GASSM: global attention and state space model based end-to-end hyperspectral change detection.

LI, Y., REN, J., FU, H. and SUN, G.

2025

Journal of the Franklin Institute 362 (2025) 107424

Contents lists available at ScienceDirect

Journal of the Franklin Institute

journal homepage: www.elsevier.com/locate/fi

GASSM: Global attention and state space model based end-to-end hyperspectral change detection

Yinhe Li [a], Jinchang Ren [b,a,*], Hang Fu [a,c], Genyun Sun [c]

[a] National Subsea Centre, School of Computing, Engineering and Technology, Robert Gordon University, AB21 0BH Aberdeen, UK
[b] School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China
[c] College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China

ARTICLE INFO

Keywords:
Hyperspectral images (HSI)
Hyperspectral change detection (HCD)
Feature extraction
Global attention mechanism
State space model (SSM)

ABSTRACT

As an essential task to identify anomalies and monitor changes over time, change detection enables detailed earth observation in remote sensing. By combining both the rich spectral information and spatial image, hyperspectral images (HSI) have offered unique and significant advantages for change detection. However, traditional hyperspectral change detection (HCD) methods, predominantly based on convolutional neural networks (CNNs), struggle with capturing long-range spatial-spectral dependencies due to their limited receptive fields. Whilst transformers based HCD methods are capable of modeling such dependencies, they often suffer from quadratic growth of the computational complexity. Considering the unique capabilities in offering robust long-range sequence modeling yet with linear computational complexity, the emerging Mamba model has provided a promising alternative. Accordingly, we propose a novel approach that integrates the global attention (GA) and state space model (SSM) to form our GASSM network for HCD. The SSM based Mamba block has been introduced to model global spatial-spectral features, followed by a fully connected layer to perform binary classification of detected changes. To the best of our knowledge, this is the first to explore using the Mamba and SSM for HCD. Comprehensive experiments on two publicly available datasets, compared with eight state-of-the-art benchmarks, have validated the efficacy and efficiency of our GASSM model, demonstrating its superiority of high accuracy and stability in HCD.

1. Introduction

Change detection (CD) is a crucial tool in remote sensing (RS) earth observation, aiming to identify differences in bi-temporal images within the same geographic area [1]. Conventional RS imaging techniques, including the Synthetic Aperture Radar (SAR), multi-spectral and red-green-blue (RGB) color imagery, often face limitations in feature discrimination. By combining the pixel-wise 1-D spectral signature with integrated spectroscopy as a standard 2-D image, the advent of HSI based RS can successfully address these limitations and enable more accurate identification and analysis of land cover and associated change [2]. The basic process of bi-temporal HCD is illustrated in Fig. 1, which typically includes four key stages: i.e. pre-processing, change feature extraction, change map segmentation and accuracy evaluation. In HCD tasks, the dominant tasks involve the extraction of change features and the segmentation of the difference images to derive the change map. These two critical steps are the primary focus of existing research in

* Corresponding author.
E-mail address: jinchang.ren@ieee.org (J. Ren).

https://doi.org/10.1016/j.jfranklin.2024.107424
Received 22 August 2024; Received in revised form 22 October 2024; Accepted 24 November 2024
Available online 14 January 2025

this area.

In addition, compared with conventional RS techniques, HSI boasts two primary advantages: 1) high spectral resolution and broad spectral range that encompasses the visible, near-infrared (NIR), and even short-wave infrared (SWIR) and ultraviolet (UV) ranges, with spectral resolutions as fine as 10 nm or less across hundreds of continuous bands [3]; and 2) rich spatial and spectral information that enables more effective and precise characterization and detection of objects, even under subtle differences between similar objects [4]. As a result, hyperspectral change detection (HCD) has been successfully applied in a wide range of applications, such as geological surveys [5] and disaster monitoring [6] in RS as well as vegetable quality nondestructive inspection [7] and precision agriculture [8].

Based mainly on convolution neural networks (CNNs), traditional deep learning networks optimize the parameters through gradient descent but face limitations in modelling long-range spatial-spectral feature dependencies, which restricts their contextual reasoning capabilities in understanding of global context. Furthermore, methods based on transformers, while capable of modeling long-range dependencies, suffer from quadratic growth in computational complexity. Recently, there is a trend to integrate the state space models (SSMs) into deep learning networks for sequence modeling [9]. A notable advancement involves introducing a selection mechanism into SSM, allowing them to selectively propagate or discard information based on the current token in the sequence or scan path for effective modeling long-range dependencies with linear computational complexity [10]. Subsequently, selective SSMs have been integrated with simplified linear transformation and CNN architecture, namely Mamba. Mamba combines the modeling power of transformers with linear scalability by considering sequence length to analyze the long-range dependencies of features, positioning it as a promising foundation model for various sequence modeling tasks. The versatility of Mamba has been demonstrated across multiple RS applications, including semantic segmentation [11], HSIs classification [12] and HSIs dehazing [13].

