



Article

SSA-LHCD: A Singular Spectrum Analysis-Driven Lightweight Network with 2-D Self-Attention for Hyperspectral Change Detection

Yinhe Li ¹, Jinchang Ren ^{1,*} , Yijun Yan ², Genyun Sun ³ and Ping Ma ¹

¹ National Subsea Centre, Robert Gordon University, Aberdeen AB21 0BH, UK; y.li24@rgu.ac.uk (Y.L.); p.ma2@rgu.ac.uk (P.M.)

² School of Science and Engineering, University of Dundee, Dundee DD1 4HN, UK; yijun.yan@ieee.org

³ College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China; genyunsun@163.com

* Correspondence: jinchang.ren@ieee.org

Abstract: As an emerging research hotspot in contemporary remote sensing, hyperspectral change detection (HCD) has attracted increasing attention in remote sensing Earth observation, covering land mapping changes and anomaly detection. This is primarily attributable to the unique capacity of hyperspectral imagery (HSI) to amalgamate both the spectral and spatial information in the scene, facilitating a more exhaustive analysis and change detection on the Earth's surface, proving to be successful across diverse domains, such as disaster monitoring and geological surveys. Although numerous HCD algorithms have been developed, most of them face three major challenges: (i) susceptibility to inherent data noise, (ii) inconsistent accuracy of detection, especially when dealing with multi-scale changes, and (iii) extensive hyperparameters and high computational costs. As such, we propose a singular spectrum analysis-driven-lightweight network for HCD, where three crucial components are incorporated to tackle these challenges. Firstly, singular spectrum analysis (SSA) is applied to alleviate the effect of noise. Next, a 2-D self-attention-based spatial-spectral feature-extraction module is employed to effectively handle multi-scale changes. Finally, a residual block-based module is designed to effectively extract the spectral features for efficiency. Comprehensive experiments on three publicly available datasets have fully validated the superiority of the proposed SSA-LHCD model over eight state-of-the-art HCD approaches, including four deep learning models.

Keywords: hyperspectral imagery (HSI); hyperspectral change detection (HCD); deep learning; singular spectral analysis (SSA); residual block



Citation: Li, Y.; Ren, J.; Yan, Y.; Sun, G.; Ma, P. SSA-LHCD: A Singular Spectrum Analysis-Driven Lightweight Network with 2-D Self-Attention for Hyperspectral Change Detection. *Remote Sens.* **2024**, *16*, 2353. <https://doi.org/10.3390/rs16132353>

Academic Editors: Akira Iwasaki, Pedro Melo-Pinto and Salah Bourennane

Received: 30 April 2024

Revised: 22 June 2024

Accepted: 26 June 2024

Published: 27 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a pivotal task in remote sensing (RS) Earth observation, change detection (CD) facilitates the discernment of disparities in bitemporal RS imageries [1], especially using hyperspectral imagery (HSI). Actually, hyperspectral change detection (HCD) has numerous advantages over conventional color and multispectral images. The first is a higher spectral resolution and a wider spectral range spanning from visible near-infrared to short-wave infrared, resulting in hundreds of continuous spectral bands, offering a detailed understanding of the spectral characteristics of the observed object [2]. The second is the higher spatial resolution, which empowers it to capture intricate details, including object boundaries, texture features, and subtle surface changes [3]. As a result, HCD has emerged as a prominent area of research, proving to be successful across diverse domains, such as disaster monitoring [4], geological surveys [5], precision agriculture [6], and quality control [7].

In the last few decades, substantial progress has been achieved in the advancement of HCD tasks, including both unsupervised and supervised algorithms, as detailed below.

1.1. Unsupervised HCD Algorithms

In general, conventional methods for unsupervised HCD can be primary categorized into image algebra-based and image transformation-based approaches, depending on whether the raw data or the data in the transformed domain are used. Often the pixel-wise difference or spectral similarity is calculated first, followed by a thresholding- or clustering-based classification to determine the changed pixels. For instance, in image algebra-based methods, change vector analysis (CVA) [8] differs the magnitudes for calculating the Euclidean distance between each pair of spectral vectors, whilst spectral angle mapping (SAM) [9] measures the angular difference between the paired spectral vectors. Although these methods are straightforward and entail a relatively low computational cost, they tend to be noise sensitive and can hardly achieve satisfactory accuracy.

By converting the raw spectral data to a different domain, image transformation-based methods enhance the changed features, while reducing the data dimension and redundancy, before calculating the pixel difference or similarity. Typical methods include principal component analysis (PCA) [10] and linear discriminant analysis (LDA) [11], where the original high-dimensional HSI data are converted into a lower-dimensional representation, whilst preserving the key information to a generate difference image with a much-reduced data redundancy. Multivariate alteration detection (MAD) [12], developed on the basis of canonical correlation analysis [13], maximizes the correlation between the spatial bands and employs statistical methods, e.g., the chi-squared distribution, to determine significant changes. To further improve the detection accuracy, an iteratively reweighted (IR) MAD method [14] was proposed, determining changed and unchanged pixels with weights being updated during iterations. In [15], slow feature analysis was proposed to extract the most temporally invariant component from bitemporal images. As the unchanged features should be spectrally invariant and vary slowly, the differences in between are suppressed and easily separated from the changed pixels. Despite their effectiveness in reducing the dimension and redundancy of the spectral data, image transformation-based methods often fail to preserve spectral continuity, and hence damage the similarity between adjacent pixels during the transformation.

In addition, many advanced methods were proposed for unsupervised HCD, including unmixing-based [16,17], low-rank, and sparse representation-based [18,19] methods. In [20], a novel three-order Tucker decomposition and reconstruction detector was proposed for tensor processing across different domains and mitigating the impact of diverse factors present in the multi-temporal dataset, followed by spectral angle-based change detection. In [21], a sparse representation-based HCD method was proposed, by joint considering the background dictionary and the neighboring pixels around the test pixel. In [22], the kernel density estimation-based spectral distribution difference of adaptive regions after band selection is used to measure the change magnitude for HCD. In [23], a novel accumulated band-wise binary distancing method was proposed, where binary distancing only indicated whether a pixel was changed or not in a certain band, which could alleviate the adverse effect of a noise-induced inconsistency of measurements. The band-wise binary distance map is then created to form a grayscale change map, on which the simple k -means was applied for final binary decision making. Although these advanced unsupervised HCD methods can achieve good detection accuracy, they have high sensitivity and poor versatility regarding data pre-processing [24] and are lacking when considering prior spectral information in the settings [25]. Therefore, further efforts are needed to enhance the discriminability between changed pixels and the background and improve the accuracy in detecting multi-scale changed regions.

1.2. Supervised HCD Algorithms

With the wide application of machine learning techniques, they have been successfully applied in HCD, using spatial or spectral feature extractions and representations. Among these, one of the most commonly used is the support vector machine (SVM), where the binary classifier can be trained to detect changed pixels in bitemporal images [26–28].

In recent years, there has been a growing trend toward employing DL-based methods for supervised HCD that train the networks based on the true label information that comes from the ground truth map, because of their strong capability of adaptive feature extraction and thus more accurate HCD. In [29], an end-to-end 2-D convolutional neural network (CNN) was proposed that performs spectral unmixing on the input HSIs to obtain a mixed affinity matrix, followed by CNN-based feature mining. In [30], Recurrent 3-D CNNs are proposed to extract spatial–spectral features, incorporating a combined long short-term memory (LSTM) module to capture bi-temporally changed features. In [31], an end-to-end bilinear CNN (BCNN) was proposed with two symmetric CNNs for learning feature representations from bitemporal images. In [3], a multi-scale diff-changed feature fusion network was proposed to enhance feature representation by learning refined changing components between bitemporal HSIs at different scales. In [32], a dual-branch difference amplification graph convolutional network was proposed that fully extracts and effectively amplifies the difference spatial features of bitemporal images.

More recently, as the self-attention mechanism can focus on key information and its powerful modeling capabilities, it has been widely employed with deep learning [33]. It allows the DL network to independently learn a set of weighting coefficients, dynamically emphasizing the regions of interest within the data [34]. Therefore, the self-attention mechanism is widely used in HCD tasks. In [35], the cross-temporal interaction symmetric attention (CSA) network was proposed, where a self-attention module was employed for supporting the extraction and integration of joint spatial–spectral–temporal features to enhance feature representation. In [36], a joint spectral, spatial, and temporal transformer (SST-Former) was proposed for feature integration and the extraction of relevant change detection features from bitemporal HSIs. In [37], a new gate spectral–spatial–temporal attention network was proposed with a spectral similarity filtering module to reduce spectral redundancy whilst capturing intra-image spatial features and extracting inter-image temporal changes. In [38], a domain adaptive and interactive differential attention network was proposed that incorporated domain adaptive constraints to mitigate the pseudo-variation interface by mapping bitemporal images to a shared deep feature space for alignment. The proposed differential attention module could effectively improve feature representation and promote the interactive coupling of differential discriminant information.

