transformer architecture effectively models pixel relationships across an entire image. For example, Zhang designed a pure transformer network with a Siamese U-shaped structure to solve CD problems and named it SwinSUNet [20]; it employs a U-shaped Siamese encoder architecture and achieves satisfactory CD performance. Based on Zhang's work, Jiang constructed the Forest-CD model using an encoder–decoder structure with the Swin transformer as the backbone to extract change features and efficiently model global information [21]. However, despite the strong performance of transformers [22,23], the self-attention mechanism demands quadratic computational resources proportional to the image size, resulting in a substantial computational burden as the input sequence length increases or the network depth grows, which creates challenges in large-scale remote sensing tasks. In contrast, the Mamba network enhances the S4 model [24] by incorporating time-varying parameters into a standard state space model (SSM) in combination with hardware optimization to boost training and inference efficiency. In particular, an SSM offers near-linear complexity by establishing long-range connections through state transitions executed via convolutional computation. Based on the Mamba architecture, Chen proposed the ChangeMamba model to address remote sensing image change detection [25]. As the core module of the ChangeMamba model, the cross-scan module (CSM) expands

image patches in different spatial directions to realize the effective modeling of the global context information of images.

However, supervised deep CD methods require substantial computational time, resources, and suitable training image datasets to obtain satisfactory results. In a specific scene, the training datasets and CD target images are too semantically and visually different. The accuracy of the above deep CD methods cannot be guaranteed.

In order to resolve this challenge, researchers have proposed unsupervised deep learning CD methods [26,27] that do not require an a priori annotated training dataset, thereby automating CD tasks. Saha proposed a novel unsupervised context-sensitive framework for deep change vector analysis (DCVA). To achieve an unsupervised system, DCVA begins with a suboptimal pre-trained multilayer CNN to extract deep features that model spatial relationships between neighboring pixels. However, DCVA depends on pre-trained network parameters, and if these parameters cannot be adapted to the features of the target change detection (CD) image, the desired level of performance may not be achieved. Additionally, the DCVA method lacks feedback from the CD results, which limits its generalization performance. Tang et al. [28] proposed an unsupervised CD method based on a graph convolutional network (GCN) and metric learning, GMCD, which yields satisfactory change detection (CD) results. However, similarly to DCVA, this method is also constrained by the semantic and spectral differences between the pre-training dataset and the target images. If the pre-trained network parameters cannot accommodate the characteristics of the target CD images, the observed CD performance cannot be achieved.

An alternative solution for overdependence on the training dataset is to collect training data within the target CD images. For instance, Yuan proposed a synthetic aperture radar (SAR) image change detection method based on weighted summation to generate a difference image and an optimized random forest [29]. The proposed disparity map generation algorithm has a better suppression effect on noise. However, the randomness introduced in the sparrow search algorithm may fall into a local optimal solution, affecting the final detection results. Chen et al. [30] proposed change vector analysis in the posterior probability space (CVAPS), which only requires the pixels of the target CD images as training samples and does not excessively rely on the training image dataset to obtain satisfactory CD accuracy. However, it is important to mention that the original CVAPS method relies on a support vector machine (SVM) for estimating pixel-level posterior probability vectors in remote sensing (RS) imagery, and the SVM cannot model the multiple-to-multiple correspondences between typical pixel spectra (signal classes) and ground-cover types. Therefore, Yang et al. [31] proposed a CVAPS method (FCM-SBN-CVAPS) coupled with fuzzy C-means clustering (FCM) and a simple Bayesian network (SBN). It establishes many-to-many stochastic links between signal classes and ground-cover types using the SBN. Then, the posterior probability vectors of each pixel in the bitemporal RS images are estimated to facilitate the CVAPS. Nevertheless, the SBN estimates the posterior probability vectors based solely on individual pixels without considering spatial information, i.e., neighboring pixels. This approach overlooks the influence of spatial information on the posterior probability estimation. Although the FCM-SBN-CVAPS method does not require semantically and visually similar training images, it requires users or domain experts to manually outline the training pixels within the target image. The reliance on manual intervention greatly limits the application potential of FCM-SBN-CVAPS because of its high labor and time costs, particularly when dealing with extensive datasets. To tackle the aforementioned issues, it is necessary to develop a highly accurate unsupervised HRRS segmentation algorithm.

Therefore, we propose an unsupervised CVAPS (UCVAPS) framework for the CD method, RFCC, where we use a context-sensitive Bayesian network (CSBN) [32] to estimate posterior probability vectors at the pixel level. This method is better suited to the HRRS CD task because of the inclusion of spatial information. To reduce the algorithm's dependence on manual input, this study proposes an unsupervised segmentation algorithm based on RVMamba with differentiable feature clustering, which automatically generates training samples for CSBN models. It also introduces a loss function incorporating feature similarity loss and spatial continuity loss [33,34] to mitigate the constraints of fixed boundaries and eliminate the need for training data. Therefore, the proposed RFCC method requires no human-annotated training data. It significantly lowers both labor and time expenses and simultaneously guarantees CD accuracy. The experimental findings confirm the superior performance of the proposed method over state-of-the-art unsupervised deep CD methods. The primary contributions of this research are presented below.

(1) We propose an unsupervised remote sensing image CD framework, RFCC, based on RVMamba and UCVAPS, which enables higher-accuracy CD prediction.
(2) We propose an unsupervised remote sensing image classification algorithm, RVMamba, with differentiable feature clustering, which introduces a visual state space model (SSM) to construct a network model that is substantial in terms of representational power and efficiency. The algorithm provides training samples for a CSBN in an unsupervised way.
(3) This is the first use of a CSBN in unsupervised remote sensing image change detection. A CSBN is used to estimate the posterior probability vector to obtain a change intensity map, which is more suitable for the HRRS CD task due to the incorporation of spatial information.
(4) A series of experiments on different datasets and comparisons with novel competitive CD methods demonstrate the excellent performance of our method.

The subsequent sections of this article are structured as follows: Section 2 presents an overview of the proposed methodology; Section 3 covers the datasets, experimental setup, performance evaluation, and comparisons with other methods; Lastly, Section 4 provides the conclusion of the study.

## 2. Methodology

### *2.1. Overview*

As shown in Figure 1, this method involves four distinct stages.

Step 1: Unsupervised classification of bitemporal images was performed to obtain training samples for the CSBN.

Step 2: The FCM and CSBN are coupled to estimate the posterior probability vector of pixels in the bitemporal images.

Step 3: The change-magnitude map is calculated by UCVAPS.

Step 4: Global thresholding is applied to binarize the change-magnitude map into a change map, and the morphology operation is used to remove small holes.

In the proposed method, the RVMamba model was integrated with a differentiable feature clustering method [34] to achieve unsupervised image classification. The classification method was used to automatically provide training samples for the CSBN to achieve UCVAPS CD without manual intervention. The CSBN incorporates spatial information within images; thus, the proposed CD method can effectively process HSSR imagery. Although the major shortcoming of the CSBN model is the high time cost caused by large neighboring windows, this can be alleviated by the parallel computing technique.

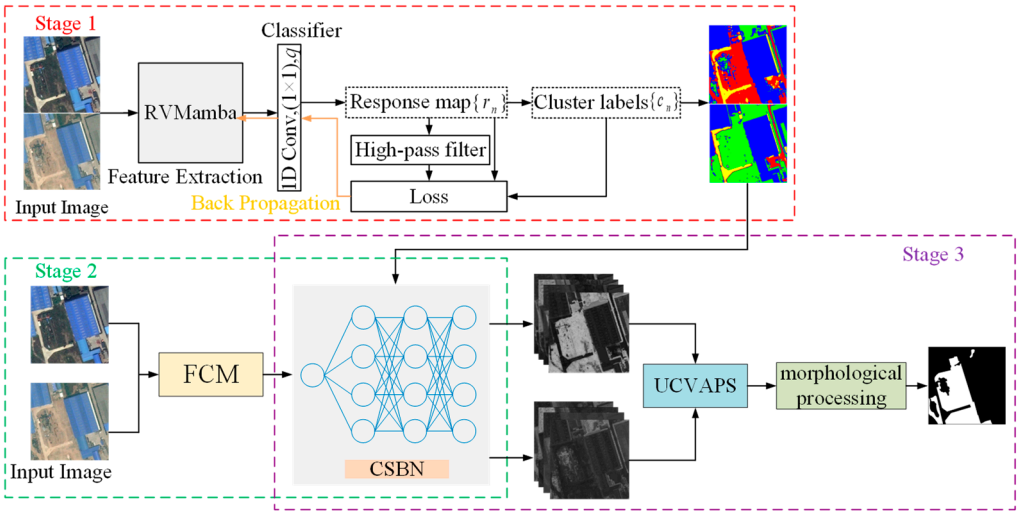


**Figure 1.** Unsupervised change detection process based on RVMamba and Posterior Probability.

### *2.2. Unsupervised RS Image Classification Based on RVMamba*

Kim [34] summarized the unsupervised image classification problem in the following conceptual model. Let $\{\}_{n=1}^{N}$ denote an RGB image composed of N pixels I = $\{p_n \in R^3\}_{n=1}^{N}$ or a d-dimensional feature image $\left\{f_n \in R^d\right\}_{n=1}^{N}$. Let $E : R^3 \rightarrow R^d$ be a feature extraction function satisfying $f_n = E(p_n)$. A classification label map $\{l_n \in Z\}_{n=1}^{N}$ is obtained from the classification function $C : R^d \rightarrow Z$ and satisfies $l_n = C(f_n)$. For the unsupervised classification problem, it is necessary to train the parameters of functions E and C with the unknown label image $\{l_n \in Z\}_{n=1}^{N}$ in a fully unsupervised manner. Therefore, the following two steps must be solved: (1) fixing the parameters of feature extraction function E and classification function C to estimate the labeled image $\{l_n \in Z\}_{n=1}^{N}$; (2) fixing the label images $\{l_n \in Z\}_{n=1}^{N}$ to update the parameters of feature extractor E and classifier C. The first uses known model parameters to estimate the labeled image $\{l_n \in Z\}_{n=1}^{N}$ (i.e., the labels of the image) based on the parameters of the feature extraction function $E$ and the classification function $C$. This process involves predicting the labels of the image according to the current model parameters. The second step, given the labeled image $\{l_n \in Z\}_{n=1}^{N}$, focuses on updating the parameters of the feature extraction function $E$ and the classification function $C$ to better fit the data. This optimization step aims to improve the model performance by adjusting the parameters based on the labeled data. These two steps are key to the model optimization process; the first step relies on fixed function parameters for label prediction, while the second step adjusts the model parameters based on the fixed labels. Thus, they are both crucial for optimizing unsupervised deep segmentation.

According to related studies [34], good unsupervised RS image classification results should be as close as possible to manual classification results. Humans are very likely to identify an image region corresponding to all or part of a ground-cover type with similar spectral and spatial features. In addition, one ground-cover type often contains several subregions with similar spectral and spatial features. Therefore, to mimic human classification, unsupervised classification algorithms must group spatially contiguous pixels with similar spectral and spatial features into a single ground-cover label. To distinguish regions from different ground-cover types, pixels with different spectral and textural features must be classified into different ground-cover labels. Kim [34] claimed that the classification strategy for generic imagery must be adjusted to generate a larger number of classified regions that may not be suitable for RS imagery. Therefore, in the RS context, three criteria must be introduced when predicting clustering labels: (a) pixels exhibiting comparable spectral and spatial features ought to be assigned to the same ground-cover label; (b) spatially contiguous pixels should be consolidated under a common ground-

cover label; (c) the number of ground-cover labels should reflect the ground-cover types within the RS images. However, these three criteria are often incompatible with actual RS image classification processes. They are difficult to satisfy simultaneously because suboptimal classical unsupervised classification algorithms (e.g., K-means and graph classification algorithms) can only be performed on a fixed feature image $\{f_n\}_{n=1}^{N}$. The parameters of the feature extraction function cannot be optimized because the parameters of classification function C can only be optimized based on a fixed feature image. Thus, their classification accuracy is limited. Therefore, Kim [34] proposed an unsupervised classification algorithm based on a CNN. This algorithm can optimize the feature image $\{f_n\}_{n=1}^{N}$ and labeled image $\{l_n\}_{n=1}^{N}$ in an iterative manner and, thus, better balance the above three classification criteria.

Due to their ability to learn hierarchical representations of image features, CNNs are widely employed to classify RS images [35]. They can automatically extract features and identify regions of interest. However, they have a limited ability to capture classification information in detail. Thus, although convolutional neural networks (CNNs) have shown strong performance in remote sensing image segmentation tasks, they exhibit notable limitations, particularly in capturing fine-grained changes. Key factors that hinder their ability to detect detailed changes include their restricted receptive fields, information loss caused by pooling operations, their inability to effectively process multi-scale information, and a limited understanding of the global context [19]. when processing HRRS images with enriched detailed information, CNNs often underclassify RS imagery. To overcome these limitations, many scholars have utilized Mamba-based networks for image classification problems [36], and these are built on state space models (SSMs). To balance the model's efficiency and effectiveness, selection mechanisms are introduced to control the propagation or interaction of information along specific sequence dimensions. By making the parameters that affect sequence interactions dependent on inputs, the model can establish long-distance dependencies while maintaining linear computational complexity. Therefore, this study utilizes the proposed RVMamba network as a feature extractor (E) for unsupervised classification. Compared with a CNN, the RVMamba architecture encodes and decodes images through top-down and bottom-up pathways, allowing for more accurate extraction of spectral and spatial features across multiple scales. Leveraging the S6 model, it effectively models long-range relationships with linear computational complexity, thereby achieving higher classification accuracy. We adopted the differentiable clustering argmax function and spatial continuity loss function proposed by Kim [34] to achieve end-to-end unsupervised network training. Unlike its CNN-based counterpart, the RVMamba-based unsupervised image classification algorithm can encode and decode images along top-down and bottom-up paths, which helps accurately extract more spectral and spatial features at multiple scales, thus yielding classification results with higher accuracy.

Aiming at generic images, the original CNN-based unsupervised image classification algorithm is susceptible to producing a high volume of labels and classification regions to improve clustering separability [34]. However, in our CD method, the CSBN model has a high computational complexity, and its computation time increases rapidly with an increasing number of labels. When a large number of labels are generated, too few pixels may be assigned to some labels. Hence, the training samples of certain labels are insufficient for the CSBN model, which may decrease the final CD accuracy. Therefore, the proposed classification algorithm removes the BatchNorm layer to generate a moderate number of classification labels for subsequent CSBN models.