In this paper, we present a pioneering attempt to combine the SSM based Mamba with the global attention (GA) and spatial-spectral feature extraction modules in the proposed end-to-end deep learning network, GASSM for HCD tasks. To enable the Mamba block to integrate spatial and spectral understanding, the introduced GA module can capture and consolidate the complex relationships in spatial and spectral domains, enhancing spatial-spectral feature interactions. To ensure efficient feature extraction of Mamba block, we designed a spatial-spectral feature extraction module, aiming to reduce feature redundancy while extracting high-quality features. This has not only preserved key information but also reduced computational complexity. Finally, the Mamba block extracts long-range features and models global spatial-spectral features, followed by a fully connected layer to determine the HCD outcomes. Experiments on two publicly available datasets have validated its superior performance when benchmarked with eight state-of-the-art methods.

The remainder of this article is organized as follows. Section 2 introduces the related work for HCD. Section 3 describes the details of the proposed GASSM. Section 4 presents the experimental results and assessments. Finally, some remarkable conclusions are summarized in Section 5.

2. Related work

Despite the rich spatial information and high spectral resolution, HSI often suffers from high redundancy and various types of noise, due mainly to the sensor limitations and atmospheric effects during the acquisition of RS data [14]. Moreover, with the increasing number of spectral bands, differences in external imaging conditions can lead to more pseudo changes, thereby affecting the detection accuracy. Over the past decades, numerous HCD methods have been proposed to address these challenges, as detailed below, including both unsupervised and supervised approaches.

2.1. Unsupervised HCD methods

Early conventional methods for HCD are algebra-based approaches, which relied on measuring pixel-wise spectral differences to assess the extent of change. These methods primarily include image difference, image ratio, absolute distance (AD) [15]. The most representative method is change vector analysis (CVA) [16], which all the spectral bands of bi-temporal HSIs are adopted to construct a vector, and the Euclidean distance is utilized to measure the change magnitude between the corresponding pixel-wise vector. A compressed CVA (C^2VA) [17] was proposed to obtain a 2-D compressed representation of the multi-dimensional change vectors. Additionally, spectral angle mapper (SAM) [18] was developed to compare the angle between corresponding spectral vectors to determine their similarity. The pixel-wise changes extracted by these algebra-based methods are processed by using thresholding or clustering methods to generate a binary change map to distinguishing changed and unchanged pixels. While these algebra-based methods are computationally fast, they fail to consider the spatial characteristics of raw images and are sensitive to noise, which causes these algebra-based methods inefficient.

Instead of relying on raw spectral features, image transformation-based methods compute the pixel differences in the transformed spatial feature domain in order to highlight the changed features whilst reducing the redundancy and dimension of the data. One

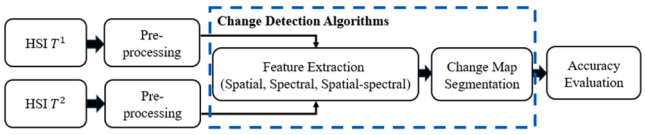

Fig. 1. Basic change detection process of bi-temporal HSIs.

typical method is the principal component analysis with *k*-means (PCAKM) [19], which projects linearly the high-dimensional HSI data to a low-dimensional space whilst preserving most of the energy or information. Multivariate change detection (MAD) used canonical correlation analysis [20] to maximize the correlation between the change and unchanged features of bi-temporal images. Subsequently, iteratively reweighted MAD (IRMAD) [21] was proposed that conducts the weighted iteration according to the chi-square distance. Slow feature analysis (SFA) [22] was proposed that extract invariant and slowly varying features from bi-temporal from the difference image after subtracting bi-temporal images and transform the image into a new low-dimensional feature space. Image transformation-based methods are effective in utilizing spatial information and reduce the dimensionality and redundancy of high-dimensional data. By projecting the original data into a transformed space, these methods can distinguish between changed and unchanged features. However, these transformation-based methods suffer from poor explainability. The complex mathematical transformations involved make it difficult to directly relate the transformed results back to the physical meaning of the original data [23]. Additionally, these transformation-based methods damage the similarity between adjacent pixels and ignore spectral continuity [24].