Despite the remarkable efficacy in general, supervised-based algorithms often rely on a substantial volume of training data, which may not be readily available in real-world scenarios. In addition, these models typically entail high computational costs and have a large number of hyperparameters [39]. Therefore, how to address the lingering concerns regarding data scarcity and the substantial computational load remains the major challenges within existing DL-based models.

1.3. Remaining Challenges and Our Contributions

Although various models and approaches have been proposed with certain progress, HCD tasks still face major challenges as summarized below.

- Encompassing atmospheric effects and sensor limitations, HSI frequently contains various forms of noise and interference, which have a significant impact on the image quality and accuracy of change detection, especially for unsupervised algorithms [40].
- Existing deep learning (DL)-based HCD models often suffer from a considerable number of hyperparameters and highly redundant information in both the spatial and spectral domains, resulting in substantial computational costs [39].

- Existing models often fail to detect multi-scale changed regions, especially when sparsely distributed [41].

To tackle these challenges, a new, lightweight DL-based HCD model, SSA-LHCD, is proposed, which can produce higher CD accuracy, but has fewer hyperparameters, by combining singular spectrum analysis (SSA) and the change detection (CD) task. The major contributions of our work are summarized as follows.

- To apply the 1-D SSA for spectral domain denoising and mitigating the effect of noise on the tasks of feature extraction and change detection [42];
- To propose an efficient spectral feature-extraction module, which utilizes a residual block and an extra 1×1 convolutional layer to restrict the gradient propagation range via skip connections, and to adeptly capture the spectral features with instance normalization, further benefiting the greatly increased non-linear characteristics with fewer hyperparameters and computational costs [43];
- To employ a 2-D self-attention module to capture local spatial–spectral features. By dynamically adjusting the attention across diverse positions with multi-scale changing areas [44], feature representation and discrimination capability are improved through strategic weight allocation, resulting in significantly enhanced module reliability.

This paper is organized as follows. Section 2 elucidates the particulars of the proposed SSA-LHCD model. Section 3 discusses the experimental results for three publicly available datasets. Section 4 discusses the ablation experiment results regarding the parameter setting of the SSA-LHCD model. Then, a comprehensive discussion on the benchmark methods and all experiment results is summarized in Section 5. Finally, some concluding remarks are presented in Section 6.

2. Methodology

The SSA-LHCD network was designed in four main steps: (1) SSA-based pre-processing for noise removal; (2) spectral feature-extraction module, (3) 2-D self-attention based local spatial–spectral feature-extraction module, and (4) decision making. The details of the SSA-LHCD network are presented in Figure 1 and are further discussed in the following subsections.

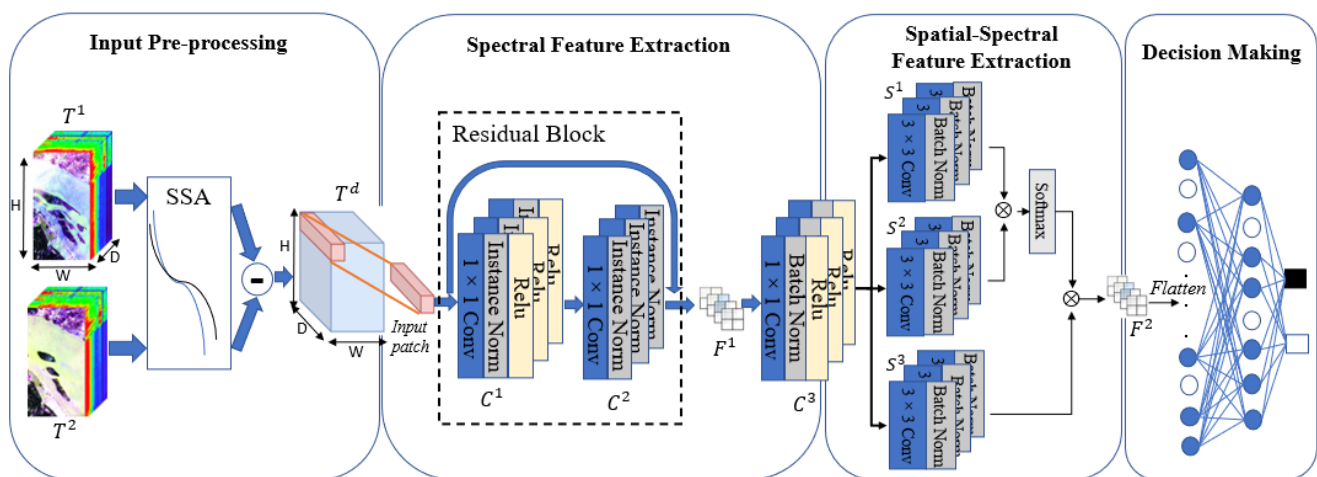


Figure 1. The architecture of the proposed end-to-end SSA-LHCD network.

2.1. SSA-Based Pre-Processing

In the conventional task of land-mapping, SSA was used to extract the representative spectral information from the HSI data [42]. For this purpose, each spectral profile was decomposed into several independent components, including the trend, oscillations, and noise [19], followed by spectral reconstruction using selected components whilst discarding the noisy ones. In HCD, a pair of bitemporal hypercubes, $T^1 \in \mathcal{R}^{W \times H \times B}$ and $T^2 \in \mathcal{R}^{W \times H \times B}$,

are presented, where W and H denote the width and height in the spatial domain, respectively, and B is the number of spectral bands. SSA is applied to reduce the inheritable noise in each hypercube, aiming to mitigate the noise caused by outliers in the differentiation process, as detailed below.

2.1.1. Embedding

Let $x = [x_1, x_2, \dots, x_B]$ denote a pixel-wise spectral vector; it will be firstly embedded to form a trajectory matrix, X , by an embedding window, L , $L \in [1, B]$.

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_B \end{pmatrix} \quad (1)$$

where $K = B - L + 1$, and each column of X is a lagged vector that can be considered as a Hankel matrix as it has equal values along the antidiagonals.

2.1.2. Eigen Decomposition

The singular value decomposition (SVD) is applied for eigen decomposition on the matrix x , where the eigenvalues and eigenvectors of XX^T are denoted as $(\lambda_1, \lambda_2, \dots, \lambda_L)$ and (U_1, U_2, \dots, U_L) , respectively. The trajectory matrix can be reconstructed as the sum of elementary matrices as follows.

$$X = X_1 + \dots + X_i + \dots + X_L, X_i = \sqrt{\lambda_i} U_i V_i^T, V_i = X^T U_i / \sqrt{\lambda_i} \quad (2)$$

2.1.3. Grouping and Projection

By dividing the total set of L components into M disjoint subsets (I_1, I_2, \dots, I_M) , where $\sum |I_m| = L$ and $m \in [1, M]$. Let $I = [i_1, i_2, \dots, i_p]$ represent a divided subset as $X_I = X_{I1} + X_{I2} + \dots + X_{Ip}$. Then, the trajectory matrix is represented by:

$$X = X_{I1} + X_{Ii} + \dots + X_{IM} \quad (3)$$

Let $Z_m = [Z_{m1}, Z_{m2}, \dots, Z_{mN}] \in \mathfrak{R}^N$ denote the 1-D signal projected from X_{Im} , which can be obtained via the diagonal averaging of each X_{Im} . Finally, the original 1-D signal, x , can be reconstructed using its eigenvalues in one or more principal groups, with highly noisy and less significant components discarded by:

$$SSA_{(x)} = z_1 + z_2 + \dots + z_M = \sum_{m=1}^M Z_m \quad (4)$$

An example of a 1-D SSA application is shown in Figure 2, which shows a pair of corresponding non-changed pixels from the bitemporal images on the River dataset, including the original spectral vectors and their difference, as well as the SSA-smoothed results and the new difference. As can be seen, both the original corresponding spectral signals and the differential spectral signal preserve the basic trend of the profile whilst smoothing out the noise and thus the outliers in the difference signal for more robust change detection.

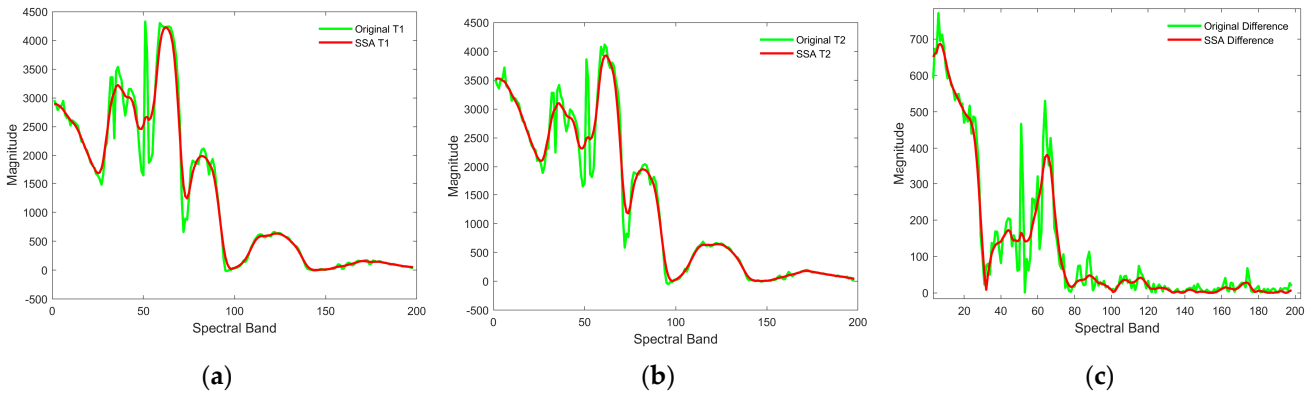


Figure 2. Examples showing a pair of unchanged pixels from the River dataset, where the outliers in the difference signal have been mitigated via SSA-based noise removal from the original profiles. (a) Original T1/SSA T1, (b) original T2/SSA T2, and (c) original difference/SSA difference.