#### 2.2.1. Network Architecture

##### Feature Similarity Constraint

The network architecture of the RVMamba-based RS image classification algorithm is illustrated in Stage 1 of Figure 2a. Taking into account the characteristics of remote sensing images, we developed a feature extraction network framework built on the Mamba architecture, as demonstrated in Figure 2. The Mamba architecture, along with its effi-

cient 2D cross-scanning mechanism (shown in Figure 3), enables the full extraction of robust and representative features from the input image. We developed an asymmetric encoder–decoder structure based on the Mamba architecture, in which the encoder uses four ResBlock modules to sample data and output features incrementally at each stage. Each ResBlock consists of three 3 × 3 convolutional layers, three batch normalization layers, and three ReLU functions. Conversely, the decoder is primarily responsible for stepwise feature fusion from the corresponding multilevel features extracted by the four-stage encoder. At the start of each stage, the VSS module models the global spatial context of the input data. Subsequently, the fusion module upsamples the feature maps and integrates them with higher-resolution low-level feature information. Finally, the output of the feature by the VSS block is upsampled and utilized to predict the corresponding land-cover maps. The structural diagram of the VSS block is illustrated in Figure 2b. The output is initially passed through a linear embedding layer and then divided into two information streams. One stream is processed by a 3 × 3 depthwise convolution (DWConv) layer. After applying the SiLU activation function, it moves into the core SS2D module. The output from the SS2D module passes through a normalization layer, which performs element-wise multiplication with the output from the first branch to merge the two paths. Finally, the features are blended using a linear layer, and the sub-results are combined with residual connections to generate the output of the VSS block.

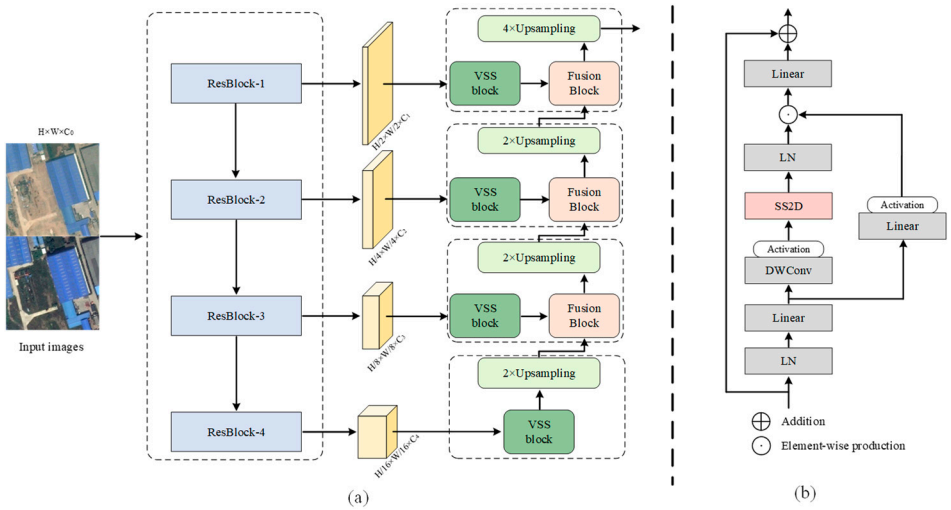


**Figure 2.** Feature extraction network for visual state space modeling. (**a**) The overarching design of RVMamba. (**b**) VSS block; SS2D is the core operation in VSS block.

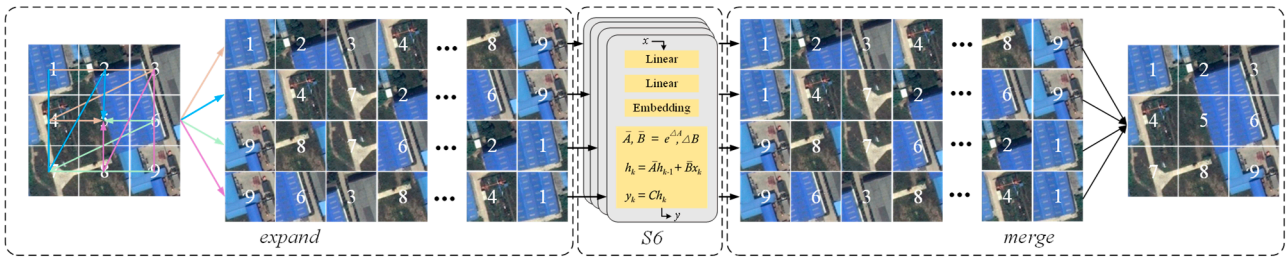


**Figure 3.** Data flow of SS2D. It expands the inputs in four directions according to the serial number, scans them one by one through S6, and then merges them.

In terms of noise processing, the encoder in the RVMamba model utilizes a network structure based on four ResBlock modules. In the presence of strong noise, the residual structure effectively prevents overfitting to the noise, thereby maintaining the stability of the classification performance. Additionally, at the beginning of each encoding stage,

the VSS module smooths local noise in the image through global context modeling, reducing the impact of noise on model performance. The VSS module processes features in detail by combining depthwise convolution with the SiLU activation function, effectively suppressing noise during feature extraction.

The VSS module structure includes a normalization layer, through which feature maps are processed by the deep convolutional layer. This normalization helps standardize the input features, reducing the influence of outliers during model training. Furthermore, the VSS module employs two information stream branches: one passes through a $3 \times 3$ deep convolutional layer, while the other stream is processed through the core of the SS2D module. The interaction between these branches allows the model to fuse features from different sources and effectively smooth outliers. This structure enables the model to adapt to outliers through flexible information flow channels, making the classification results robust to outliers.

$$\begin{aligned} h'(t) &= Ah(t) + Bx(t) \\ y(t) &= Ch(t) \end{aligned} \tag{1}$$

Equation (1) describes the system state over time. Here, $h(t) \in \mathbb{R}^N$ represents the state vector of the system at time $t$, and $x(t)$ is the input vector, which corresponds to the external signal or control input. The matrix $A \in \mathbb{R}^{M \times N}$ is the state transition matrix that governs the evolution of the system's internal state, while $B \in \mathbb{R}^{N \times 1}$ is the input matrix that dictates how the external input $x(t)$ influences the state $h(t)$. This equation can be interpreted as a linear ordinary differential equation (ODE) for modeling the system state dynamics over time. $y(t) \in \mathbb{R}^1$ is the output of the system at time $t$, and $C \in \mathbb{R}^{1 \times N}$ is a projection matrix that maps the high-dimensional state $h(t)$ to a scalar output.

In deep learning methods, ODEs must be discretized. The Selective Scanning Spatial State Sequence Model (S6) [37] is a discrete counterpart of a continuous system. The time-scale parameter $\Delta$ converts the continuous parameters $A$ and $B$ into discrete parameters $\overline{A}$ and $\overline{B}$ using the commonly applied zero-order hold and first-order Taylor series expansion.

$$\begin{aligned} \overline{A} &= e^{\Delta A} \\ \overline{B} &= \left(e^{\Delta A} - I\right) A^{-1} B \approx (\Delta A)(\Delta A)^{-1} B = \Delta B \end{aligned} \tag{2}$$

$\Delta$ is a time-scale parameter that defines the discretization step of the control system, representing the transition from continuous to discrete time. The ODEs of (2) can be rewritten as follows.

$$\begin{aligned} h_k &= \overline{A} h_{k-1} + \overline{B} x_k \\ y_k &= C h_k \end{aligned} \tag{3}$$

Equation (3) represents a discretized state transfer equation that describes the system state update process at time step $k$. Here, $h_k$ denotes the state vector of the system at time step $k$, and $\overline{A}$ is a discretized state transition matrix, which is obtained by discretizing the exponential function $e^{\Delta A}$ of the state transfer matrix $A$ from the continuous time system. $\overline{B}$ is a discretized input matrix that describes how the external input $x_k$ affects the current state $h_k$. $y_k$ is the output at time step $k$, and $C$ is the projection matrix. In S6, the matrices B, C, and 1 are derived from the input $x_k$ [37]. In S6, the matrices B, C, and 1 are derived from the input $x_k$ [37].

In Equation (2), the time-scale parameter $\Delta$ is used to discretize the state transition matrix $\overline{A}$ and the input matrix $\overline{B}$ of the continuous system into corresponding parameters for the discrete time system. This discretization is achieved using common methods such as zero-order hold and first-order Taylor expansion, which transform the continuous model into a form suitable for numerical computation. This approach simplifies the continuous system's dynamics, enabling efficient training and inference on a computer.

Spatiotemporal variations in remotely sensed images are often highly detailed and exhibit multiscale characteristics. Continuous systems are generally better suited to capturing these subtle spatiotemporal changes. In the case of complex remote sensing images,

overly simplistic discretization methods may result in the loss of critical spatiotemporal information.

Liu et al. [38] introduced SSMs to visual tasks and proposed the two-dimensional selective scanning (SS2D) method, as shown in Figure 3. SS2D extends image patches in four directions, generating four independent sequences. These sequences are then individually processed by an SSM, and the extracted features are combined to generate a complete 2D feature map. Based on the input feature map $x$, the output feature map $\overline{x}$ from SS2D is expressed as follows:

$$\begin{aligned} x_v &= expand(x, v) \\ \overline{x}_v &= S6(x, v) \\ \overline{x} &= merge(\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4) \end{aligned} \tag{4}$$

Here, $v \in \{1, 2, 3, 4\}$ denotes four distinct scan directions. Furthermore, expand(.) and merge(.) are the scan-expand and scan-merge operations, respectively.

To group pixels with similar features into identical labels, a classifier is used to classify each pixel into q ground-cover labels. Suppose that the input RGB image is $I = \{p_n \in R^3\}$. All its pixel values are normalized to the interval [0, 1]. To extract the d-dimensional features $\{f_n\}$ from $\{p_n\}$, we use RVMamba (Figure 2) as the feature extractor, and the input image $I$ goes through a total of nine convolutional components, each comprising four 2D convolution layer blocks, two normalization functions, and two ReLU functions. The encoder uses multilayer convolution and pooling operations to downsample the image, gradually decreasing the resolution of the feature map in spatial terms and extracting high-level feature representations. Each convolutional component is downsampled using maximum pooling to expand the receptive field. The decoder then uses deconvolution and upsampling operations to recover the feature map, which has been resized to match the original image dimensions, whereas jump connections are used to preserve the details. Subsequently, responsive mapping is obtained using the classifier $\{r_n = W_c f_n\}$, where $W_c \in R^{q \times d}$, and $W_c$ is the parameter of the classifier. Finally, based on the argmax clustering function, the ground-cover labels $l_n$ of pixel $p_n$ are indexed to the maximum value in $r_n$; thus, the feature vector $\{f_n\}$ is clustered into $q$ ground-cover labels, and its clustering formula is as follows:

$$l_n = \left\{ i \middle| r_{n,i} \geq r_{n,j}, r_n \in R^q \forall j \right\} \tag{5}$$

Constraints with the Number of Label Clusters

The number of labels $q$ for unsupervised segmentation reflects the spectral and spatial characteristics of the classified imagery. Thus, the number of segmentation labels varies with different images, where too large of a value of $q$ indicates over-segmentation, and too small of a value of $q$ indicates under-classification. As described in Section 2.2.1, the proposed method classifies a target image using a predetermined number of labels $q$. The value of $q$ is initialized to the maximum permitted number of ground-cover labels. The number of labels is constrained by the loss function, reflecting the feature similarity and spatial continuity during network training. The clustering function based on the argmax classification corresponds to $q$-class clustering. Typically, the response map $\{r_n\}$ is normalized by the BatchNorm module before cluster labels are assigned using the argmax classification. However, BatchNorm converts the original response map $\{r_n\}$ into $\{r'_n\}$. This gives each clustering index $i$ of $r'_{n,i}$ $(i = 1, \cdots, q)$ an equal chance to be the maximum clustering index, which causes a large clustering number $q$. Therefore, this normalization process causes the proposed RVMamba to tend towards generating a large number of pseudo-ground-cover labels and slow its convergence speed. However, a remote sensing image normally contains a limited number of ground-cover types. Therefore, it is necessary to ensure that the number of pseudo-ground-cover labels is in line with the number of real ground-cover types, which are needed to eliminate the BatchNorm layer.

2.2.2. Loss Function

The loss function comprises feature similarity loss and spatial continuity loss (Equation (6)), and the network weights are updated by minimizing this loss function.

$$L = L_{sim}(\{r_n, l_n\}) + \mu L_{con}(\{r_n\}) \tag{6}$$

$\mu$ denotes the weights that balance the feature similarity loss ($L_{sim}$) and the spatial continuity loss ($L_{con}$), which are described as follows.

Unsupervised ground-cover labels $l_n$ are obtained using the argmax function based on the response $\{r_n\}$. The ground-cover labels are then used as training labels to calculate the feature similarity constraints based on $\{r_n\}$ and $\{l_n\}$.

$$\begin{gathered} L_{sim}(\{r_n, l_n\}) = \sum_{n=1}^{N} \sum_{i=1}^{q} -\delta(i - l_n) \ln r_{n,i} \\ \delta(i - l_n) = \begin{cases} 1, & if\ i - l_n = 0 \\ 0, & otherwise \end{cases} \end{gathered} \tag{7}$$

$\delta$ is a selection function that is used to select the corresponding response value $ln\ r'_{n,i}$ of the pseudo-ground-cover label $l_n$ of the pixel $n$.

The feature similarity loss facilitates the similarity of features; the features of pixels corresponding to the same label are similar, and the features of pixels related to different labels are different. The feature similarity loss is minimized using network backpropagation.

In image segmentation, pixels corresponding to identical labels are spatially continuous. Therefore, the introduction of spatial continuity loss is prone to generating identical labels for neighboring pixels. This process is implemented using differential operators, based on which the spatial continuity loss $L_{con}$ can be defined as follows:

$$L_{con}(\{r_n\}) = \sum_{\xi=1}^{W-1} \sum_{\eta=1}^{H-1} ||r_{\xi+1,\eta} - r_{\xi,\eta}||_1 + ||r_{\xi,\eta+1} - r_{\xi,\eta}||_1 \tag{8}$$

where $W$ and $H$ represent the height and width of the input image, and $r_{\xi,\eta}$ denotes the response mapping value $\{r_n\}$ at the location $(\xi, \eta)$. The utilization of the spatial continuity loss $L_{con}$ can prevent the algorithm from overestimating the number of labels due to the complex spectral and textural characteristics of RS imagery.

2.2.3. Backpropagation

This section describes the training strategy used in the proposed RVMamba-based unsupervised RS image segmentation. The segmentation of target images involves two steps: (1) fixating network parameters to estimate ground-cover labels and (2) fixating the ground-cover labels to update the network parameters. In general, the first procedure is accomplished using forward propagation, whereas the second is accomplished using backward propagation. The proposed method performs backpropagation based on the loss function L described in Section 2.2.2 to update the convolution filter parameter and linear classifier parameter. Similarly to [28], the parameter initialization algorithm is the Xavier initialization algorithm [33]. This forward–backward process iterates $J$ times to obtain the final ground-cover labels {}, which have a low computational cost and time, as indicated by the experiments.