2.2. Spectral Feature Extraction

From the SSA-smoothed hypercubes of T^1 and T^2 , their absolute difference can be obtained as a new hypercube, T^d :

$$T^d = \left| SSA_{(T^2)} - SSA_{(T^1)} \right| \quad (5)$$

where $T^d \in \mathfrak{R}^{W \times H \times B}$. To produce more training samples, T^d is divided into 3-D overlapped patches, $P_{(\alpha, \beta)} \in \mathfrak{R}^{O \times O \times B}$, with a window size, O ; (α, β) denote the coordinates of the patch center in the spatial domain, where $\alpha \in [1, W]$, $\beta \in [1, H]$, and the truth label is decided by the centered pixel. In our experiments, 20% of pixels from both the changed and unchanged regions were randomly selected for training, while the remaining were used for testing.

The spectral feature-extraction module is structured into two main components. The initial part is a residual block composed of $C^{(1)}$ and $C^{(2)}$, aiming to extract features in the spectral domain. $C^{(1)}$ is constructed with a 1×1 convolutional layer, serving as an initial extraction for spectral features, followed by an instance normalization (IN) layer and the rectified linear unit (ReLU) activation function. The incorporated IN layer [45] independently normalizes each pixel rather than the entire batch, thereby ensuring that the features of each sample have similar means and variances, which aids in speeding up the convergence and enhancing the model's generalization capability. Additionally, the non-linear properties of the ReLU [46] activation function foster the acquisition of more intricate functions and alleviate the issue of vanishing gradients, consequently amplifying the model's capacity for non-linear representation.

For $C^{(2)}$, it also has a 1×1 convolutional layer and an IN layer for deep spectral feature extraction, enhancing the network's representation capacity. Unlike $C^{(1)}$, the difference lies in the absence of a ReLU activation function, aiming to preserve the feature information for compatibility with the residual connection. The incorporated skip connections facilitate the addition of input features to the output features. This design enables the network to effectively capture the residual information between the input, $P_{(\alpha, \beta)}$, and output, F^1 , as follows.

$$F^1 = C^{(2)} \left(C^{(1)} \left(P_{(\alpha, \beta)} \right) \right) + P_{(\alpha, \beta)} \quad (6)$$

Utilizing the 1×1 convolution operation on the input layer enables the linear combination of features across different channels, resulting in the generation of novel feature representations. This process enhances the network's representational capacity and the overall performance by extracting more expressive features. The subsequent component is a profound spectral feature-extraction layer, denoted as the 1×1 convolutional layer, $C^{(3)}$, accompanied by the batch normalization (BN) [47] and ReLU functions. $C^{(3)}$ is adept at adjusting the number of convolutional kernels, thereby effectively reducing the number

of spectral channels. This dimensionality reduction serves a dual purpose, i.e., trimming down the number of parameters and computational complexity, whilst simultaneously preserving crucial spectral features. The outcomes are the improved computational efficiency of the network and dimension reduction-mitigated challenges associated with gradient vanishing, fostering improved information propagation within the network [48]. In other words, the combination of these features has not only refined the network's efficiency, but also addressed key issues related to gradient flow and parameter optimization.

2.3. Spatial–Spectral Feature Extraction

Inspired by the work in [44], here we integrate a 2-D self-attention module into the proposed SSA-LHCD model, serving local spatial–spectral feature extraction [49], aiming to boost the stability of the feature extractor within the model.

Taking the output of the spectral feature extractor as the input, after traversing through three successive 2-D convolutional layers of $S^{(1)}$, $S^{(2)}$, and $S^{(3)}$, it can generate a novel spatial–spectral feature map. By converting the feature map into a 2-D attention matrix, it can facilitate the creation of a refined spatial–spectral feature representation. This multi-step process enriches the model's ability to capture intricate relationships and latent dependencies within the input. This comprehensive representation encapsulates both spatial and spectral information, offering a robust foundation for subsequent stages of the model. Upon completing the deep spatial–spectral feature extraction, the final extracted feature map is derived as $F^{(2)}$ as follows:

$$F^{(2)} = (S^{(3)}(Z_{(\alpha,\beta)}))^T \times \text{Softmax}\left(S^{(1)}(Z_{(\alpha,\beta)})^T \times S^{(2)}(Z_{(\alpha,\beta)})\right) \quad (7)$$

Incorporating the self-attention module empowers the SSA-LHCD model to capture intricate spatial–spectral dependencies, fostering enhanced stability and robustness in feature extraction for a diverse range of applications.

2.4. Decision Making

Change detection can be regarded as a binary classification problem for distinguishing changed and unchanged pixels. Firstly, the spatial–spectral features, $F^{(2)}$, obtained from the previous spatial–spectral feature-extraction stage are flattened into a one-dimensional vector. This transformation prepares the features for input into a fully connected neural network suitable for decision making. The flattened feature vector is then fed into a series of fully connected layers. Each layer performs linear transformations, followed by nonlinear activations to learn complex patterns and relationships within the input features, $F^{(2)}$. Subsequently, the final layer of the fully connected network employs a SoftMax activation function, which converts the network's outputs into a probability distribution over two classes. The final classification decision is made by selecting the class with the highest probability, thus achieving a binary classification.

The selected optimizer is adaptive momentum (Adam) [50] and the selected loss function is cross-entropy [51], with the initial learning rate of 0.0001. The specific details of each layer in the end-to-end SSA-LHCD model are summarized in Table 1.

Table 1. Architecture details of each layer in the SSA-LHCD model.

Layers	Type	Channels	Kernel
T^1	SSA Pre-processing	B	-
T^2	SSA Pre-processing	B	-
T^d	Difference	B	-
C^1	Conv2D + IN + Relu	B	1×1
C^2	Conv2D + IN	B	1×1

Table 1. *Cont.*

Layers	Type	Channels	Kernel
C^3	Conv2D + BN + Relu	64	1×1
S^1	Conv2D + BN	32	3×3
S^2	Conv2D + BN	32	3×3
S^3	Conv2D + BN	32	3×3
Flatten	Flatten	288	-
FC^1	Linear (Dropout = 0.4)	64	-
FC^2	Linear (Dropout = 0.4)	8	-
FC^3	Linear (Dropout = 0.4)	2	-

3. Experiments

3.1. Dataset Description

The datasets utilized in our experiments were obtained from the Hyperion sensor installed on the Earth Observing-1 (EO-1) satellite, which offers a total of 242 bands in the range of 0.4–2.5 μm , with a spatial resolution of 10 m [52]. The three datasets used are River [53], Yancheng [54], and Hermiston [55]. The pseudo-colored images and detailed information about these datasets are presented in Figure 3 and Table 2, respectively.

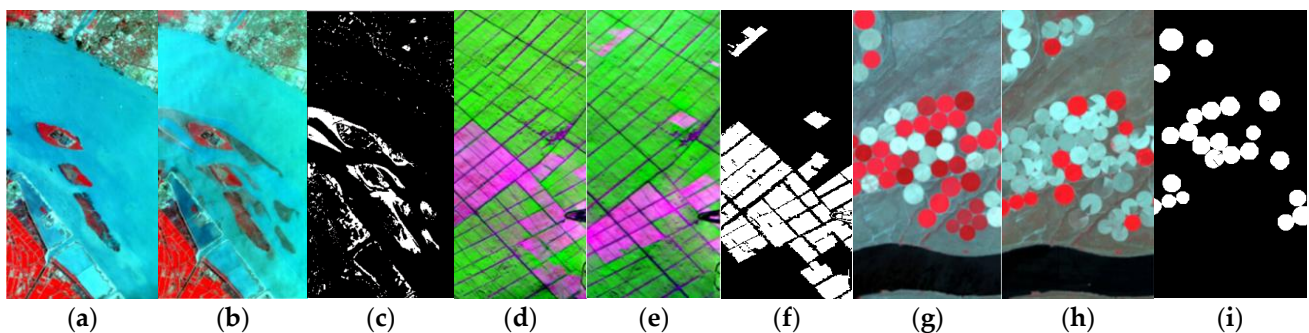


Figure 3. Pseudo-colored and ground truth images of the three datasets. (a) River on 3 May 2013. (b) River on 31 December 2013. (c) Ground truth of River. (d) Yancheng on 3 May 2006. (e) Yancheng on 23 April 2007. (f) Ground truth of Yancheng. (g) Hermiston on 1 May 2004. (h) Hermiston on 8 May 2007. (i) Ground truth of Hermiston.

Table 2. Details of the datasets used in our experiments.

	River	Yancheng	Hermiston
Date for T^1	3 May 2013	3 May 2006	1 May 2004
Date for T^2	31 December 2013	23 April 2007	8 May 2007
Location	Jiangsu, China	Yancheng, China	Oregon, US
Spatial Size	463×241	420×140	390×200
Bands	198	154	242
Unchanged Pixels	101,885	40,417	68,014
Changed Pixels	9698	18,383	9986

3.2. Evaluation Criteria

By considering the change detection task as a binary classification problem, where changed and unchanged pixels are denoted as 1 (positive) and 0 (negative), respectively,

the overall accuracy (OA) and Kappa coefficient (KP) are used for quantitative performance evaluations. OA here indicates the percentage of correctly classified pixels as defined below:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

where TP , TN , FP , and FN denote the correctly detected changed pixels, correctly detected unchanged pixels, incorrectly detected changed pixels, and incorrectly detected unchanged pixels, respectively.

The KP is used to measure the inter-rater reliability as the degree of similarity between the change map and the ground truth:

$$KP = \frac{OA - PRE}{1 - PRE} PRE = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + TN + FP + FN)^2} \quad (9)$$

Recall (Re) represents the ratio of the number of TP observations to the total number of actual positives.

$$Re = \frac{TP}{TP + FN} \quad (10)$$

The $F1$ score ($F1$) defines a balanced index that can be considered as the harmonic mean of precision (Pre) and Re , where Pre is defined as the ratio of the number of TP observations to the total number of predicted positive observations.

$$Pre = \frac{TP}{TP + FP} \quad (11)$$

$$F1 = 2 \times \frac{Pre \times Re}{Pre + Re} \quad (12)$$

3.3. Results and Comparison

In this section, we assess the efficacy of the proposed method by comparing it with three state-of-the-art unsupervised change detection methods, as well as five supervised methods. A brief summary of these compared approaches is presented as follows.

- AD [41]: The absolute difference between spectral values is accumulated as the change map, followed by the k -means binary classification.
- CVA [13]: The Euclidean distance between two spectral vectors is calculated, followed by OTSU thresholding to determine the change map.
- PCA-KM [15]: PCA to reduce the data dimension and redundancy, followed by k -means clustering for the binary classification of changed pixels.
- SVM [26]: Supervised machine learning-based method that extracts pixel-wise spectral information as feature vectors, with the raw SVM used as the binary classifier.
- Two-dimensional CNN [42]: Deep spatial feature extracted by using multi-layer and multi-scale 2-D CNNs.
- CSANet [30]: Self-attention-based method that extracts and integrates joint spatial-spectral-temporal features by incorporating a traditional self-attention module to enhance feature representation within each temporal one.
- ML-EDAN [56]: A two-stream encoder-decoder model to integrate hierarchical features from convolutional layers in bitemporal images, enhanced by a contextual information-guided attention module for improved spatial-spectral feature transfer and an LSTM subnetwork to analyze temporal dependencies.
- CBANet [12]: Self-attention-based method integrating a cross-band feature-extraction module and a 2-D self-attention module, thereby enhancing the feature representation and discrimination capability.

The benchmarks are established according to the specified parameters in the default settings, where DL-based methods are trained using PyTorch on NVIDIA RTX A2000, with a batch size of 32 and 200 training epochs. For training, 20% of pixels from both changed

and unchanged regions are randomly selected, while the remaining pixels are used for testing. To ensure fairness and reliability, each supervised method is repeated ten times in our experiments, and the averaged results of OA and KP are reported for comparison. In addition, comparisons of the resulting change maps as well as quantitative evaluations using *Pre*, *Re*, and *F1* for all supervised methods are conducted. In the resulting change maps, false alarms and missing pixels are highlighted in red and green, respectively, while correctly detected changed areas are presented in white, and true negatives are depicted in black for an easy visual comparison.

3.3.1. Results of the River Dataset

The extracted change maps and quantitative results from the River dataset for all benchmarks are shown in Figure 4 and Table 3, respectively. Although all three unsupervised algorithms achieve *Re* values no less than 99.5%, they exhibit quite a low *Pre*, i.e., excessive false alarms. These false alarms are visibly concentrated in the upper and lower left corners of the change maps, as seen in Figure 4a–c, due to the misclassification of subtle sporadic change pixels in the River dataset. Consequently, the *Pre* of all unsupervised algorithms drops to below 66%, where all *KP* values fall below 0.75. For the supervised methods, however, the extracted maps exhibit much less false alarms, yet there is a prevalent issue of missing detections, especially for 2-D CNN and ML-EDAN approaches. These results demonstrate a relatively low detection accuracy, as indicated by *OA* values below 97% and *KP* values hovering around 0.80. Interestingly, the SVM performs marginally better than the 2-D CNN and ML-EDAN, with the *OA* boosted to 97.02% and *KP* to 0.8109. However, the SVM has the highest standard deviation of 0.0078 in *OA* among all supervised methods. Not surprisingly, thanks to the SSA pre-processing and proposed feature-extraction modules, our SSA-LHCD model outperforms all benchmarks on the River dataset, surpassing the CBANet by 0.24% in the *OA* and 0.0144 in the *KP*.

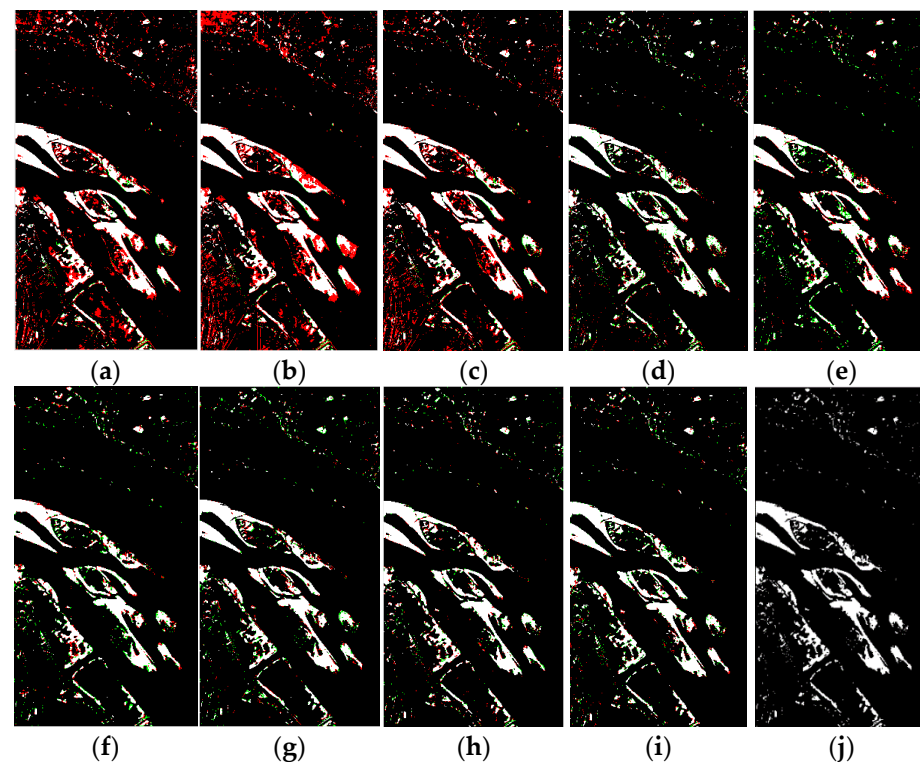


Figure 4. Extracted change maps on the River dataset from different methods of AD (a), CVA (b), PCA (c), SVM (d), 2-D CNN (e), CSANet (f), ML-EDAN (g), CBANet (h), and SSA-LHCD (i) in comparison to the ground truth map (j), where the false alarms and missing pixels are labelled in red and green.

Table 3. Quantitative assessment of different methods on the River dataset.