In addition, the learning rate and momentum values in the stochastic gradient descent (SGD) algorithm play a crucial role in achieving a balance between parameter updates and label segmentation. However, if these values are not appropriately tuned, this can lead to inconsistent segmentation results. Therefore, based on our experimental analysis, setting the learning rate to 0.1 and the momentum value to 0.9 can significantly improve the overall segmentation performance.

### *2.3. FCM-CSBN-UCVAPS Framework*

In this section, the mechanism of the FCM-CSBN-UCVAPS and change detection procedures and methods are elucidated.

Firstly, the FCM enables reliable unsupervised detection of signal classes by leveraging the abundant pixel-level data available in HRRS images. Secondly, each signal class is associated with all pixels in an image through fuzzy memberships, which facilitates the training of the CSBN. As a result, information from a limited number of training pixels becomes linked to the entire image, significantly improving training efficiency. Thirdly, the FCM allows for the compact representation of a large number of spectral vectors at the pixel level using a smaller set of signal classes. This feature enhances the robustness of RFCC by reducing the training sample randomness associated with RVMamba and random training sample selection. Fourthly, with the incorporation of fuzzy memberships, the CSBN can effectively model different ground-cover labels that share the same spectrum, as well as identical ground-cover labels with different spectra. Finally, the soft-clustering nature of the FCM, combined with the exploitation of all fuzzy memberships by the CSBN, leads to a substantial reduction in cumulative clustering errors within the proposed method.

Related studies have demonstrated the effectiveness of incorporating spatial information to address spectral variability, reduce uncertainty in HRRS images, and improve change detection (CD) accuracy [32,39]. In line with these findings, we developed a CSBN to establish random associations between signal pairs representing various spectra and types of ground cover. This innovative approach allows for the modeling of scenarios where the same ground-cover type exhibits various spectra or where multiple ground-cover types have identical spectral signatures. Moreover, by incorporating signal pairs within a specific neighborhood, the CSBN effectively incorporates pixel-level spatial information, enabling it to handle the high spectral variability and complex structural characteristics of ground objects in HRRS images. Consequently, within the UCVAPS framework, the synergistic utilization of the FCM and CSBN significantly enhances the performance of unsupervised HRRS CD.

#### 2.3.1. Unsupervised CVAPS

Given that $p_{i,j}$ is a pixel in row $i$ and column $j$ within an RS image I, the mixed pixel $p_{i,j}$ contains the automatically classified ground-cover labels $L_1$ and $L_2$, and the spectrum of the pixel changes with a small degree at times $t_1$ and $t_2$, insignificantly perturbing the estimated posterior probability vectors. Suppose that the posterior probabilities $P(L_1|p_{i,j})$ and $P(L_2|p_{i,j})$ at time $t_1$ are 51% and 49%, respectively. Then, under the maximum a posteriori (MAP) principle, the pixel should be labeled as $L_1$. Similarly, at time $t_2$, the posterior probabilities $P(L_1|p_{i,j})$ and $P(L_2|p_{i,j})$ are 49% and 51%, respectively. Therefore, the pixel $p_{i,j}$ is classified as $L_2$. Therefore, the MAP principle identifies pixels $p_{i,j}$ as changed pixels, regardless of their insignificant spectral differences in bitemporal images. A false alarm must be attributed to unsupervised classification errors that cannot be effectively handled by the MAP principle.

Inspired by the CVAPS framework proposed by Chen et al. [29], this study proposes a novel UCVAPS method that assumes that $\alpha_v^1$ and $\alpha_v^2$ are equal to the posterior probability $P(L_v|p_{i,j})$ at times $t_1$ and $t_2$, respectively. Then, the posterior probability vectors of the pixel at time $t_1$ and $t_2$ are denoted as $\boldsymbol{\alpha}^1 = (\alpha_1^1, \ldots, \alpha_v^1, \ldots, \alpha_m^1)$ and $\boldsymbol{\alpha}^2 = (\alpha_1^2, \ldots, \alpha_v^2, \ldots, \alpha_m^2)$, respectively, where m is the number of unsupervised classification labels in the bitemporal images. Therefore, the unsupervised change vector $\Delta\boldsymbol{\alpha}$ in the posterior probability space of pixel $p_{i,j}$ is defined as follows:

$$\Delta\boldsymbol{\alpha} = \boldsymbol{\alpha}^1 - \boldsymbol{\alpha}^2 \tag{9}$$

The change magnitude $\|\Delta\boldsymbol{\alpha}\|$ of the pixel $p_{i,j}$ can be measured using Euclidean distance, which is expressed as follows.

$$\|\Delta\boldsymbol{\alpha}\| = \sqrt{\sum_{v=1}^{m}\left(\boldsymbol{\alpha}_v^1 - \boldsymbol{\alpha}_v^2\right)^2} \tag{10}$$

According to Equation (10), a slight change in the posterior probability vectors does not produce a high change magnitude $\|\Delta\boldsymbol{\alpha}\|$, thus alleviating the elevated false alarm rate caused by the MAP principle. Instead of comparing ground-cover labels, UCVAPS compares posterior probability vectors to reduce the cumulative classification error caused by RVMamba. Although CVAPS is designed for multiple change detection, it requires accurately detected change regions to guarantee its multiple change detection performance, because the change types have to be predicted from previously detected change regions. The proposed UCVAPS exhibits good binary change detection performance, superior to state-of-the-art competitive approaches, and, thus, can provide sound binary change detection results for subsequent multiple change detection. Moreover, CVAPS can normalize the change-magnitude map in a supervised way, while UCVAPS can normalize the change-magnitude map without supervision, which has high application value in certain real-world scenarios.

#### 2.3.2. Context-Sensitive Bayesian Network Model

Li and Bretschneider [39] initially introduced the CSBN, which effectively leverages spatial information for posterior probability estimation. However, it cannot deal with cumulative clustering errors and mixed pixels when it uses the K-means algorithm to fulfill CD tasks. While the FCM can decompose a single pixel into multiple signal classes with varying degrees of membership, integrating a space-sensitive CSBN with the FCM during the estimation of posterior probabilities and the training stages can address the limitations associated with mixed pixels. In this study, a signal class refers to a cluster of pixels with similar spectral characteristics. This section offers a concise overview of the CSBN model.

We built a CSBN model, as shown in Figure 4. The construction of this model involves four levels to estimate the posterior probability $P\left(L_v \middle| p_{i,j}\right)$: (1) the pixel $p_{i,j}$; (2) the pixel pairs $\left(p_{i,j}, p_{u,v}\right)$; (3) the signal class pairs $(\omega_m, \omega_n)$; (4) the ground-cover label of the remotely sensed image $L_v$.

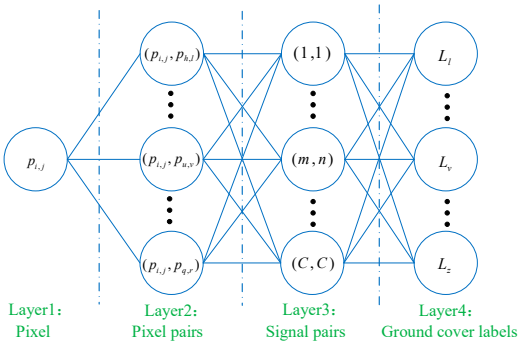


**Figure 4.** Context-sensitive Bayesian network model.

For a given remotely sensed image I, based on the pixel $p_{i,j} \in$ I, its adjacent pixels $p_{u,v}$ in neighborhood $N_{i,j}$, and the indices $(m, n)$ of the signal class pairs $(\omega_m, \omega_n)$, we compute $P\left(L_v \middle| p_{i,j}\right)$ using Equation (11).

$$P\left(L_v \middle| p_{i,j}\right) = \sum_{p_{u,v} \in N_{i,j}}^{n} P\left(L_v \middle| \left(p_{i,j}, p_{u,v}\right)\right) P\left(\left(p_{i,j}, p_{u,v}\right) \middle| p_{i,j}\right) \tag{11}$$

According to the conditional independent assumption and chain rule,

$$P\left(L_v\middle|\left(p_{i,j},p_{u,v}\right)\right)=\sum_{1\le m,n\le C}P(L_v|(m,n))P\left((m,n)\middle|\left(p_{i,j},p_{u,v}\right)\right) \tag{12}$$

Therefore, $P\left(L_v\middle|p_{i,j}\right)$ can be rewritten as:

$$\begin{aligned}P\left(L_v\middle|p_{i,j}\right)=&\sum_{p_{u,v}\in N_{i,j}}^{n}P\left(\left(p_{i,j},p_{u,v}\right)\middle|p_{i,j}\right)\\&\times\sum_{1\le m,n\le C}P(L_v|(m,n))P\left((m,n)\middle|\left(p_{i,j},p_{u,v}\right)\right)\end{aligned} \tag{13}$$

According to the Bayesian rule, $P(L_v|(m,n))$ can be expressed as:

$$P(L_v|(m,n))=\frac{P((m,n)|L_v)}{P((m,n))} \tag{14}$$

Therefore, $P\left(L_v\middle|p_{i,j}\right)$ can be rewritten as:

$$\begin{aligned}P\left(L_v\middle|p_{i,j}\right)&=\sum_{p_{u,v}\in N_{i,j}}^{n}P\left(\left(p_{i,j},p_{u,v}\right)\middle|p_{i,j}\right)\times\sum_{1\le m,n\le C}^{n}\frac{P((m,n)|L_v)}{P((m,n))}P\left((m,n)\middle|\left(p_{i,j},p_{u,v}\right)\right)\\&=P(L_v)\sum_{p_{u,v}\in N_{i,j}}^{n}P\left(\left(p_{i,j},p_{u,v}\right)\middle|p_{i,j}\right)\times\sum_{1\le m,n\le C}^{n}\frac{P((m,n)|L_v)}{P((m,n))}P\left((m,n)\middle|\left(p_{i,j},p_{u,v}\right)\right)\end{aligned} \tag{15}$$

$P((m,n))$ represents the prior probability of signal pairs $(\omega_m,\omega_n)$, while the conditional probability $P((m,n)|L_v)$ models the stochastic relationships between signal pairs $(\omega_m,\omega_n)$ and the ground-cover label $L_v$. These labels are obtained from randomly selected training samples in unsupervised classification images. In this study, we assume that the prior probability $P(L_v)$ follows a uniform distribution. The probability $P\left((m,n)\middle|\left(p_{i,j},p_{u,v}\right)\right)$ is determined using the corresponding fuzzy memberships.

$$P\left((m,n)\middle|\left(p_{i,j},p_{u,v}\right)\right)=P\left(m\middle|p_{i,j}\right)P(n|p_{u,v})=u_m(i,j)\times u_n(u,v) \tag{16}$$

Here, $P\left(m\middle|p_{i,j}\right)$ and $P(n|p_{u,v})$ denote the probability of pixels $p_{i,j}$ and $p_{u,v}$, belonging to signal classes $\omega_m$ and $\omega_n$. It is obvious that $P\left(\omega_m\middle|p_{i,j}\right)$ and $u_m(i,j)$, as well as $P(\omega_n|p_{u,v})$ and $u_n(u,v)$, share the same constraints. Therefore, it is reasonable to approximate $P\left(\omega_m\middle|p_{i,j}\right)$ by $u_m(i,j)$ and $P(\omega_n|p_{u,v})$ by $u_n(u,v)$, which effectively integrates the CSBN with the FCM.

This study assumes that each pixel $p_{u,v}$ in neighborhood $N_{i,j}$ is equally important for estimating $P\left(L_v\middle|p_{i,j}\right)$. Thus, $P\left(\left(p_{i,j},p_{u,v}\right)\middle|p_{i,j}\right)=1/\left|N_{i,j}\right|$, where $\left|N_{i,j}\right|$ denotes the number of pixels in the local neighborhood $N_{i,j}$.

In order to compute the posterior probability $P\left(L_v\middle|p_{i,j}\right)$, we need to learn the probability $P((m,n)|L_v)$ based on a pre-provided training set $T_v$. This training set is randomly selected from the unsupervised classification image, which contains pixels associated with the label $L_v$. The estimation equation depends on the frequency of occurrence of the signal pair $(\omega_m,\omega_n)$ corresponding to the label $L_v$ in the classification image.

#### 2.3.3. FCM Algorithm

The FCM can decompose a single pixel into different signal classes with varying fuzzy memberships, addressing the mixed-pixel problem that is commonly encountered in remote sensing images. Therefore, we use the FCM to calculate the fuzzy membership, establishing random links between pixels and signal classes to enhance the accuracy of change detection.

Let $I=\left\{p_{i,j}\middle|1\le i\le N,1\le j\le M\right\}$ be an $M\times N$ remote sensing image, which is fuzzily categorized into C signal classes; $u_k(i,j)(1\le k\le C)$ denotes the fuzzy member-

ship degree of $p_{i,j}$ to $\omega_k$, and the fuzzy membership degree $U = \{u_k(i,j)\}$ satisfies the following constraint:

$$\begin{gathered} u_k(i,j) \in [0,1] \quad \forall(i,j,k) \\ \sum_{k=1}^{C} u_k(i,j) = 1 \quad \forall(i,j) \\ 0 < \sum_{\substack{1 \le i \le N \\ 1 \le j \le M}} u_k(i,j) < N \times M \quad \forall k \end{gathered} \tag{17}$$

The FCM algorithm iteratively minimizes this cost function.

$$J(U,\Psi) = \sum_{\substack{1 \le i \le N \\ 1 \le j \le M}} \sum_{k=1}^{C} (u_k(i,j))^q \left\| p_{i,j} - \psi_k \right\|^2 \tag{18}$$

Here, $\Psi = \{\psi_1, \cdots, \psi_k, \cdots \psi_n\}$ denotes the set of signal class centers, and $\psi_k$ denotes the center of signal class $\omega_k$. $q$ determines the fuzziness of the clustering result. We can see from Equations (17) and (18) that $P(\omega_n | p_{i,j})$ and $u_k(i,j)$ satisfy the same constraints. Therefore, we use $u_k(i,j)$ to estimate $P(\omega_n | p_{i,j})$, thus realizing the combination of the CSBN and FCM.