	<i>OA (%)</i>	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
AD	94.31	0.7137	0.6108	0.9515	0.7440
CVA	92.53	0.6528	0.5393	0.9635	0.6915
PCA-KM	95.17	0.7478	0.6524	0.9506	0.7738
SVM	97.02 ± 0.0078	0.8109 ± 0.0049	0.8358	0.8417	0.8387
2-D CNN	96.82 ± 0.0007	0.7946 ± 0.0033	0.9073	0.8888	0.8978
CSANet	97.43 ± 0.0012	0.8360 ± 0.0049	0.9130	0.9175	0.9152
ML-EDAN	96.96 ± 0.0014	0.8009 ± 0.0049	0.9220	0.8975	0.9093
CBANet	97.65 ± 0.0036	0.8526 ± 0.0036	0.9405	0.9119	0.9256
SSA-LHCD	97.89 ± 0.0007	0.8670 ± 0.0026	0.9322	0.9343	0.9332

3.3.2. Results of the Yancheng Dataset

Similar to the results in the River dataset, the inadequate performance of all three unsupervised methods is evident, as shown in Figure 5a–c. The quantitative results on Yancheng dataset are shown in Table 4. These methods have a notable number of missing detection of pixels, coupled with the presence of false alarms, particularly in striped lines and other field regions. As a result, the *KP* values remain consistently low, hovering around 0.71, with the *OA* dropping below 90%; both *Pre* and *Re* are below 90%. Here, the SVM becomes the poorest performer among all supervised algorithms, with *OA* and *KP* values of only 94.87% and 0.8806, respectively, due mainly to the SVM’s limitation in pixel-wise learning without considering the spatial features. In contrast, deep learning-based approaches have an *OA* exceeding 96% and a *KP* over 0.92. Nevertheless, our SSA-LHCD model remains the best, showcasing the highest average values for both the *KP* and *OA*. Furthermore, the standard deviation of the *KP* is only 0.0012, the lowest among all supervised methods, along with the highest *F1* score compared with all benchmark tests. These outcomes serve as compelling evidence, substantiating the effectiveness and robustness of our proposed SSA-LHCD model.

Table 4. Quantitative assessment of different methods on the Yancheng dataset.

	<i>OA (%)</i>	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
AD	87.80	0.7074	0.8430	0.7494	0.7935
CVA	87.55	0.7025	0.8327	0.7529	0.7908
PCA	88.28	0.7180	0.8557	0.7519	0.8004
SVM	94.87 ± 0.0013	0.8806 ± 0.0029	0.9063	0.9110	0.9086
2-D CNN	96.67 ± 0.0014	0.9223 ± 0.0030	0.9608	0.9557	0.9582
CSANet	97.15 ± 0.0009	0.9335 ± 0.0023	0.9658	0.9641	0.9650
ML-EDAN	97.15 ± 0.0012	0.9316 ± 0.0034	0.9685	0.9517	0.9598
CBANet	97.13 ± 0.0006	0.9332 ± 0.0014	0.9645	0.9633	0.9639
SSA-LHCD	97.16 ± 0.0011	0.9365 ± 0.0012	0.9680	0.9701	0.9691

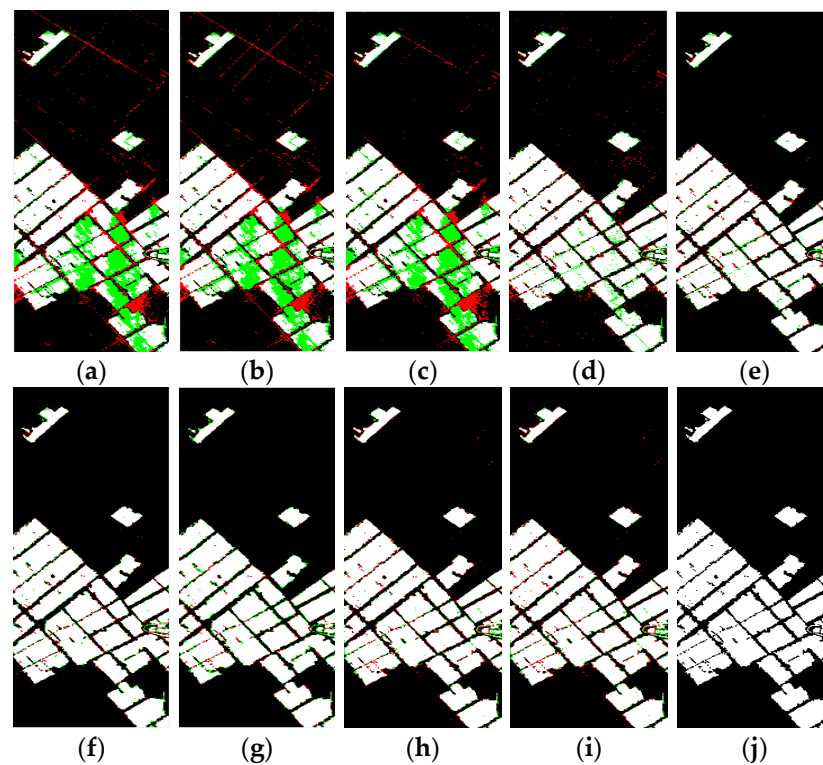


Figure 5. Extracted change maps of the Yancheng dataset from the different methods of AD (a), CVA (b), PCA (c), SVM (d), 2-D CNN (e), CSANet (f), ML-EDAN (g), CBANet (h), and SSA-LHCD (i) in comparison to the ground truth map (j), where the false alarms and missing pixels are labelled in red and green.

3.3.3. Results of the Hermiston Dataset

For the Hermiston dataset, the extracted change maps and quantitative assessment are shown in Figure 6 and Table 5, respectively. Due to the absence of scattered variation pixels and the distinct visibility of all changed features, the *OA* values of all benchmarks surpassed 97%, or are over 99% for all supervised methods, though the SVM remains the worst supervised model due to the lack of spatial features. Here, our SSA-LHCD model emerges as the second-highest performer among all deep learning models, while the *OA* is only 0.11% lower than the top-performing ML-EDAN, and the *KP* is merely 0.0067 lower than the leading CBANet. This is attributed to the relatively homogeneous change type in this dataset, limiting the prominence of deep spectral feature extraction. The primary disparity lies in the detected edges of changed regions. CBANet, with its incorporation of deep spatial feature learning and small kernels, accurately identifies pixels along the edges of each change region through spatial feature extraction. In contrast, our SSA-LHCD model focuses solely on extracting spectral features by utilizing the 1×1 convolutional layer and residual block and does not explicitly learn deep spatial features like the CBANet. As a result, the *OA* and *KP* of our model on the Hermiston dataset are slightly lower than those of the CBANet and ML-EDAN.

Table 5. Quantitative assessment of different methods on the Hermiston dataset.

	<i>OA</i> (%)	<i>KP</i>	<i>Pre</i>	<i>Re</i>	<i>F1</i>
AD	97.28	0.8824	0.8625	0.9367	0.8981
CVA	98.43	0.9035	0.8978	0.9351	0.9161
PCA	97.89	0.9068	0.9060	0.9322	0.9189

Table 5. Cont.

	OA (%)	KP	Pre	Re	F1
SVM	99.07 ± 0.0002	0.9581 ± 0.0012	0.9519	0.9759	0.9638
2-D CNN	99.12 ± 0.0004	0.9662 ± 0.0077	0.9819	0.9779	0.9799
CSANet	99.23 ± 0.0006	0.9659 ± 0.0031	0.9822	0.9705	0.9763
ML-EDAN	99.32 ± 0.0001	0.9669 ± 0.0008	0.9806	0.9820	0.9813
CBANet	99.28 ± 0.0010	0.9745 ± 0.0030	0.9808	0.9883	0.9845
SSA-LHCD	99.21 ± 0.0009	0.9678 ± 0.0008	0.9781	0.9909	0.9844

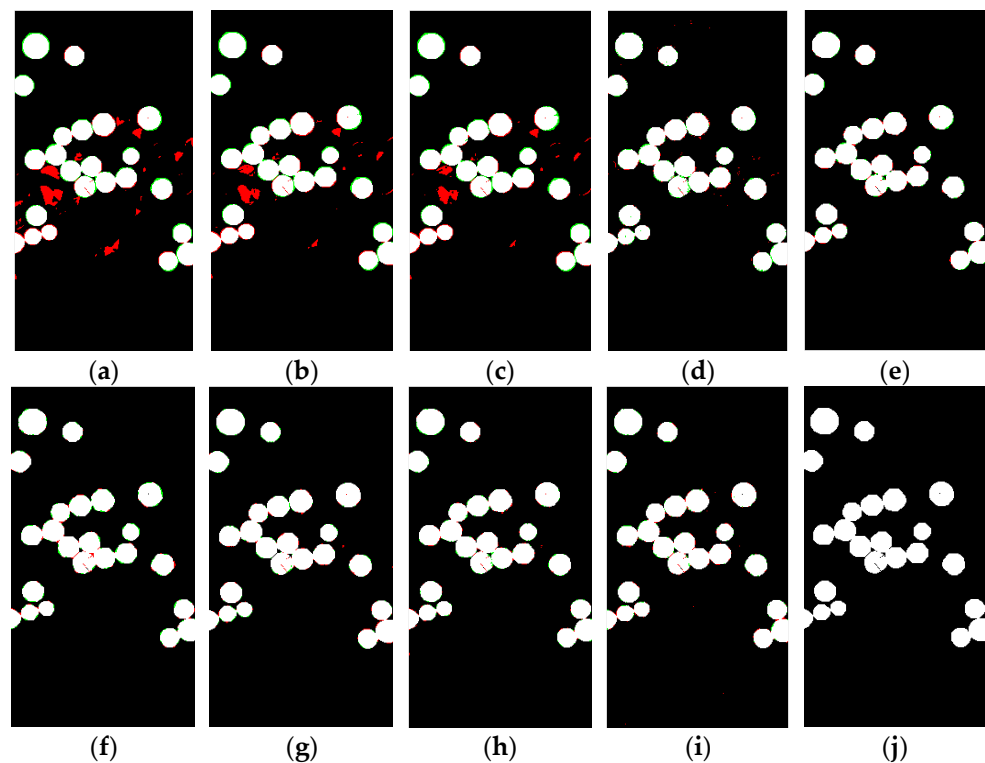


Figure 6. Extracted change maps of the Hermiston dataset from the different methods of AD (a), CVA (b), PCA (c), SVM (d), 2-D CNN (e), CSANet (f), ML-EDAN (g), CBANet (h), and SSA-LHCD (i) in comparison to the ground truth map (j), where the false alarms and missing pixels are labelled in red and green.

4. Ablation Study

To comprehensively validate the effectiveness of our proposed SSA-LHCD model, we conducted a series of experiments covering computational hyperparameters, the effect of modular block, and different numbers of spectral or spatial–spectral feature-extraction kernels.

4.1. Hyperparameter Analysis

In Table 6, we compare the numbers of hyperparameters and floating-point operations (FLOPs) and the overall running time in minutes (m), including both training time and testing time, for all the DL-based models, including ours, on the River dataset. For those using multi-layer CNNs, such as the 2-D CNN, CSANet, and ML-EDAN, the numbers are much higher, i.e., over two magnitudes, than ours. The inclusion of the 1×1 convolutional kernel in the spectral feature module and the residual block contribute to the lightweight nature of our SSA-LHCD model, which also outperforms other benchmarking methods.

Table 6. Complexity comparison of DL methods on the River dataset.

	2-D CNN	CSANet	ML-EDAN	CBANet	SSA-LHCD
Parameters (k)	607.43	2452.88	88,933.34	319.36	167.24
FLOPs (M)	368.21	144.44	590.22	6.66	2.80
Running Time (m)	35.42	53.43	76.27	18.53	14.21

4.2. Effect of Modular Blocks and Patch Size

In this section, we conducted three sets of experiments: (i) SSA-LHCD without SSA pre-processing, (ii) SSA-LHCD without the residual block, and (iii) SSA-LHCD with both module blocks with different patch sizes on all three datasets. We tested five patch sizes of $\{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, \text{ and } 11 \times 11\}$, and the results are presented in Figure 7a–c, where the training ratio is set as 20%. First, the performance of the SSA-LHCD model degrades when either the SSA or the residual block is absent, showing their unique value to the proposed model. Second, with the increasing patch size, the KP values on the Hermiston dataset increase and reach 0.9697 at the patch size of 11×11 . However, the varying trends of the KP on the River and Yancheng datasets are different, which show an initial increase followed by a decrease when the patch size exceeds 5×5 . The observed phenomenon can be attributed to the abundance of sparsely distributed change or non-change pixels in the River and Yancheng datasets, whilst the Hermiston dataset only contains large, regular regions. Smaller patch sizes are better suited for extracting these scattered pixels effectively, and when using a large patch size, may lead to false alarms. However, in the case of the Hermiston dataset with distinct spectral features, the designed deep spectral feature module loses its advantage. Conversely, larger patch sizes encompass more spatial edge information, leading to improved edge detection accuracy for small changed areas. For balancing the detection accuracy and computational efficiency, we chose a patch size of 5×5 for our SSA-LHCD model.

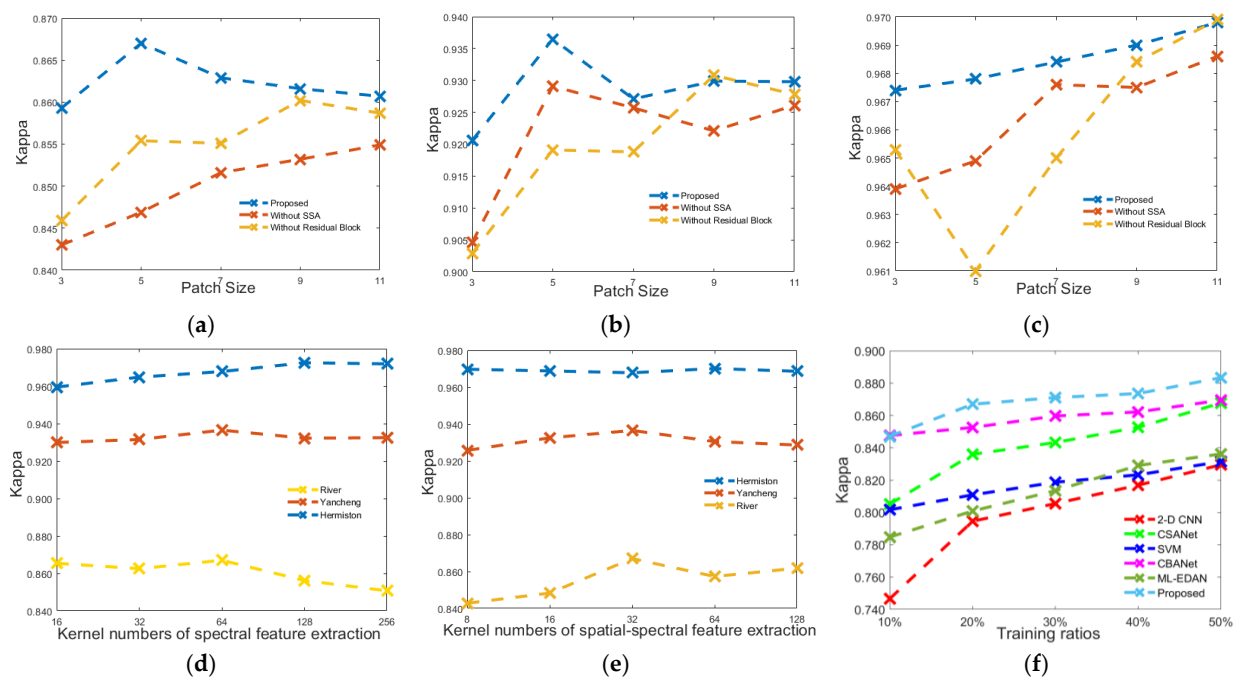


Figure 7. Ablation experiments and results of the SSA-LHCD model in different settings on the three datasets, including the Kappa values of different patch sizes on River (a), Yancheng (b), and Hermiston datasets (c); different kernel numbers of the spectral feature-extraction module (d); different kernel numbers of the 2-D self-attention module (e); and different training ratios of all DL-based benchmarks on the River dataset (f).

4.3. Number of Spectral Feature-Extraction Kernels

To determine the optimal number of kernels in the spectral feature-extraction module, five different settings of 16, 32, 64, 128, and 256 were tested. As shown in Figure 7d, the varying trends of the *KP* on the three datasets appear similar to those from increasing the patch size. When the kernel number of the spectral feature-extraction module is set to 64, the highest *KP* values can be achieved on the River and Yancheng datasets. It is worth noting that, when the kernel number is set to 128, the average *KP* value on the Hermiston dataset is 0.9735, which is very close to the *KP* value of the CBANet. However, for the overall performance of the proposed network, we decided to set the kernel number of the spectral feature-extraction module to 64 for all datasets.

4.4. Number of Spatial–Spectral Feature-Extraction Kernels

We also evaluated the selection of the number of 2-D self-attention kernels by conducting experiments using five different settings, including 8, 16, 32, 64, and 128. The variation trend of *KP* on the three datasets is shown in Figure 7e. Similarly, the kernel number is set to 32 to balance the model's parameters and robustness.

4.5. Training Ratios

To further validate the efficacy of our SSA-LHCD model, its performance is assessed on the River dataset, considering varying percentages of training ratios from 10% to 50%. As shown in Figure 7f, a larger training ratio generally leads to an improved detection accuracy, where our model consistently achieves the highest *KP*. Specifically, when the training ratio is 50%, our model can achieve a *KP* of 0.8843, surpassing the second best, CBANet, by a margin of 1.39%.

5. Further Discussion

The proposed SSA-LHCD network demonstrates significant advantages in terms of higher detection accuracy and fewer hyperparameters compared to the benchmarked state of the arts. This are mainly due to the residual block-based spectral feature-extraction module and the 2-D self-attention-based spatial–spectral feature-extraction module, as well as SSA-based pre-processing to effectively reduce noise whilst preserving valuable features, enabling our lightweight DL network to extract spectral and spatial–spectral features more effectively.

As shown in the compared results for the three datasets, image algebra-based, CVA and AD and image transformation-based PCA are all noise sensitive. Furthermore, the threshold segmentation or clustering processes in these methods fail to accurately classify subtle changes, leading to numerous false alarms or missing pixels. The SVM, as a classical supervised binary classifier, is trained using pixel-wise spectral vectors. Consequently, its detection accuracy is significantly lower compared to methods based on spatial features and spatial–spectral feature extraction using deep learning models. This is due to the SVM's inability to capture spatial features, which are crucial for precise classification.