#### 2.3.4. Fuzzy Membership-Based CSBN Model Training

To compute the posterior probability $P(L_v | p_{i,j})$, it is necessary to learn the probability $P((m,n)|L_v)$ based on the training set $T_v$ provided by the unsupervised segmentation of RVMamba, which contains the pixels $p_{x,y}$ belonging to pseudo-ground-cover type $L_v$. The estimation equation is based on the frequency of the signal pair $(\omega_m, \omega_n)$ within the pseudo-ground-cover type $L_v$, which is defined as follows:

$$SPF_v(m,n) = \sum_{\substack{p_{x,y} \in T_v \\ p_{u,v} \in N_{x,y}}} u_m(x,y) \times u_n(u,v) \tag{19}$$

Unlike hard membership, fuzzy memberships $u_m(x,y)$ and $u_n(u,v)$ take values ranging from 0 to 1. Therefore, it is necessary to sum the fuzzy membership values of the signal pairs $(\omega_m, \omega_n)$ corresponding to all pixel pairs $\left(p_{x,y}, p_{u,v}\right)$ in the training set $T_v$ in order to derive the signal pair frequency $SPF_v(m,n)$.

Based on the signal pair frequency $SPF_v(\vartheta,\mu)$ for each signal class pair $\left(\omega_\vartheta, \omega_\mu\right)$ $(\vartheta, \mu = 1, \cdots C)$, the conditional probability $P((m,n)|L_v)$ can be approximated as follows.

$$P((m,n)|L_v) \approx P((m,n)|L_v, T_v) = \frac{SPF_v(m,n)}{\sum_{\vartheta,\omega} SPF_v(\vartheta,\mu)} \tag{20}$$

Furthermore, in computing $P(L_v|(m,n))$, the prior probability $P((m,n))$ is calculated using the following total probability formula:

$$P((m,n)) = \sum_v P((m,n)|L_v) P(L_v) \tag{21}$$

The pseudo-code for the RFCC Model for CD is presented in Algorithm 1.

**Algorithm 1. Inference of RFCC Model for CD**

Input: $I = \{(I^1, I^2)\}$ (a pair of registered images)
Output: M (a prediction change mask)
//step1: RVMamba unsupervised image segmentation
For j = 1 to J
$\{f_n\}$ Extract Features $(\{p_n\}, \{w_m, b_m\})$
$\{r_n\}\{W_c f_n\}$ Generate Responsive Map
$\{l_n\} argmax(r_n)$ Generate Label Image
$L$ $L_{sim}(\{r_n, l_n\}) + \mu L_{con}(\{r_n\})$
$\{w_m, b_m\}_{m=1}^{9}, \{W_c\}$
Backpropagation
//step2: FCM-CSBN model
$P(L_v|p_{i,j}) = P(L_v)\sum_{p_{u,v}\in N_{i,j}}^{n} P((p_{i,j}, p_{u,v})|p_{i,j}) \times \sum_{1\le m,n\le C}^{n} \frac{P((m,n)|L_v)}{P((m,n))} P((m,n)|(p_{i,j}, p_{u,v}))$
//step3: CVAPS
$\Delta\alpha = \alpha^2 - \alpha^1$
$M = \sqrt{\sum_{v=1}^{m}(\alpha_v^2 - \alpha_v^1)^2}$

## 3. Experimental Validation

### *3.1. Dataset Description*

As shown in Figure 5, seven datasets were used in this experiment, with detailed analyses conducted on datasets 1, 2, and 3 to evaluate the CD performance across various metrics. The characteristics of these three datasets range from simple to complex, allowing us to examine the relationship between the optimal window size of the RFCC method and the spatial complexity of remote sensing imagery. In particular, this analysis aims to validate the importance of considering neighboring pixels in the CSBN. DS1, DS2, and DS3 used in this study were obtained from the Shangtang SenseEarth Archive [40]. We validated the adaptivity of the algorithm using the DS4 (GF-2), DS5 (Landsat8), DS6 (CD_Data_GZ) [41], and DS7 (Shangtang SenseEarth Archive [40]) datasets. To guarantee CD accuracy, the bitemporal images in each dataset were preprocessed using radiometric correction and coregistration. The details of each dataset are as follows.

(1) DS1: The first dataset contained two RGB 512 × 512 HRRS images characterized by a spatial resolution of 3 m per pixel. The images covered three ground-cover types: farmland I, farmland II, and buildings.
(2) DS2: The second dataset comprised two RGB 512 × 512 HRRS images characterized by a spatial resolution of 3 m per pixel. The images contained four ground-cover types: grassland, open ground, and buildings I and II.
(3) DS3: The third dataset comprised two RGB 512 × 512 HRRS images characterized by a spatial resolution of 3 m per pixel. The images comprised four ground-cover types: buildings, wastelands, rivers, and grasslands.
(4) DS4: To validate the performance of the proposed method in detecting irregular change regions and various change types in medium- and high-resolution imagery, we selected the DS4 dataset, which consisted of images from Tashkurgan Tajik Autonomous County in Xinjiang, China, captured by Gaofen 2 in 2021 and 2022, with a resolution of 1 m and a pixel resolution of 512 × 512.
(5) DS5: To validate the performance of the proposed method on low-resolution imagery, we selected the DS5 dataset, which consisted of two Landsat8 RGB images of Lanzhou New District, Lanzhou City, Gansu Province, captured in 2016 and 2017, with a spatial resolution of 30 m and a pixel resolution of 650 × 650.
(6) DS6: To verify the performance of the method proposed in this study on very-high resolution imagery with significant spectral differences, we selected DS6 from bitemporal remote sensing images of Guangzhou, China, in the CD_Data_GZ dataset. The DS6 dataset contained two remote sensing RGB images with a spatial resolution of

0.55 m and dimensions of 1708 × 1708, and the spectral features of the bitemporal images exhibited noticeable differences.

(7) In order to validate the performance of the proposed method on medium- and high-resolution images, we chose 512 × 512 bitemporal images in the DS7 dataset with a resolution of 3 m.

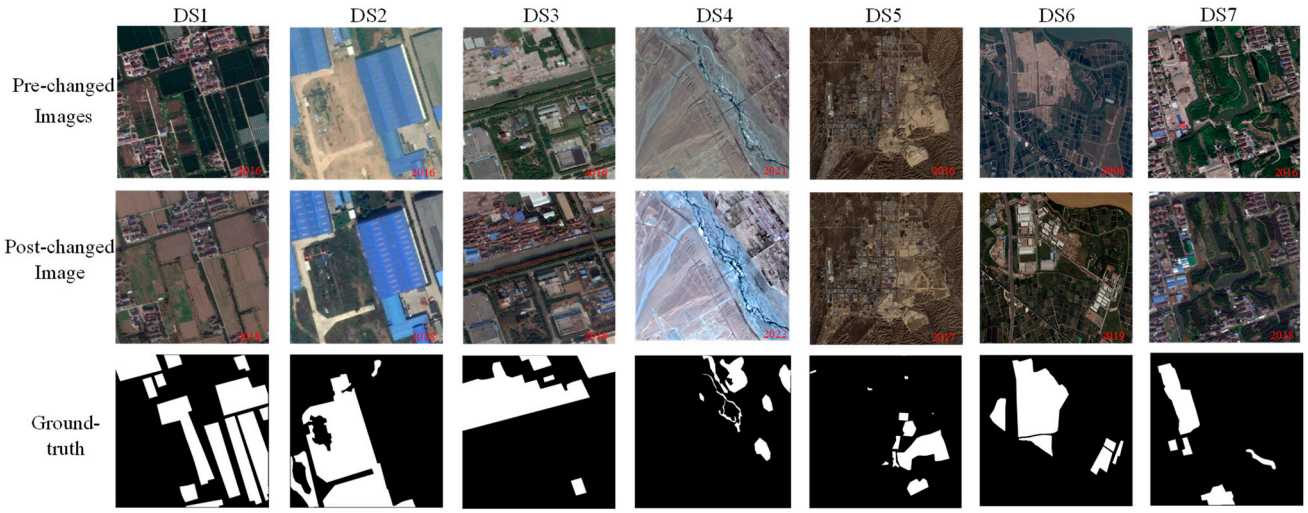


**Figure 5.** Experimental datasets and ground truth.

### *3.2. Criteria for Evaluating Results*

The performance of the proposed unsupervised RVMamba-based image segmentation algorithm was quantitatively assessed and compared using the following three metrics.

(1) The pixel accuracy (PA) is the ratio of the number of correctly classified pixels to the total number of pixels.
(2) The mean pixel accuracy (MPA) is the average of the proportions of corrected classified pixels for each ground label.
(3) The mean intersection over union (mIOU) is the IOU for all ground labels. Here, the IOU indicates the intersection over the union of the true and classified ground labels.

The effectiveness of the CD methods was evaluated based on the following four metrics.

(1) The false alarm (FA) rate measures the percentage of pixels mistakenly determined to be changed pixels.
(2) The misdetection (MD) rate measures the percentage of pixels falsely determined to be unchanged pixels.
(3) The overall accuracy (OA) represents the ratio of the total number of correctly detected changed and unchanged pixels to the total number of pixels.
(4) The Kappa coefficients.

### *3.3. Analysis of the Experimental Results*

#### 3.3.1. Unsupervised Segmentation Accuracy Evaluation

The segmentation accuracies of the unsupervised RVMamba, UNet, and KMeans algorithms were compared based on datasets DS1, DS2, and DS3, respectively. As demonstrated in Figure 6, the unsupervised RVMamba achieved the highest PA, MPA, and mIOU compared with the unsupervised UNet and KMeans. In dataset DS1, the mIOU of the unsupervised RVMamba algorithm was 0.7856, which was 0.0274 and 0.0800 higher than those of the unsupervised UNet and KMeans algorithms, respectively. In dataset DS2, the unsupervised RVMamba algorithm obtained an mIOU value of 0.8726, which was 0.0334 and 0.0876 higher than those of the unsupervised UNet and KMeans algorithms, respectively. In DS3, the unsupervised RVMamba algorithm achieved an mIOU of 0.7963,

which was 0.0405 and 0.2404 higher than those of the UNet and KMeans algorithms, respectively. The superior segmentation accuracy of the proposed RVMamba algorithm must be attributed to the RVMamba structure. The RVMamba-based unsupervised segmentation delivers high-precision results due to the architecture's ability to extract multi-scale spectral and spatial features, maintain linear computational complexity, and establish long-range dependencies.

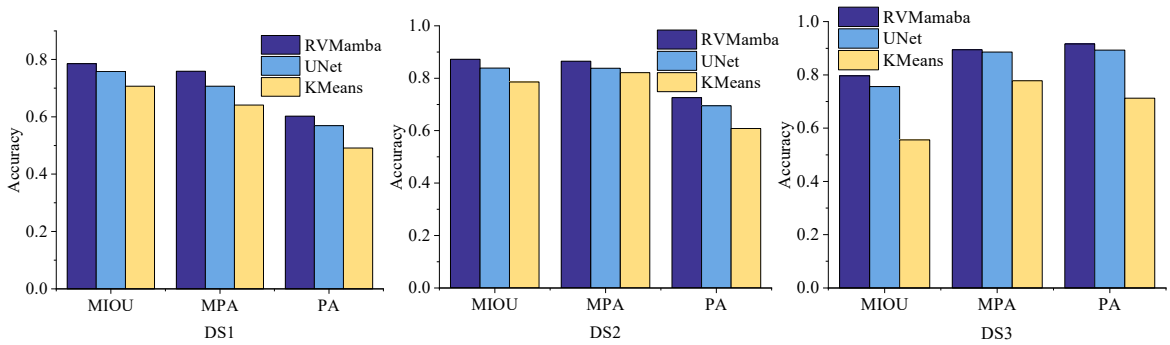


**Figure 6.** Segmentation accuracies of RVMamba, UNet, and KMeans.

### 3.3.2. Contrast Experiment

In this study, the five most advanced CD algorithms were performed to assess the effectiveness of the proposed RFCC from different perspectives. A concise overview of the implemented algorithms is provided in Table 1.

**Table 1.** The Brief descriptions of the implemented comparison algorithms.

| Algorithm | Description |
|---|---|
| ASEA | A change detection algorithm based on adaptive spatial context extraction from VHR remote sensing images [42] |
| PCANet | Group of PCA filters combined into an unsupervised PCANet [43] model, which is used to generate change maps robust to speckle noise |
| KPCAMNet | Based on convolutional kernel PCA and deep Siamese mapping network, KPCAMNet [44] is capable of binary and multiclass change detection |
| DeepCVA | An unsupervised context-sensitive framework based on pre-trained convolution neural network features [27] |
| GMCD | An unsupervised change detection method utilizing graph convolutional networks and metric learning [28] |

#### Performance Analysis on Dataset DS1

The visual results of the five most advanced CD algorithms (listed in Table 1) on dataset DS1 are shown in Figure 7. From the observations in Figure 7, the B2B distance introduced by ASEA successfully enhanced the homogeneity of the change region, yielding improved results. With PCA-based deep learning technologies, PCANet significantly suppressed the FA areas at the expense of significantly enlarged MD areas. KPCAMNet severely overestimated the FA and MD areas, indicating that its nonlinear kernel function could not accommodate this dataset. DeepCVA also produced large FA and MD areas because its pre-trained network parameters were unsuitable for processing this dataset. Due to the graph convolutional network and metric learning, the GMCD yielded satisfactory CD results with fewer FA and MD areas than the above four methods. Compared to the GMCD method, the proposed RFCC method obtained fewer FA and MD areas because of its spatially sensitive CSBN model and highly accurate RVMamba segmentation model.

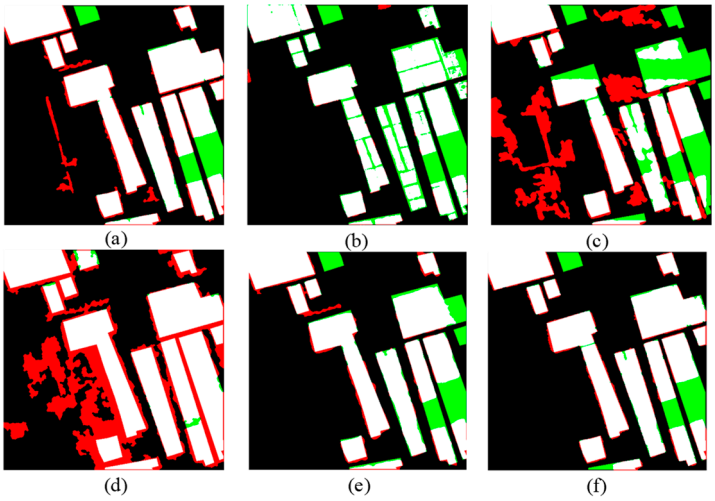


**Figure 7.** Change maps obtained by the different most advanced methods on the dataset DS1. (**a**) ASEA, (**b**) PCANet, (**c**) KPCAMNet, (**d**) DeepCVA, (**e**) GMCD, (**f**) RFCC. (Black is TN, white is TP, red is FA, and green is MD).