In comparison to DL-based state-of-the-art approaches, our proposed SSA-LHCD method outperforms almost all of them, offering higher detection accuracy with reduced hyperparameters. The 2-D CNN method uses multi-layer 2-D convolutions with large kernels to extract local spatial features from input patches, yet it fails to account for spectral features. This limitation results in the lowest detection accuracy for the three datasets, especially the River dataset, which contains many sporadic pixels. The influence of neighboring pixels due to the large kernel used leads to the misclassification of many sporadic changed pixels as unchanged ones, resulting in a high number of missing pixels. Conversely, on the Yancheng and Hermiston datasets, which consist of connected regions, a large number of false alarms are detected at the edges of the connected areas. By using multi-level spatial–spectral feature extraction via encoder–decoder and LSTM subnetworks, the ML-EDAN becomes the most complex network among all the compared models, with the number of parameters and FLOPs being approximately 531-times and 210-times greater than those of

our proposed SSA-LHCD model, respectively. Based on the Siamese 2-D CNN structure, the CSANet extracts the joint spatial–spectral–temporal features of corresponding patches, along with the cross-temporal self-attention module utilized to integrate the jointed features oriented from each temporal embedding. Similarly, also as a self-attention-based network, the CBANet can effectively extract spectral and spatial–spectral features. The detection accuracy of the CSANet and CBANet, two self-attention mechanism-based algorithms, rank second and third, respectively, in all benchmark tests on the River and Yancheng datasets. However, it is still inferior to our proposed SSA-LHCD model, due to the inability of the SSA module to mitigate noise for the effective extraction of spectral features.

From the ablation experiments, as shown in Figure 7a–c, it can be found that the SSA pre-processing step and residual block-based spectral feature-extraction module significantly improve the detection accuracy under different patch sizes. For the River and Yancheng datasets with varying scales of change, a smaller patch size results in higher detection accuracy. However, for the Hermiston dataset, characterized by a regular change scale and single change type, larger patch sizes increase the detection accuracy. Regarding the kernel numbers of the spectral feature-extraction module and the spatial–spectral feature-extraction module, the SSA-LHCD model achieves the best detection accuracy for the Hermiston dataset when set to 128 and 64, respectively. For a balanced performance across different datasets, we set the kernel numbers of these two modules to 64 and 32, which yielded the best results for both the River and Yancheng datasets. Furthermore, the SSA-LHCD model can achieve the best detection performance across various training settings, indicating that a higher detection accuracy can be achieved with fewer training samples.

In summary, DL-based methods tend to outperform SVM and unsupervised approaches in HCD. As shown in all the quantitative results, the three self-attention-based models, CSANet, CBANet, and our SSA-LHCD model, outperform the two models that only use multi-scale 2-D convolutional layers and the Siamese autoencoder-based network, as shown in both the visual map comparison results and quantitative results. Notably, when examining the change maps, the three self-attention-based models demonstrate a superior detection performance, particularly for sparsely distributed change regions. For the River dataset with many sporadic pixels, the SSA-LHAD model achieves the best detection accuracy. Overall, our approach exhibits significant advantages over other existing models, especially for detecting different scales of changes.

There remain certain limitations to our proposed method. Currently, the difference of the presented HSI pairs after SSA pre-processing is taken as the input, followed by a single channel 1×1 convolutional layer for the deep extraction of the spectral features. In current implementations, only the trend signal of the SSA is used. Considering that the other components can also be potentially useful, their effects will be explored further in our future work. Although the proposed SSA-LHCD network has surpassed the state-of-the-art benchmarks in overall accuracy when using fewer training samples, it still requires manually labeled data due to its supervised nature. This dependency on manual labeling is a significant limitation in practical applications.

6. Conclusions

In this paper, a novel, lightweight end-to-end DL-based network (SSA-LHCD) is proposed for HCD. First, bitemporal HSIs were pre-processed using SSA for noise reduction. Initial change features are then extracted through subtraction. Following this, a residual block-based spectral feature-extraction module is employed to refine these initial change features by effectively capturing spectral information. Subsequently, a 2-D self-attention mechanism is integrated to capture local spatial–spectral features, enhancing both feature representation and discrimination capabilities. Finally, a fully connected layer serves as the classifier, facilitating binary HCD decision making.

SSA-based noise reduction, 1×1 convolutional layers, and the residual block significantly improve the model's overall performance of change detection by enabling efficient

spectral feature learning. Moreover, the inclusion of the 2-D self-attention module is crucial for capturing complex spatial–spectral features, further enhancing the model’s ability to discriminate changed regions, thus improving HCD accuracy. Comprehensive experiments demonstrate SSA-LHCD’s superiority over eight state-of-the-art methods on three publicly available datasets, highlighting its capability to produce higher detection accuracy with fewer hyperparameters. This innovative approach offers significant advancements in HCD by enhancing noise reduction, multi-scale change handling, and computational efficiency, setting a new benchmark in supervised HCD.

Our next step is to enhance the feature-extraction process by incorporating the features of the remaining principal components. By comparing different degrees of components, we aim to extract change features and generate pseudo-ground truths. Using the existing network as the basis, we will utilize these pseudo-labels to train the network, achieving self-supervised learning and thereby completely eliminating the need for manually labeled data. Additionally, we plan to enhance the SSA-LHCD model by incorporating other advanced techniques, such as multi-scale deformable attention modules and adaptive fusion [57], to cope with various sizes of changed regions.

Author Contributions: Conceptualization, J.R., Y.Y. and G.S.; methodology, Y.L., J.R. and Y.Y.; software, Y.L., J.R. and P.M.; validation, Y.Y. and P.M.; formal analysis, J.R. and G.S.; investigation Y.L., Y.Y. and P.M.; resources, J.R.; data curation, Y.L. and P.M.; writing—original draft preparation, Y.L.; writing—review and editing, J.R., Y.Y., P.M. and G.S.; visualization, Y.Y. and P.M.; supervision, J.R. and G.S.; project administration, J.R.; funding acquisition, J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by the Sea Sense project, funded by Net Zero Technology Centre, UK and the Robert Gordon University PhD scholarship.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

1. Yan, Y.; Ren, J.; Liu, Q.; Zhao, H.; Sun, H.; Zabalza, J. PCA-Domain Fused Singular Spectral Analysis for Fast and Noise-Robust Spectral-Spatial Feature Mining in Hyperspectral Classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *20*, 5505405. [[CrossRef](#)]
2. Ma, P.; Ren, J.; Sun, G.; Zhao, H.; Jia, X.; Yan, Y.; Zabalza, J. Multiscale Superpixelwise Prophet Model for Noise-Robust Feature Extraction in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5508912. [[CrossRef](#)]
3. Luo, F.; Zhou, T.; Liu, J.; Guo, T.; Gong, X.; Ren, J. Multiscale Diff-Changed Feature Fusion Network for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–13. [[CrossRef](#)]
4. Zhang, N.; Zhang, X.; Yang, G.; Zhu, C.; Huo, L.; Feng, H. Assessment of Defoliation during the Dendrolimus Tabulaeformis Tsai Et Liu Disaster Outbreak using UAV-Based Hyperspectral Images. *Remote Sens. Environ.* **2018**, *217*, 323–339. [[CrossRef](#)]
5. Fu, H.; Sun, G.; Zhang, L.; Zhang, A.; Ren, J.; Jia, X.; Li, F. Three-Dimensional Singular Spectrum Analysis for Precise Land Cover Classification from UAV-Borne Hyperspectral Benchmark Datasets. *ISPRS J. Photogramm. Remote Sens.* **2023**, *203*, 115–134. [[CrossRef](#)]
6. Ang, K.L.; Seng, J.K.P. Big Data and Machine Learning with Hyperspectral Information in Agriculture. *IEEE Access* **2021**, *9*, 36699–36718. [[CrossRef](#)]
7. Yan, Y.; Ren, J.; Sun, H.; Williams, R. Nondestructive Quantitative Measurement for Precision Quality Control in Additive Manufacturing using Hyperspectral Imagery and Machine Learning. *IEEE Trans. Ind. Informat.* **2024**, 1–13. [[CrossRef](#)]
8. Malila, W.A. Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat. In Proceedings of the LARS Symposia, West Lafayette, IN, USA, 3–6 June 1980; p. 385.
9. Yang, C.; Everitt, J.H.; Bradford, J.M. Yield Estimation from Hyperspectral Imagery using Spectral Angle Mapper (SAM). *Trans. ASABE* **2008**, *51*, 729–737. [[CrossRef](#)]
10. Deng, J.S.; Wang, K.; Deng, Y.H.; Qi, G.J. PCA-based Land-use Change Detection and Analysis using Multitemporal and Multisensor Satellite Data. *Int. J. Remote Sens.* **2008**, *29*, 4823–4838. [[CrossRef](#)]
11. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of Hyperspectral Images with Regularized Linear Discriminant Analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [[CrossRef](#)]
12. Nielsen, A.A.; Conradsen, K.; Simpson, J.J. Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies. *Remote Sens. Environ.* **1998**, *64*, 1–19. [[CrossRef](#)]

13. Yang, X.; Liu, W.; Liu, W.; Tao, D. A Survey on Canonical Correlation Analysis. *IEEE Trans. Knowled. Data Eng.* **2019**, *33*, 2349–2368. [[CrossRef](#)]
14. Nielsen, A.A. The Regularized Iteratively Reweighted MAD Method for Change Detection in Multi-and Hyperspectral Data. *IEEE Trans. Image Process.* **2007**, *16*, 463–478. [[CrossRef](#)]
15. Wu, C.; Du, B.; Zhang, L. Slow Feature Analysis for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 2858–2874. [[CrossRef](#)]
16. Li, Q.; Mu, T.; Tuniyazi, A.; Yang, Q.; Dai, H. Progressive Pseudo-Label Framework for Unsupervised Hyperspectral Change Detection. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *127*, 103663. [[CrossRef](#)]
17. Liu, W.; Ma, Y.; Wang, X.; Huang, J.; Chen, Q.; Li, H.; Mei, X. UADNet: A Joint Unmixing and Anomaly Detection Network Based on Deep Clustering for Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5511419. [[CrossRef](#)]
18. Wu, C.; Du, B.; Zhang, L. Hyperspectral Anomalous Change Detection Based on Joint Sparse Representation. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 137–150. [[CrossRef](#)]
19. Gao, L.; Hong, D.; Yao, J.; Zhang, B.; Gamba, P.; Chanussot, J. Spectral Superresolution of Multispectral Imagery with Joint Sparse and Low-Rank Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 2269–2280. [[CrossRef](#)]
20. Hou, Z.; Li, W.; Tao, R.; Du, Q. Three-Order Tucker Decomposition and Reconstruction Detector for Unsupervised Hyperspectral Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6194–6205. [[CrossRef](#)]
21. Ertürk, A.; Iordache, M.; Plaza, A. Sparse Unmixing with Dictionary Pruning for Hyperspectral Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *10*, 321–330. [[CrossRef](#)]
22. Lv, Z.; Lei, Z.; Xie, L.; Falco, N.; Shi, C.; You, Z. Novel Distribution Distance Based on Inconsistent Adaptive Region for Change Detection using Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4404912. [[CrossRef](#)]
23. Li, Y.; Ren, J.; Yan, Y.; Maher, A.; Gao, Z. ABBD: Accumulated Band-wise Binary Distancing for Unsupervised Parameter-Free Hyperspectral Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 9880–9893. [[CrossRef](#)]
24. Lu, Y.; Saeys, W.; Kim, M.; Peng, Y.; Lu, R. Hyperspectral Imaging Technology for Quality and Safety Evaluation of Horticultural Products: A Review and Celebration of the Past 20-Year Progress. *Postharvest Biol. Technol.* **2020**, *170*, 111318. [[CrossRef](#)]
25. Liu, L.; Lei, S.; Shi, Z.; Zhang, N.; Zhu, X. Hyperspectral Remote Sensing Imagery Generation from RGB Images Based on Joint Discrimination. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7624–7636. [[CrossRef](#)]
26. Bovolo, F.; Bruzzone, L.; Marconcini, M. A Novel Approach to Unsupervised Change Detection Based on a Semisupervised SVM and a Similarity Measure. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2070–2082. [[CrossRef](#)]
27. Demir, B.; Bovolo, F.; Bruzzone, L. Detection of Land-Cover Transitions in Multitemporal Remote Sensing Images with Active-Learning-Based Compound Classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 1930–1941. [[CrossRef](#)]
28. Ahlqvist, O. Extending Post-Classification Change Detection using Semantic Similarity Metrics to Overcome Class Heterogeneity: A Study of 1992 and 2001 US National Land Cover Database Changes. *Remote Sens. Environ.* **2008**, *112*, 1226–1241. [[CrossRef](#)]
29. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 3–13. [[CrossRef](#)]
30. Song, A.; Choi, J.; Han, Y.; Kim, Y. Change Detection in Hyperspectral Images using Recurrent 3D Fully Convolutional Networks. *Remote Sens.* **2018**, *10*, 1827. [[CrossRef](#)]
31. Lin, Y.; Li, S.; Fang, L.; Ghamisi, P. Multispectral Change Detection with Bilinear Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1757–1761. [[CrossRef](#)]
32. Qu, J.; Xu, Y.; Dong, W.; Li, Y.; Du, Q. Dual-Branch Difference Amplification Graph Convolutional Network for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5519912. [[CrossRef](#)]
33. Xie, G.; Ren, J.; Marshall, S.; Zhao, H.; Li, R.; Chen, R. Self-Attention Enhanced Deep Residual Network for Spatial Image Steganalysis. *Digital Signal Process.* **2023**, *139*, 104063. [[CrossRef](#)]
34. Zheng, X.; Wang, B.; Du, X.; Lu, X. Mutual Attention Inception Network for Remote Sensing Visual Question Answering. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5606514. [[CrossRef](#)]
35. Song, R.; Ni, W.; Cheng, W.; Wang, X. CSANet: Cross-Temporal Interaction Symmetric Attention Network for Hyperspectral Image Change Detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6010105. [[CrossRef](#)]
36. Wang, Y.; Hong, D.; Sha, J.; Gao, L.; Liu, L.; Zhang, Y.; Rong, X. Spectral-Spatial-Temporal Transformers for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5536814. [[CrossRef](#)]
37. Yu, H.; Yang, H.; Gao, L.; Hu, J.; Plaza, A.; Zhang, B. Hyperspectral Image Change Detection Based on Gated Spectral-Spatial-Temporal Attention Network with Spectral Similarity Filtering. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5511313. [[CrossRef](#)]
38. Ji, Y.; Sun, W.; Wang, Y.; Lv, Z.; Yang, G.; Zhan, Y.; Li, C. Domain Adaptive and Interactive Differential Attention Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5616316. [[CrossRef](#)]
39. Sun, W.; Du, Q. Hyperspectral Band Selection: A Review. *IEEE Geosci Remote Sens. Mag.* **2019**, *7*, 118–139. [[CrossRef](#)]
40. Rasti, B.; Scheunders, P.; Ghamisi, P.; Licciardi, G.; Chanussot, J. Noise Reduction in Hyperspectral Imagery: Overview and Application. *Remote Sens.* **2018**, *10*, 482. [[CrossRef](#)]
41. Wen, D.; Huang, X.; Bovolo, F.; Li, J.; Ke, X.; Zhang, A.; Benediktsson, J.A. Change Detection from very-High-Spatial-Resolution Optical Remote Sensing Images: Methods, Applications, and Future Directions. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 68–101. [[CrossRef](#)]

42. Zabalza, J.; Ren, J.; Wang, Z.; Marshall, S.; Wang, J. Singular Spectrum Analysis for Effective Feature Extraction in Hyperspectral Imaging. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1886–1890. [[CrossRef](#)]
43. Yang, J.; Wang, X.; Wang, R.; Wang, H. Combination of Convolutional Neural Networks and Recurrent Neural Networks for Predicting Soil Properties using Vis–NIR Spectroscopy. *Geoderma* **2020**, *380*, 114616. [[CrossRef](#)]
44. Li, Y.; Ren, J.; Yan, Y.; Liu, Q.; Ma, P.; Petrovski, A.; Sun, H. CBANet: An End-to-End Cross Band 2-D Attention Network for Hyperspectral Change Detection in Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5513011. [[CrossRef](#)]
45. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv* **2016**, arXiv:1607.08022. [[CrossRef](#)]
46. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
47. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
48. Zhang, H.; Dauphin, Y.N.; Ma, T. Fixup Initialization: Residual Learning without Normalization. *arXiv* **2019**, arXiv:1901.09321. [[CrossRef](#)]
49. Tolie, H.F.; Ren, J.; Elyan, E. DICAM: Deep Inception and Channel-Wise Attention Modules for Underwater Image Enhancement. *Neurocomputing* **2024**, *584*, 127585. [[CrossRef](#)]
50. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.6980. [[CrossRef](#)]
51. Zhang, Z.; Sabuncu, M. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *Adv. Neural Inf. Process. Syst.* **2018**, *31*.
52. Henry, P.; Chander, G.; Fougny, B.; Thomas, C.; Xiong, X. Assessment of Spectral Band Impact on Intercalibration Over Desert Sites using Simulation Based on EO-1 Hyperion Data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1297–1308. [[CrossRef](#)]
53. 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](#)]
54. Li, X.; Yuan, Z.; Wang, Q. Unsupervised Deep Noise Modeling for Hyperspectral Image Change Detection. *Remote Sens.* **2019**, *11*, 258. [[CrossRef](#)]
55. Seydi, S.T.; Hasanlou, M.; Amani, M. A New End-to-End Multi-Dimensional CNN Framework for Land Cover/Land use Change Detection in Multi-Source Remote Sensing Datasets. *Remote Sens.* **2020**, *12*, 2010. [[CrossRef](#)]
56. Qu, J.; Hou, S.; Dong, W.; Li, Y.; Xie, W. A Multilevel Encoder–Decoder Attention Network for Change Detection in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5518113. [[CrossRef](#)]
57. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable Detr: Deformable Transformers for End-to-End Object Detection. *arXiv* **2020**, arXiv:2010.04159. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.