As shown in Table 2, the performance of RFCC was the best in terms of OA (93.54%) and Kappa (0.8544). Compared with other unsupervised CD methods, the Kappa values of our method were 0.0392 (over ASEA), 0.0509 (over PCANet), 0.3900 (over KPCAMNet), 0.2624 (over DeepCVA), and 0.0429 (over GMCD). Although our FA and MD rates are not the highest, RFCC can balance the FA and MD areas and achieve the best overall performance out of all the methods.

**Table 2.** Performance comparison of the most advanced CD algorithms (DS1).

| Algorithms | FA | MD | OA | Kappa | Time (s) |
|---|---|---|---|---|---|
| ASEA | 12.71% | 11.35% | 91.69% | 0.8162 | 19.7151 |
| PCANet | **1.79%** | 23.10% | 91.61% | 0.8034 | 3257.9549 |
| KPCAMNet | 36.90% | 32.00% | 75.43% | 0.4643 | **7.3249** |
| DeepCVA | 37.81% | **1.18%** | 79.03% | 0.5919 | 9.7145 |
| GMCD | 7.71% | 17.27% | 91.72% | 0.8114 | 9.2575 |
| **RFCC** | 6.79% | 12.50% | **93.54%** | **0.8544** | 92.5694 |

The data in Figure 7 and Table 2 validate the performance of our method for dataset DS1. Notably, the proposed method outperformed DeepCVA and GMCD. The inferior performance of DeepCVA and GMCD must be attributed to their dependency on the pre-trained network parameters, which may not be suitable for dataset DS1. In contrast, our method performs CD based solely on target images and does not rely on pre-trained parameters derived from different datasets, which enhances its adaptivity. Moreover, PCANet extracts features based on a linear transformation and cannot effectively process nonlinearly correlated image data. Although KPCANet can extract features from nonlinearly correlated image data, its efficacy is heavily influenced by the kernel function used. If the selected kernel function cannot accommodate the spectral–spatial characteristics of the image data, KPCANet may not achieve satisfactory CD results. By contrast, our method is automatically trained based on the data within the target images. It does not make any assumptions regarding the probabilistic distribution of image data, which explains its superior performance over PCANet and KPCANet.

However, the CSBN in RFCC requires the computation of spatial information, which results in a higher runtime for the method. In contrast, the KPCAMNet method uses the

PCA filter as a convolutional filter to capture the representative neighborhood features of each pixel, leading to the longest processing time, as shown in the table. The PCANet, DeepCVA, and GMCD methods, which involve only the forward propagation process, require considerably less time.

Performance Analysis on Dataset DS2

The visual results for the five most advanced CD algorithms (listed in Table 1) on dataset DS2 can be seen in Figure 8. As shown in Figure 8a,b, ASEA and PCANet suppress the MD areas at the cost of overestimated FA areas. Similarly to dataset DS1, KPCAMNet severely overestimated both the FA and MD areas, indicating that it could not accommodate this dataset. DeepCVA produces large FA areas due to its pre-trained network parameters, which are unsuitable for the current dataset. Notably, GMCD yielded CD results inferior to those of PCA-K-means because of its unsuitable pre-trained network parameters. Relative to the other competing methods, the proposed RFCC method obtained considerably fewer FA and MD areas.

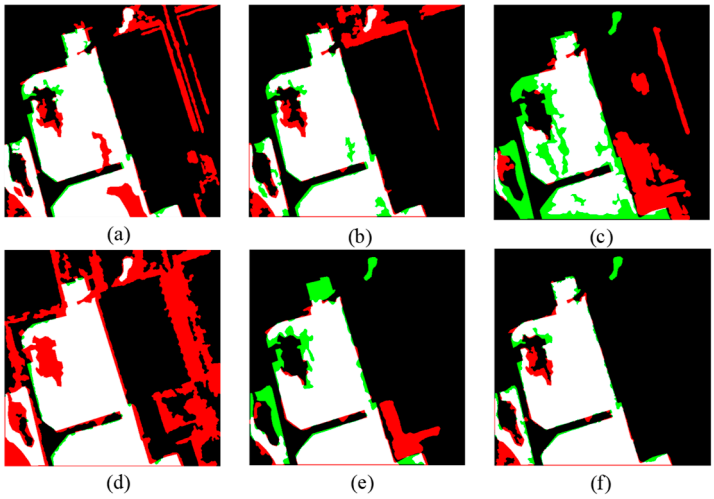


**Figure 8.** Change maps obtained by the different most advanced methods on the dataset DS2. (**a**) ASEA, (**b**) PCANet, (**c**) KPCAMNet, (**d**) DeepCVA, (**e**) GMCD, (**f**) RFCC. (Black is TN, white is TP, red is FA, and green is MD).

As shown in Table 3, RFCC had the lowest FA (9.05%), the third lowest MD (6.13%), and the highest OA (95.42%) and Kappa (0.8911). Relative to the other unsupervised CD methods, the Kappa of our method is 0.1467 (over ASEA), 0.1104 (over PCANet), 0.3735 (over KPCAMNet), 0.3666 (over DeepCVA), and 0.1098 (over GMCD) higher. In addition, our OA values are 6.80%, 5.05%, 15.58%, 19.88%, and 4.44% higher than those of ASEA, PCANet, KPCAMNet, DeepCVA, and GMCD, respectively. In contrast to dataset DS1, our method obtained satisfactory FA and MD rates due to the simple spectral and spatial features of the images in dataset DS2. In addition, the lowest FA of our method was due to the accurate posterior probability vectors estimated using the CSBN method.

The running time of the RFCC algorithm is directly related to the size of the window in the CSBN. Since the feature complexity of DS2 differs from that of DS1, DS2 requires a smaller window, resulting in a significantly shorter processing time.

**Table 3.** Performance comparison of the most advanced CD algorithms (DS2).

| Algorithms | FA | MD | OA | Kappa | Time (s) |
|---|---|---|---|---|---|
| ASEA | 26.66% | 5.12% | 88.62% | 0.7444 | 19.3562 |
| PCANet | 19.39% | 9.88% | 90.37% | 0.7801 | 3249.2589 |
| KPCAMNet | 32.70% | 35.66% | 79.84% | 0.5176 | **7.1965** |
| DeepCVA | 44.36% | **2.07%** | 75.54% | 0.5245 | 9.5269 |
| GMCD | 14.05% | 16.92% | 90.98% | 0.7813 | 9.1469 |
| **RFCC** | **9.05%** | 6.13% | **95.42%** | **0.8911** | 80.3258 |

Performance Analysis on Dataset DS3

The visual results of the five most advanced CD algorithms (listed in Table 1) on dataset DS3 can be seen in Figure 9. As shown in Figure 9a,b, ASEA effectively reduces noise, enabling it to generate small-scale FA. Similarly to datasets DS1 and DS2, PCANet and KPCAMNet severely overestimated the FA and MD areas, indicating that their kernel function was unsuitable for the current dataset. DeepCVA produced large FA areas because of its unsuitable pre-trained network parameters. Notably, GMCD yielded CD results with smaller FA areas. However, it also generated obvious MD areas because of unsuitable pre-trained network parameters. Compared with other competitors, the proposed RFCC method balanced the FA and MD areas well.

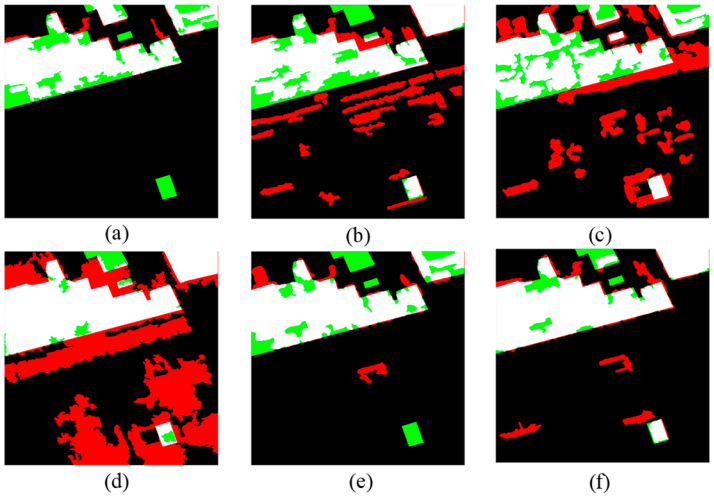


**Figure 9.** Change maps obtained by the different most advanced methods on the dataset DS3. (**a**) ASEA, (**b**) PCANet, (**c**) KPCAMNet, (**d**) DeepCVA, (**e**) GMCD, (**f**) RFCC. (Black is TN, white is TP, red is FA, and green is MD).

As displayed in Table 4, RFCC obtained the highest OA (93.84%) and Kappa (0.8296). Compared with the other unsupervised CD methods, the Kappa values of our method were 0.0518 (over ASEA), 0.2520 (over PCANet), 0.3296 (over KPCAMNet), 0.4757 (over DeepCVA), and 0.0645 (over GMCD) higher. In addition, our OA was 1.16%, 9.75%, 13.99%, 27.04%, and 1.79% higher than those of ASEA, PCANet, KPCAMNet, DeepCVA, and GMCD, respectively. Among the three datasets, our method obtained Kappa and OA values closest to those of ASEA and GMCD on dataset DS3 because of its complex spectral and spatial characteristics.

DS3 has the highest feature complexity, requiring a larger window in the CSBN to calculate the posterior probability, which makes the RFCC algorithm more time-consuming.

**Table 4.** Performance comparison of the most advanced CD algorithms (DS3).

| Algorithms | FA | MD | OA | Kappa | Time (s) |
|---|---|---|---|---|---|
| ASEA | **5.97%** | 26.84% | 92.68% | 0.7778 | 19.1963 |
| PCANet | 36.41% | 26.22% | 84.09% | 0.5776 | 3228.4692 |
| KPCAMNet | 45.16% | 24.75% | 79.85% | 0.5000 | **6.8563** |
| DeepCVA | 59.47% | **8.28%** | 66.80% | 0.3539 | 9.1587 |
| GMCD | 11.35% | 24.53% | 92.05% | 0.7651 | 9.0296 |
| **RFCC** | 14.61% | 11.34% | **93.84%** | **0.8296** | 125.3654 |

#### 3.3.3. Ablation Experiment

We also provide three ablation experiments. A concise overview of the implemented algorithms is provided in Table 5.

**Table 5.** Brief descriptions of the implemented ablation experiment algorithms.

| Algorithm | Description |
|---|---|
| UNet-FCM-CSBN-CVAPS | A posterior probabilistic change detection method that combines the UNet unsupervised segmentation algorithm with CSBN, utilizing the UNet network as a feature extractor. |
| RVMamba-FCM-SBN-CVAPS | A posterior probabilistic change detection method that combines the RVMamba unsupervised segmentation algorithm with SBN, utilizing the RVMamba network as a feature extractor. |
| RVMamba-SVM-CVAPS | A posterior probabilistic change detection method that combines the RVMamba unsupervised segmentation algorithm with SVM, utilizing the RVMamba network as a feature extractor. |

##### Performance Analysis on Dataset DS1

It can be observed in the ablation experiments (Figure 10) that for RVMamba-SVM-CVAPS, the MD rate was significantly reduced at the cost of an increased number of small FA areas (see Figure 10d) because of the low uncertainty of posterior probability vectors estimated by the SVM. Comparatively, RVMamba-FCM-SBN-CVAPS outperformed RVMamba-SVM-CVAPS with fewer FA areas at the modest cost of several enlarged MD areas (Figure 10c). This is because the SVM estimates the posterior probability vector using an optimization technique, whereas the SBN statistically estimates the posterior probability vector. Hence, the SBN is prone to producing posterior probability vectors with higher uncertainty, leading to lower FAs and higher MDs. Finally, RFCC and UNet-FCM-CSBN-CVAPS significantly outperformed RVMamba-FCM-SBN-CVAPS with fewer FA and MD areas because they reasonably reduced the uncertainty of the posterior probability vectors by incorporating pixel-level spatial information through the CSBN (see Figure 10a–c).

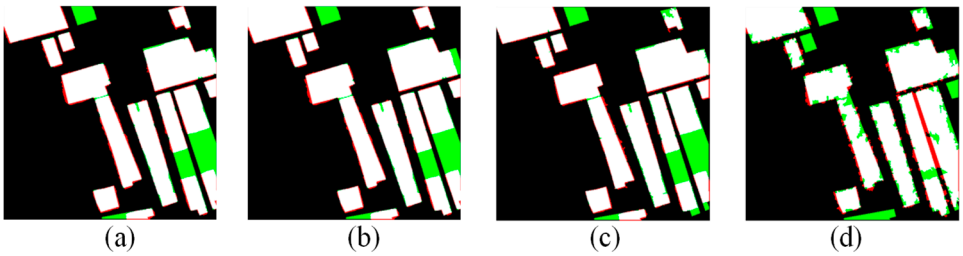


**Figure 10.** Change maps obtained with different algorithms tested on the dataset DS1. (**a**) RFCC, (**b**) UNet-FCM-CSBN-CVAPS, (**c**) RVMamba-FCM-SBN-CVAPS, (**d**) RVMamba-SVM-CVAPS. (Black is TN, white is TP, red is FA, and green is MD).

As shown in Table 6, the proposed RFCC obtained a Kappa value of 0.8544, and the OA was 93.54%. Comparatively, the Kappa improvement achieved by our method was 0.0031 (over UNet-FCM-CSBN-CVAPS), indicating that the superior segmentation accuracy of RVMamba is a positive contributor to the results achieved by the proposed method. Moreover, the Kappa value of the proposed method was 0.0059, which was higher than that of RVMamba-FCM-SBN-CVAPS. This evidence confirms that RVMamba and the CSBN can synergistically enhance the change detection performance. Our method yielded Kappa values that were 0.0314 higher than that of RVMamba-SVM-CVAPS. The significantly higher Kappa of the proposed method suggests that the SVM is not an ideal option for estimating posterior probability vectors.

**Table 6.** Performance comparison of the ablation CD algorithms (DS1).

| Algorithms | Parameters | FA | MD | OA | Kappa |
|---|---|---|---|---|---|
| **RFCC** | 50 Signal class, Samples = 1000, $q$ = 3, Window size = 9 | 6.79% | **12.50%** | **93.54%** | **0.8544** |
| UNet-FCM-CSBN-CVAPS | 50 Signal class, Samples = 1000, $q$ = 3, Window size = 9 | 6.31% | 13.38% | 93.42% | 0. 8512 |
| RVMamba-FCM-SBN-CVAPS | 30 Signal class, $q$ = 3.5, Samples = 1000 | **5.90%** | 14.13% | 93.32% | 0.8484 |
| RVMamba-SVM-CVAPS | cp = 13, gamma = 3, Samples = 2000 | 8.18% | 15.41% | 92.18% | 0.8229 |

Performance Analysis on Dataset DS2

As manifested in the ablation experiment (Figure 11), similarly to dataset DS1, RVMamba-SVM-CVAPS severely overestimated the change areas in dataset DS2 (see Figure 11d). As shown in Figure 11c, compared with RVMamba-SVM-CVAPS, RVMamba-FCM-SBN-CVAPS detected fewer FA areas but more MD areas due to the high uncertainty of the posterior probability vectors estimated by the SBN. Similarly to dataset DS1, the performance of RFCC and UNet-FCM-CSBN-CVAPS significantly exceeded that of RVMamba-FCM-SBN-CVAPS with fewer FA and MD areas due to the inclusion of spatial information by the CSBN (see Figure 11a,b).

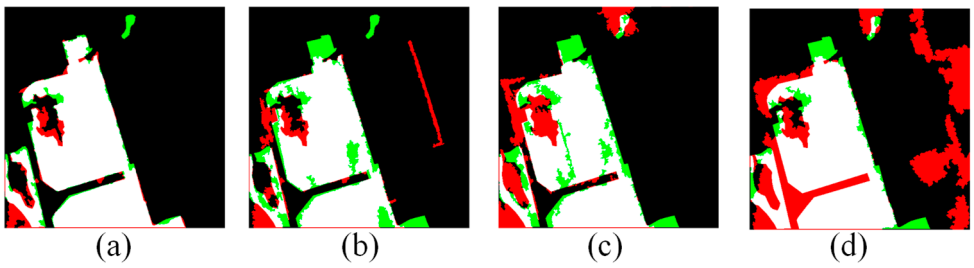


**Figure 11.** Change maps obtained with different algorithms tested on the dataset DS2. (**a**) RFCC, (**b**) UNet-FCM-CSBN-CVAPS, (**c**) RVMamba-FCM-SBN-CVAPS, (**d**) RVMamba-SVM-CVAPS. (Black is TN, white is TP, red is FA, and green is MD).

As shown in Table 7, RFCC achieved the highest Kappa and OA values and maintained a good balance between FA and MD. Similarly to dataset DS1, the Kappa of our method is 0.0193 (over UNet-FCM-CSBN-CVAPS); our method also significantly outperformed RVMamba-FCM-SBN-CVAPS, and its Kappa was 0.2820 higher than that of RVMamba-FCM-SBN-CVAPS. Finally, RVMamba-FCM-SBN-CVAPS yielded the poorest CD performance for OA and Kappa because of the very high FA rates. Numerically, the Kappa of RVMamba-SVM-CVAPS is 0.3311 lower than that of the proposed RFCC.

**Table 7.** Performance comparison of the ablation CD algorithms (DS2).

| Algorithms | Parameters | FA | MD | OA | Kappa |
|---|---|---|---|---|---|
| **RFCC** | 30 Signal class, Samples = 1000, $q$ = 3.5, Window size = 7 | **9.05%** | 6.13% | **95.42%** | **0.8911** |
| UNet-FCM-CSBN-CVAPS | 30 Signal class, Samples = 1000, $q$ = 3.5, Window size = 7 | 9.55% | 7.99% | 94.63% | 0.8718 |
| RVMamba-FCM-SBN-CVAPS | 50 Signal class, $q$ = 4, Samples = 1000, | 32.19% | 18.77% | 82.50% | 0.6091 |
| RVMamba-SVM-CVAPS | Cp = 11, gamma = 15, Samples = 1000, | 42.93% | **0.43%** | 77.71% | 0.5600 |

Performance Analysis on Dataset DS3

As shown in Figure 12, Generally, in the ablation experiments, the UNet-Net-based method caused more false alarms than its RVMamba-based counterpart. These observations are confirmed by the numerical results in Table 8; the Kappa of RFCC is 0.1267 (over UNet-FCM-CSBN-CVAPS).

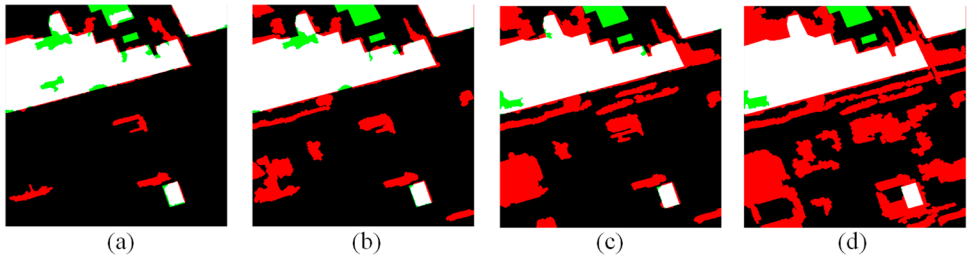


**Figure 12.** Change maps obtained with different algorithms tested on the dataset DS3. (**a**) RFCC, (**b**) UNet-FCM-CSBN-CVAPS, (**c**) RVMamba-FCM-SBN-CVAPS, (**d**) RVMamba-SVM-CVAPS. (Black is TN, white is TP, red is FA, and green is MD).

**Table 8.** Performance comparison of the ablation CD algorithms (DS3).

| Algorithms | Parameters | FA | MD | OA | Kappa |
|---|---|---|---|---|---|
| **RFCC** | 10 Signal class, Samples = 1000, $q$ = 4, Window size = 11 | 14.61% | 11.34% | **93.84%** | **0.8296** |
| UNet-FCM-CSBN-CVAPS | 10 Signal class, Samples = 1000, $q$ = 4, Window size = 5 | 31.01% | 10.01% | 88.27% | 0.7029 |
| RVMamba-FCM-SBN-CVAPS | 30 Signal class, $q$ = 3.5, Samples = 1000, | 43.26% | 8.41% | 81.81% | 0.5802 |
| RVMamba-SVM-CVAPS | cp = 1, gamma = 15, Samples = 2000, | 58.17% | 8.30% | 68.44% | 0.3751 |

3.3.4. Entropy Analysis

In this section, the performance differences among RVMamba-SVM-CVAPS, RVMamba-FCM-SBN-CVAPS, and RFCC are explained based on the posterior probability vector. As an optimization algorithm, support vector machines (SVMs) tend to produce low-uncertainty (low entropy) posterior probability vectors, where certain ground-cover labels have significantly higher posterior probabilities than others, leading to high false alarm rates in CD results. As a statistical algorithm, the simple Bayesian network (SBN) produces a posteriori probability vectors with high uncertainty (high entropy), where the differences in posterior probabilities between various ground-cover labels are much smaller than those based on SVMs. However, the SBN's high uncertainty can lead to the

underestimation of change regions. The CSBN method addresses this issue by reasonably incorporating spatial information to reduce the uncertainty.

As illustrated in Figure 13, posterior probability vectors with low uncertainty (low entropy) led to significantly higher change magnitudes and FA rates than those with high uncertainty (high entropy). As shown in Table 9, because the CSBN and SBN produced a higher average AE than the SVM, the FA rates of RFCC and RVMamba-FCM-SBN-CVAPS were 6.79% and 5.90% (DS1), 9.05% and 32.19% (DS2), and 14.61% and 43.26% (DS3), respectively, which were lower than those of RVMamba-SVM-CVAPS. However, for dataset DS2, although the CSBN produced a slightly lower AE than the SBN, the corresponding FA rate of RFCC was 12.64% lower than that of RVMamba-FCM-SBN-CVAPS, indicating that the CSBN can reasonably consider spatial information and may reduce the uncertainty and FA rate simultaneously.

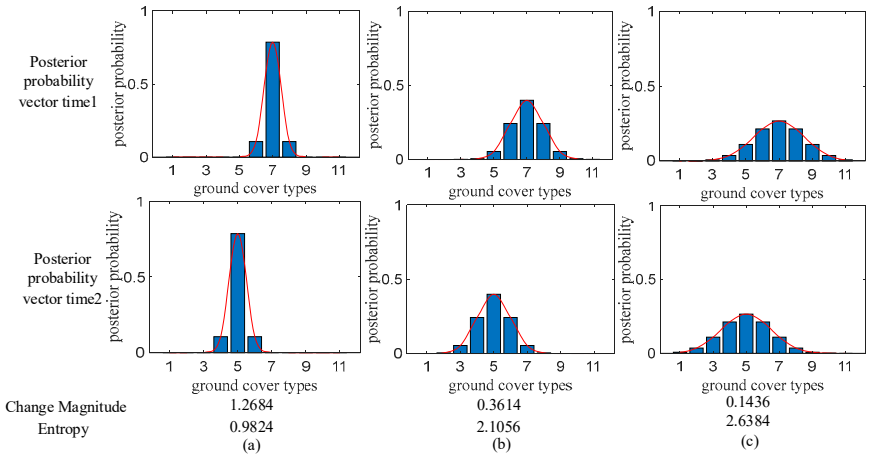


**Figure 13.** Evaluation of change magnitude and entropy in bitemporal simulated posterior probability vectors. (**a**): Low uncertainty. (**b**): Appropriate reduction in certainty. (**c**): High uncertainty.

**Table 9.** Average entropy (AE) of change detection approach.

| Algorithms | Approach | AE | FA | MD |
|---|---|---|---|---|
| DS1 | RFCC | 1.2699 | 6.79% | 12.50% |
| | RVMamba-FCM-SBN-CVAPS | 1.3841 | 5.90% | 14.13% |
| | RVMamba-SVM-CVAPS | 0.9072 | **8.18%** | 15.41% |
| DS2 | RFCC | 1.6215 | 9.05% | 6.13% |
| | RVMamba-FCM-SBN-CVAPS | 1.7755 | 32.19% | 18.77% |
| | RVMamba-SVM-CVAPS | 0.3876 | **42.93%** | 0.43% |
| DS3 | RFCC | 1.7049 | 14.61% | 11.34% |
| | RVMamba-FCM-SBN-CVAPS | 1.7600 | 43.26% | 8.41% |
| | RVMamba-SVM-CVAPS | 0.1019 | **58.17%** | 8.30% |

### 3.3.5. Parameter Sensitivity Analysis

#### Fuzzy Degree $q$ and Signal Class Number

This subsection discusses the performance of the proposed methods (RFCC and RVMamba-FCM-SBN-CVAPS) with different fuzzy degrees ($q$ = 1.5, 2, 2.5, 3, 3.5, and 4) and different signal class numbers (10, 30, and 50). The SBN and CSBN were trained using 1000 training samples per label, which were randomly selected from the RVMamba-classified images. The outcomes of the experiment are presented in Figure 14.

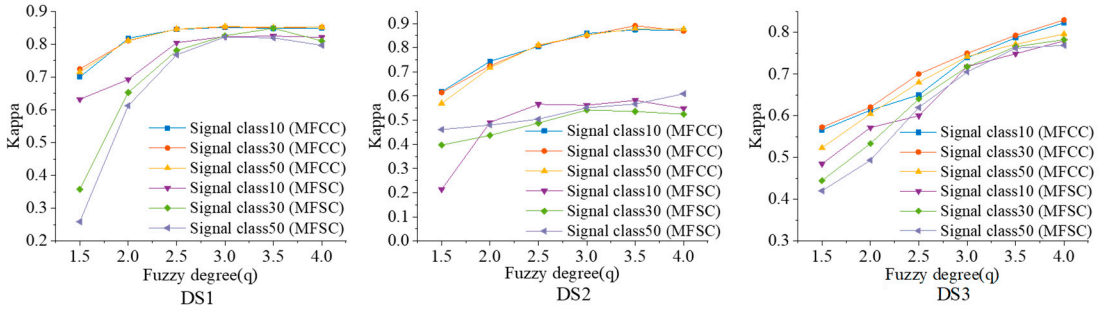


**Figure 14.** Effect of fuzziness q on algorithm results.

On all three datasets, when considering 10, 30, and 50 signal classes, both the RFCC and RVMamba-FCM-SBN-CVAPS approaches consistently achieved the lowest Kappa values when $q$ = 1.5. These results were caused by the ineffective mixed-pixel decomposition of the FCM at a low fuzzy degree. Moreover, a low fuzzy degree leads to a high accumulated clustering error and reduced uncertainty in the FCM, which decreases the uncertainty of the posterior probability vectors estimated by the SBN and CSBN. The performance of RVMamba-FCM-SBN-CVAPS and RFCC is degraded because the low uncertainty of the posterior probability vectors induces a high cumulative classification error. It can also be observed in Figure 14 that the Kappa value of RFCC is consistently higher than that of RVMamba-FCM-SBN-CVAPS for all three datasets because of the incorporation of pixel-level spatial information by the CSBN. In addition, the RFCC method achieved a more stable CD performance than RVMamba-FCM-SBN-CVAPS, as shown in Figure 14, whereas the CSBN was less sensitive to parameter variations (fuzzy degree and signal class number) due to its inclusion of spatial information.

Window Size

As illustrated in Figure 15, varying window sizes influence the effectiveness of the RFCC method. For the DS2 dataset, the simpler spatial features in the remotely sensed images lead to a smaller window size, yielding the highest Kappa coefficient. Conversely, in the DS1 and DS3 datasets, the images exhibit varying spatial complexities, necessitating larger window sizes to achieve the highest Kappa coefficients. These observations indicate that the optimal window size for the RFCC method is dependent on the spatial complexity of the remote sensing images; the more complex the feature distribution within the remote sensing image, the larger the window required to capture spatial information and enhance accuracy.

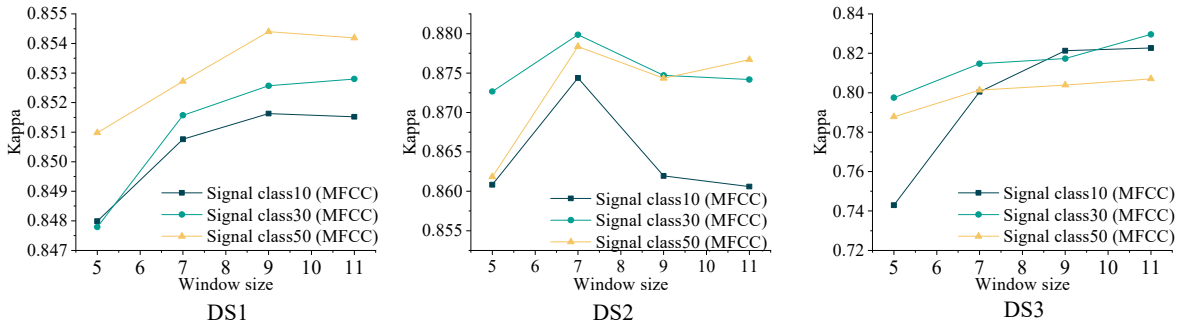


**Figure 15.** Effect of window size on algorithm results.

The Number of Segmentation Labels

In this study, an unsupervised RVMamba segmentation algorithm was employed to classify RS images into a predetermined number of ground labels, which affected the performance of the CSBN and, subsequently, UCVAPS CD. It can be observed in Figure 16

that on all three datasets, RFCC produced a slightly varied Kappa with an increasing number of ground labels because the CSBN is robust against cumulative classification errors induced by a large number of ground labels. Meanwhile, the computation time increases because of the elevated computation loads of the CSBN, caused by the increasing ground label number.

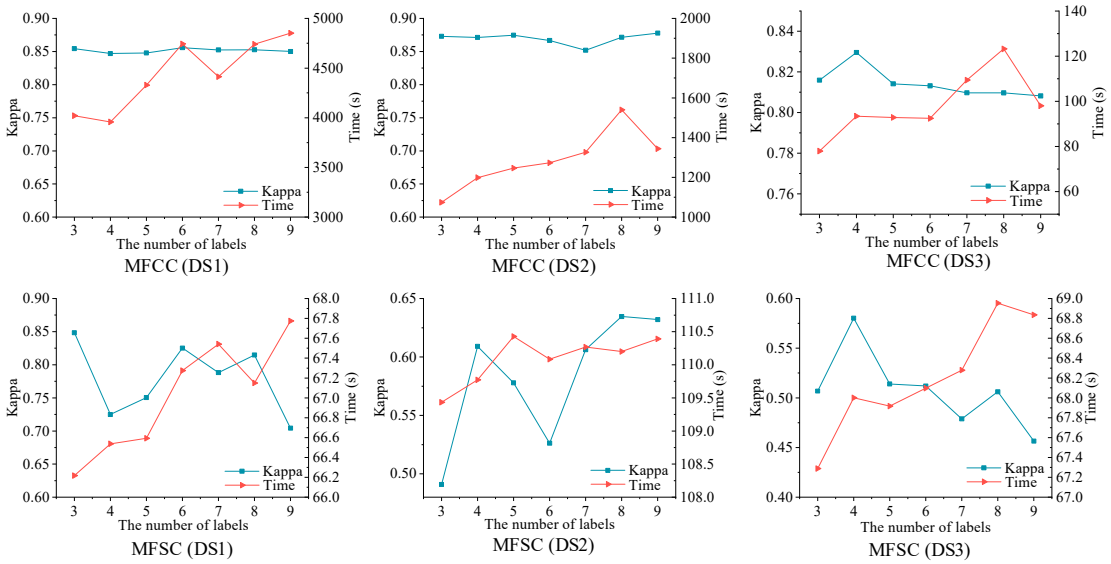


**Figure 16.** Effect of the number of segmentation labels on Kappa and algorithm timeliness.

### 3.3.6. Adaptive Experiments

In this part of the study, we used three datasets to evaluate the adaptivity of the proposed method. As shown in Figure 17 and Table 10, our proposed method delivered the highest performance with datasets DS4, DS5, and DS6, confirming its high adaptivity.

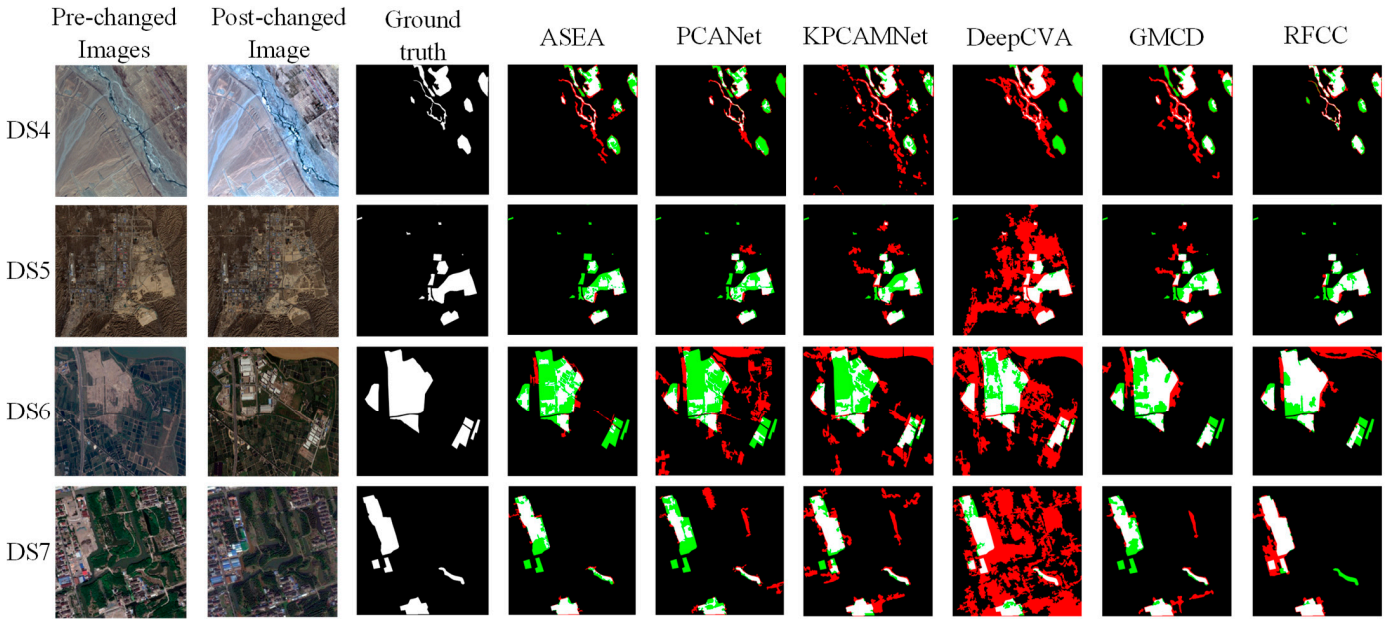


**Figure 17.** Change maps generated by different techniques in the adaptive experiments. (Black is TN, white is TP, red is FA, and green is MD).

**Table 10.** Performance comparison of the most advanced CD algorithms.

| Algorithms | ASEA | PCANet | KPCAMNet | DeepCVA | GMCD | RFCC |
|---|---|---|---|---|---|---|
| DS4 | 0.6627 | 0.5787 | 0.5384 | 0.4185 | 0.5786 | **0.8016** |
| DS5 | 0.6782 | 0.6745 | 0.6167 | 0.3246 | 0.6632 | **0.7931** |
| DS6 | 0.4340 | 0.1923 | 0.3229 | 0.3245 | 0.6509 | **0.6692** |
| DS7 | 0.7393 | 0.5022 | 0.6067 | 0.1417 | 0.6489 | **0.7583** |

As shown in Figure 17, on DS4, ASEA and PCANet produced relatively smaller FA and MD areas, yielding overall good change detection (CD) results. In contrast, KPCAMNet generated a larger FA area. While DeepCVA and GMCD achieved smaller FA and MD areas compared with RFCC, their performance was still lower than that of RFCC. This performance difference can be attributed to the fact that the pre-trained parameters of DeepCVA and GMCD were not well suited for the target images, confirming that the RFCC method has superior generalization performance and is better at detecting irregular fine-scale changes than DeepCVA and GMCD.

The results on DS5 shown in Figure 17 indicate that ASEA and PCANet detected fewer FA areas at the cost of more MD areas, while KPCAMNet detected more FA areas. DeepCVA produces larger FA areas due to its unsuitable pre-trained network parameters. Comparatively, GMCD obtains fewer FA areas due to the introduction of graph convolution and metric learning. Compared with its competitors, the RFCC method had well-balanced FA and MD rates due to the spatial sensitivity of the CSBN model, and it achieved the highest Kappa.

The results on DS6 shown in Figure 17 indicate that ASEA underestimates the change area. KPCAMNet and PCANet severely overestimate the FA and MD areas, indicating that the method is inadequate for processing large-scale remote sensing imagery. DeepCVA also produces large FA and MD areas. Notably, GMCD produces CD results with the smallest FA area. However, it also produces significant MD areas due to inappropriate pre-trained network parameters. Compared with the other competing methods, the proposed RFCC method achieved well-balanced FA and MD areas and obtained the highest Kappa value.

The results on DS7 shown in Figure 17 indicate that ASEA achieves good CD results with few FA areas at the cost of large MD areas. Comparatively, PCANet and KPCAMNet predicted more FA areas. Similarly to dataset DS4, DeepCVA overproduced large FA areas, indicating that its pre-trained parameters cannot accommodate this dataset. Similarly, GMCD also produces CD results that are inferior to those of PCANet because of its unsuitable pre-trained network parameters. In contrast, the RFCC method produced the best Kappa with well-balanced FA and MD rates.

#### 3.3.7. Change Detection with Unsupervised Segmentation

RVMamba is an unsupervised segmentation model that provides unsupervised training samples of pseudo-ground-cover labels for the CSBN. While the pseudo-ground-cover labels generated by RVMamba do not necessarily have one-to-one correspondences with real ground-cover labels in remote sensing images, they closely align with real ground-cover labels because unsupervised RVMamba can generate segmentation results that are similar to manually segmented results. Therefore, the pseudo-ground-cover labels are suitable for the unsupervised training of the CSBN.

If we simply compare two unsupervised segmentation results, the change detection performance will be adversely affected by the cumulative segmentation error. Because our UCVAPS approach can transfer the unsupervised segmentation results into posterior probability vectors, comparing the posterior probability vectors can significantly reduce the cumulative segmentation error, which considerably enhances the change detection performance [30]. The CSBN model effectively integrates spatial information at the pixel level to accurately predict posterior probability vectors, making it well suited for high-resolution

remote sensing imagery. The proposed UCVAPS framework generates a uniformly scaled change-magnitude map that is ideal for binarization with a global threshold. To prove the above statements, we conducted change detection experiments based on a direct comparison of unsupervised segmentation results. The experimental results are shown in the following.

As illustrated in the Figure 18, the change detection results derived from a direct comparison of unsupervised segmentation exhibit a high false alarm rate, resulting from cumulative segmentation errors.

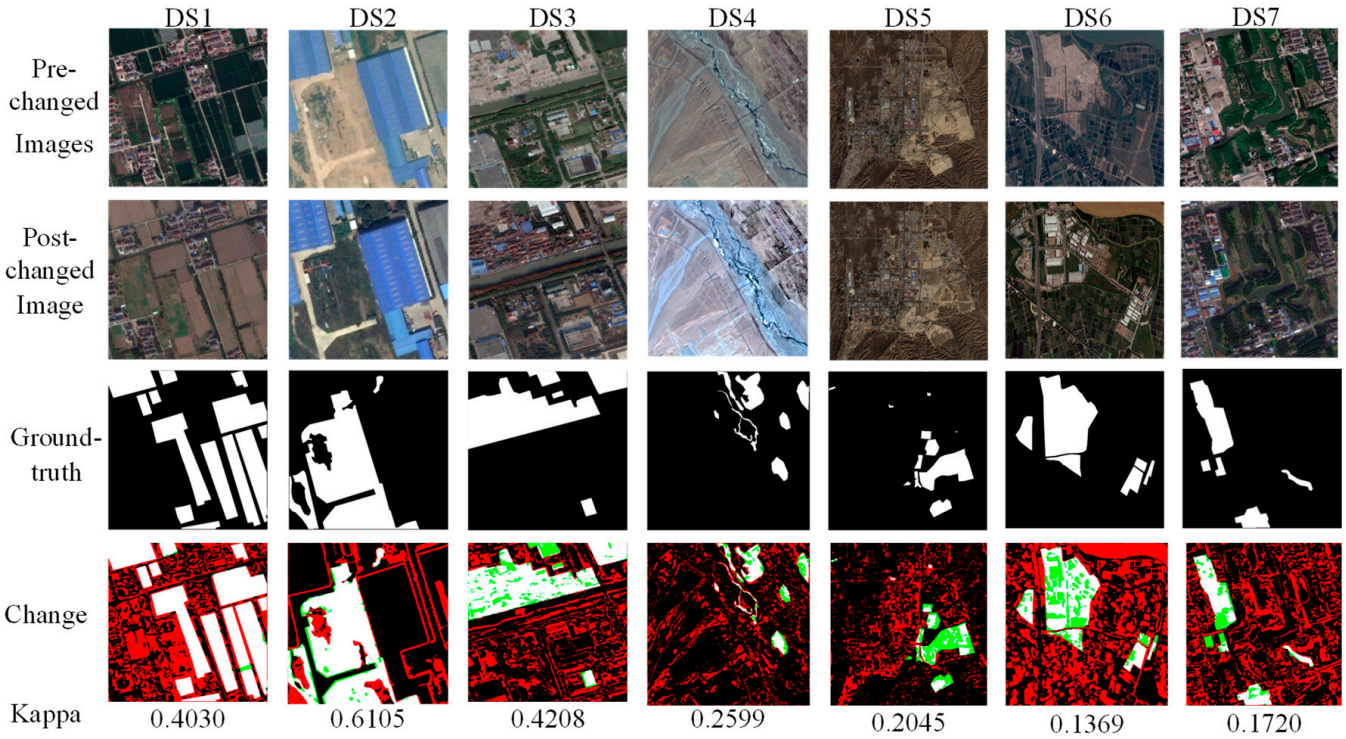


**Figure 18.** Change detection with unsupervised segmentation.

To clearly show the value of the method, we implemented two simple image segmentation methods, K-means and Fuzzy C-means, combined with our proposed RVMamba method to accomplish three change detection experiments based on a direct comparison of unsupervised segmentation results. The direct comparison of unsupervised segmentation results to achieve change detection relies heavily on the accuracy of the segmentation results, and the accumulated segmentation errors propagate into the change detection results. As shown in the Figure 19, the effect of the accumulated segmentation error leads to a large number of misdetected regions in the change detection results of all three methods. However, our proposed unsupervised segmentation algorithm for RVMamba achieves a higher Kappa due to its higher segmentation accuracy than that of K-means and Fuzzy C-means.

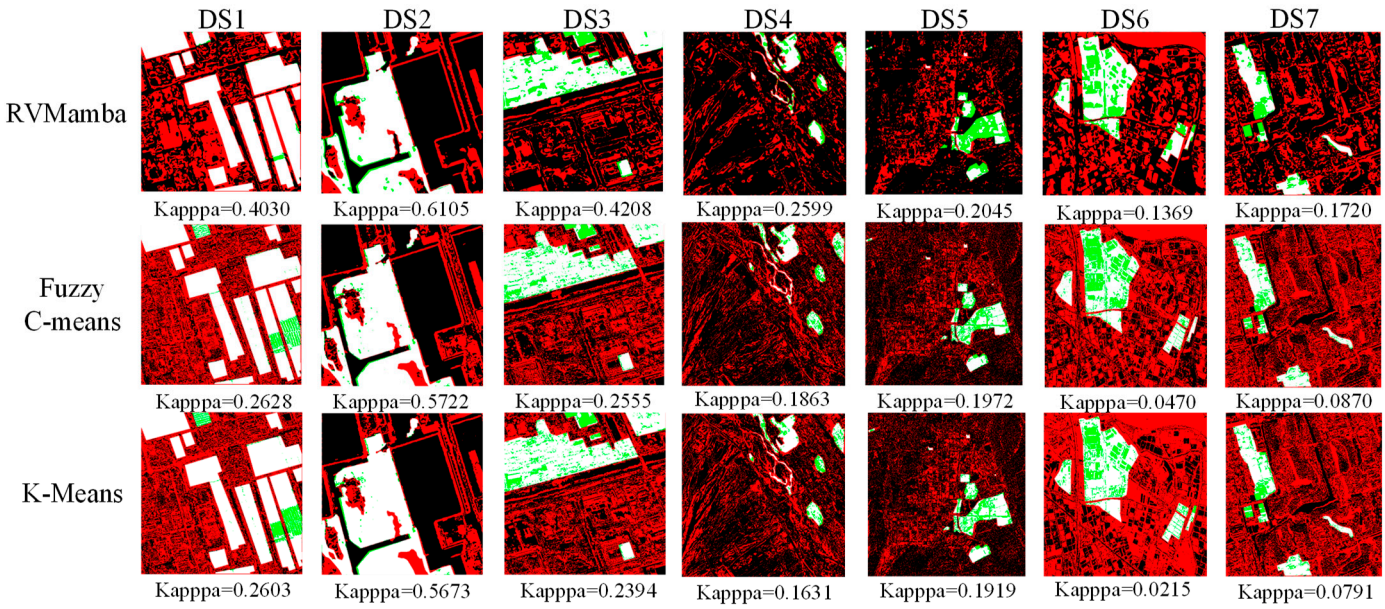


**Figure 19.** Change detection based on RVMamba, K-means, and Fuzzy C-means unsupervised segmentation.

#### 3.3.8. Statistical Tests

We used *t*-tests to assess the significance of the performance differences between RFCC and the other competitive methods. The *t*-test results for Kappa and OA were calculated for the six experimental datasets in this study and are presented in Table 11. According to the *t*-test analysis, a *p*-value of <0.05 indicates a significant difference, while a *p*-value of <0.01 indicates a highly significant difference. As shown in Table 11, the *t*-test results for Kappa indicate that RFCC differs significantly from ASEA, PCANet, and GMCD ($p < 0.05$), while RFCC differs very significantly from KPCAMNet and DeepCVA ($p < 0.01$). Similarly, for the OA results presented in Table 11, the *t*-test values for RFCC with ASEA, PCANet, and GMCD are all less than 0.05, showing significant differences, while the *t*-test values for RFCC with KPCAMNet and DeepCVA are less than 0.01, showing highly significant differences.

**Table 11.** Statistical comparison of *t*-test results for Kappa and OA between RFCC and the comparison methods.

| | ASEA | PCANet | KPCAMNet | DeepCVA | GMCD |
|---|---|---|---|---|---|
| Kappa | 0.0449 | 0.0242 | 0.0001 | 0.0072 | 0.0319 |
| OA | 0.0365 | 0.0403 | 0.0005 | 0.0051 | 0.0155 |

Therefore, we conclude that RFCC exhibits a significant performance difference compared with the other methods.

#### 3.3.9. Outliers and Noise

The RVMamba model may still be influenced by extreme or high-intensity noise. When an image contains substantial noise with a non-uniform spatial distribution, the model's processing capability may be insufficient. Furthermore, the global modeling approach could suppress important local details, reducing the model's sensitivity to local noise and outliers. Therefore, future work should focus on improving the model's robustness and adaptability to extreme noise and outliers.

In addition, the use of discretization may overlook the dynamic properties of continuous systems, particularly in remote sensing image processing, where spatiotemporal

variations are highly complex. The proposed method faces three main issues: neglecting nonlinear dynamics, using excessively large time scales, and accumulating errors. Therefore, we plan to improve the RVMamba model in the future. Specifically, to improve the model's accuracy in capturing temporal and spatial changes in remote sensing images, we can explore higher-order Taylor expansions or other more precise discretization methods beyond the simple first-order Taylor expansion. Additionally, we could design an adaptive time-scale parameter $\Delta$ that dynamically adjusts the discretization step size based on the characteristics and dynamic changes in the image. Specifically, smaller time steps could be applied in areas with significant changes to capture finer details.

## 4. Conclusions and Prospects

Supervised deep learning CD methods often require highly related training data to achieve satisfactory results. Although some unsupervised deep learning CD methods do not require training data, they rely on pre-trained network parameters to attain optimal CD performance. The observed CD performance could not be achieved if the pre-trained network parameters were unsuitable for the target CD images. To address these problems, this study combined an RVMamba-based unsupervised segmentation algorithm with the CSBN model to create a novel RFCC method. The proposed method obtained superior CD results due to the following four advantages: first, the training samples were automatically supplied to the CSBN using an RVMamba-based unsupervised segmentation algorithm. Thus, highly related training data and pre-trained network parameters are not mandatory. Second, the FCM algorithm decomposes mixed pixels into different signal classes, thereby alleviating cumulative clustering errors. Third, the CSBN model integrates pixel-level spatial information into the estimation of the posterior probability vectors. Finally, the UCVAPS framework can generate a uniformly scaled change-magnitude map that is suitable for binarization using a global threshold. The experimental results suggest that the proposed method attains high CD accuracy with low FA and MD rates, is superior to most advanced deep learning CD methods, and has high practicability and application potential.

Although the proposed model has relatively high computational costs, its performance is better than that of competitive state-of-the-art approaches, as shown in our experimental results. Moreover, we are developing parallelized computational methods to boost its change detection speed, which will be published in the future.

**Author Contributions:** All the authors have contributed substantially to the manuscript. J.S. and S.Y. proposed the methodology. Y.L. and X.L. performed the experiments and software. J.S. wrote the paper. J.S. and S.Y. analyzed the data. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work is co-funded by National Key R&D Program of China (Grant No. 2022YFB3903604), Gansu Province Basic Innovation Group Project (Grant No. 24JRRA220), and the Key Research and Development Project of Lanzhou Jiao Tong University (Grant No. LZJTU-ZDYF2301).

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Acknowledgments:** The authors are grateful to the editor and anonymous reviewers for their helpful and valuable suggestions.

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

## References

1. Li, X.; Yan, H.; Xie, W.; Kang, L.; Tian, Y. An improved pulse-coupled neural network model for Pansharpening. *Sensors* **2020**, *20*, 2764. [CrossRef] [PubMed]
2. Qin, H.; Wang, J.; Mao, X.; Zhao, Z.; Gao, X.; Lu, W. An improved faster R-CNN method for landslide detection in remote sensing images. *J. Geovisualization Spat. Anal.* **2023**, *8*, 2. [CrossRef]
3. Xu, X.; Li, X.; Li, Y.; Kang, L.; Ge, J. A novel adaptively optimized PCNN model for hyperspectral image sharpening. *Remote Sens.* **2023**, *15*, 4205. [CrossRef]

4. Jiang, X.; Huang, B.; Zhao, Y. Spatiotemporal image fusion with spectrally preserved Pre-Prediction: Tackling complex Land-Cover changes. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5406114. [CrossRef]
5. Gao, B.; He, Y.; Chen, X.; Zheng, X.; Zhang, L.; Zhang, Q.; Lu, J. Landslide risk evaluation in shenzhen based on stacking ensemble learning and InSAR. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2023**, *16*, 1–18. [CrossRef]
6. Yousif, O.; Ban, Y. Improving urban change detection from multitemporal SAR images using PCA-NLM. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2032–2041. [CrossRef]
7. Castellana, L.; D'Addabbo, A.; Pasquariello, G. A composed supervised/unsupervised approach to improve change detection from remote sensing. *Pattern Recognit. Lett.* **2007**, *28*, 405–413. [CrossRef]
8. Xu, X.; Li, J.; Chen, Z. TCIANet: Transformer-Based context information aggregation network for remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2023**, *16*, 1951–1971. [CrossRef]
9. Zuo, X.; Jin, F.; Ding, L.; Wang, S.; Lin, Y.; Liu, B. Multitask siamese network guided by enhanced change information for semantic change detection in bitemporal remote sensing images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2025**, *18*, 61–77. [CrossRef]
10. Li, Z.; Cao, S.; Deng, J.; Wu, F.; Wang, R.; Luo, J.; Peng, Z. STADE-CDNet: Spatial–Temporal attention with difference enhancement-Based network for remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5611617. [CrossRef]
11. Chen, S.; Su, X.; Zheng, L.; Yuan, Q. Statistic ratio attention-guided siamese U-Net for SAR image semantic change detection. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4004405. [CrossRef]
12. Vinholi, G.J.; Silva, D.; Machado, R.; Pettersson, M.I. CNN-Based change detection algorithm for wavelength-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4003005. [CrossRef]
13. Tan, X.; Xiao, Z.; Wan, Q.; Shao, W. Scale sensitive neural network for road segmentation in high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 533–537. [CrossRef]
14. Zhang, X.; Fan, R.; Ma, L.; Liao, X.; Chen, X. Change detection in very high-resolution images based on ensemble CNNs. *Int. J. Remote Sens.* **2020**, *41*, 4755–4777. [CrossRef]
15. Mohammadian, A.; Ghaderi, F. SiamixFormer: A Fully-Transformer siamese network with temporal fusion for accurate building detection and change detection in bi-temporal remote sensing images. *Int. J. Remote Sens.* **2023**, *44*, 3660–3678. [CrossRef]
16. Pan, Z.; Xu, J.; Guo, Y.; Hu, Y.; Wang, G. Deep learning segmentation and classification for urban village using a worldview satellite image based on U-Net. *Remote Sens.* **2020**, *12*, 1574. [CrossRef]
17. Sun, S.; Mu, L.; Wang, L.; Liu, P. L-UNet: An LSTM network for remote sensing image change detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8004505. [CrossRef]
18. Shen, Q.; Huang, J.; Wang, M.; Tao, S.; Yang, R.; Zhang, X. Semantic feature-constrained multitask siamese network for building change detection in high-spatial-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *189*, 78–94. [CrossRef]
19. Kumar, S.; Kumar, A.; Lee, D.-G. RSSGLT: Remote Sensing Image Segmentation Network Based on Global–Local Transformer. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 8000305. [CrossRef]
20. Zhang, C.; Wang, L.; Cheng, S.; Li, Y. SwinSUNet: Pure Transformer Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5224713. [CrossRef]
21. Jiang, J.; Xiang, J.; Yan, E.; Song, Y.; Mo, D. Forest-CD: Forest change detection network based on VHR images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2506005. [CrossRef]
22. Lv, P.; Li, M.; Zhong, Y. A Semi-Supervised pyramid Cross-Temporal attention transformer for change detection in high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 2503305. [CrossRef]
23. Chen, M.; Jiang, W. A hierarchical local-global-aware transformer with scratch learning capabilities for change detection. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 6000905. [CrossRef]
24. Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. *arXiv* **2021**, arXiv:2111.00396.
25. Chen, H.; Song, J.; Han, C.; Xia, J.; Yokoya, N. ChangeMamba: Remote sensing change detection with spatiotemporal state space model. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4409720. [CrossRef]
26. Khelifi, L.; Mignotte, M. Deep learning for change detection in remote sensing images: Comprehensive review and Meta-Analysis. *IEEE Access* **2020**, *8*, 126385–126400. [CrossRef]
27. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3677–3693. [CrossRef]
28. Tang, X.; Zhang, H.; Mou, L.; Liu, F.; Zhang, X.; Zhu, X.; Jiao, L. An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5609715. [CrossRef]
29. Yuan, M.; Xin, Z.; Liao, G.; Huang, P.; Li, Y. Change detection in SAR image based on weighted difference image generation and optimized random forest. *IET Image Process.* **2024**, *18*, 2754–2773. [CrossRef]
30. Chen, J.; Chen, X.; Cui, X.; Chen, J. Change vector analysis in posterior probability space: A new method for land cover change detection. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 317–321. [CrossRef]
31. Yang, Y.; Li, Y.; Song, J.; Wang, Z. A comparative study of remote sensing image change detection based on bayesian networks and different spatial fuzzy C-Means clustering. In Proceedings of the 9th International Conference on Geology Resources Management and Sustainable Development (ICGRMSD 2021), Beijing, China, 19 December 2021.

32. Li, Y.; Li, X.; Song, J.; Wang, Z.; He, Y.; Yang, S. Remote-Sensing-Based change detection using change vector analysis in posterior probability space: A Context-Sensitive bayesian network approach. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2023**, *16*, 3198–3217. [CrossRef]
33. Liu, X.; Xu, Q.; Ma, J.; Jin, H.; Zhang, Y. MsLRR: A unified multiscale low-rank representation for image segmentation. *IEEE Trans. Image Process.* **2014**, *23*, 2159–2167. [CrossRef]
34. Kim, W.; Kanezaki, A.; Tanaka, M. Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Trans. Image Process.* **2020**, *29*, 8055–8068. [CrossRef]
35. Ou, X.; Liu, L.; Tu, B.; Zhang, G.; Xu, Z. A CNN framework with slow-fast band selection and feature fusion grouping for hyperspectral image change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5524716. [CrossRef]
36. Ma, X.; Zhang, X.; Pun, M.-O. RS3Mamba: Visual state space model for remote sensing image semantic segmentation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6011405. [CrossRef]
37. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752.
38. Liu, Y. VMamba: Visual state space model. *arXiv* **2024**, arXiv:2401.10166.
39. Li, Y.; Bretschneider, T.R. Semantic-Sensitive satellite image retrieval. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 853–860. [CrossRef]
40. Yang, K.; Xia, G.S.; Liu, Z.; Du, B.; Yang, W.; Pelillo, M.; Zhang, L. Asymmetric siamese networks for semantic change detection in aerial images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5609818. [CrossRef]
41. Peng, D.; Bruzzone, L.; Zhang, Y.; Guan, H.; Ding, H.; Huang, X. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5891–5906. [CrossRef]
42. Kondmann, L.; Toker, A.; Saha, S.; Schölkopf, B.; Leal-Taixé, L.; Zhu, X.X. Spatial context awareness for unsupervised change detection in optical satellite images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5614615. [CrossRef]
43. Gao, F.; Dong, J.; Li, B.; Xu, Q. Automatic change detection in synthetic aperture radar images based on PCANet. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1792–1796. [CrossRef]
44. Wu, C.; Chen, H.; Du, B.; Zhang, L. Unsupervised change detection in multitemporal VHR images based on deep kernel PCA convolutional mapping network. *IEEE Trans. Cybern.* **2022**, *52*, 12084–12098. [CrossRef]