In recent years, some advanced unsupervised HCD methods have been proposed based on statistic [25][26], band selection [27][28] and spectral unmixing [29][30], etc. In [26], an automated image analysis method was proposed that relies on the utilization of difference images and histogram statistics for HCD. This method distinguishes alterations based on both positive and negative values within the difference image and establishes location-specific thresholds by identifying minimum points within the histogram. In [28], a spatial-contextual feature extraction method was proposed that incorporated band selection to reduce spectral redundancy and an iterative spatial-adaptive filter for noise reduction. The change magnitude was evaluated using CVA and adaptive regions, followed by a binary thresholding using the OTSU method. In [30], joint unmixing and spatial information co-guidance technologies were combined to extract endmembers and estimate abundance, incorporating spectral perturbed regularization to enhance robustness against spectral variability. In [31], a novel unsupervised method based on three-order tucker decomposition and reconstruction detector was proposed, which combines a singular value accumulation to determine the principal components in the factor matrices, tucker decomposition and reconstruction to eliminate the influence of various factors in the bi-temporal images, and SAM to analyse spectral changes after tensor processing in different domains. In [32], a new unsupervised HCD method was proposed based on discrete representation of change information, focusing on the represented change information in each spectral band. The band-wise radiation information is analysed to generate a quantized discrete representation of the change vector, followed by a tree representation to distinguish between different types of changes. In summary, these innovative unsupervised HCD methods do offer significant advantages over traditional algebra-based and transformation-based algorithms, particularly in terms of improving detection accuracy [33]. However, these methods also have notable limitations, such as requiring manual intervention to adjust parameters, affecting their usability and lacks generalization capabilities [34]. In addition, the introduction of numerous parameters increases the complexity of these methods [35].

2.2. *Supervised HCD methods*

Supervised HCD algorithms include conventional machine learning and DL-based methods that train the networks using the labelled ground truth maps [36]. In conventional shallow machine learning, support vector machine (SVM) and random forest (RF) are commonly used binary classification algorithms. SVM [37] treats each pixel as an independent sample, with the spectral values across different bands forming a feature vector for each pixel. Using these feature vectors, the SVM trains a classifier to find an optimal hyperplane that separates the pixels of different categories on a per-pixel basis. RF [38] performs classification by constructing multiple decision trees and determining the final category of each pixel through a majority voting mechanism. These shallow classifiers treat each pixel as an independent sample, which allows for parallel processing to improve detection efficiency [39]. However, they overlook the spatial features between adjacent pixels [40].

Recently, with the advancement in computing power and the development of DL technology, deep neural networks [41] have gradually become an effective alternative to conventional machine learning methods. The multi-layered structures can model tasks in finer detail and improve discriminative capabilities. In the field of HSIs processing, many current algorithms rely on convolutional neural networks (CNNs) [42] and recurrent neural networks (RNNs) [43] was proposed to extract multiscale spatial-spectral features and discriminative temporal features. In [44], 1-D and 2-D CNNs are used to extract spectral and spatial features, respectively. In [45], a general end-to-end 2-D CNN-based method was proposed, which performs spectral unmixing on the input HSIs to obtain a mixed affinity matrix, and then uses multiple 2-D CNN layers to mine the feature information. In [46], a bilinear CNN model was proposed that employs two symmetric CNNs to extract spatial-spectral features of the pair of HSI patches, respectively. In [47], a recurrent 3-D CNN fully convolutional network was proposed, in which 3-D CNN layers are employed to extract spatial-spectral features whilst multi-temporal change features are extracted by combining LSTM. In [48], a multipath convolutional LSTM and CNN through the designed parallel architecture to extract multilevel temporal dependencies and multiscale temporal–spatial-spectral features. In [49], a multiscale diff-changed feature fusion network was proposed based on multiple different scale 2-D CNN layers to enhance the feature representation by learning refined changing components between bi-temporal HSIs at different scales. As the self-attention mechanism can focus on key information and its powerful modelling capabilities, has become a widely used and effective feature extraction method in the field of HSI-CD and is usually used to enhance the feature expression ability of CNNs [50]. For instance, in [51], a deep multiscale pyramid network was proposed and enhanced with a spatial-spectral residual attention module. The multiscale pyramid convolution can effectively excavate multilevel and multiscale spatial-spectral features, and these features were further enhanced by a spatial-spectral residual attention module of each scale. In [52], the cross-temporal interaction symmetric attention (CSA) network was proposed, where a self-attention module was employed to support the extraction and integration of joint spatial-spectral-temporal

features to enhance feature representation. In [53], a joint spectral, spatial, and temporal transformer was proposed for feature integration and extraction of relevant change features from bi-temporal HSIs based on a self-attention mechanism. In [54], a new gate spectral-spatial-temporal attention network was proposed with a spectral similarity filtering module to reduce the spectral redundancy whilst capturing intra-image spatial features and extracting inter-image temporal changes. In [55], a domain adaptive and interactive differential attention network was proposed that incorporated domain adaptive constraints to mitigate pseudo-variation interface by mapping bi-temporal images to shared deep feature space for alignment. Although supervised DL-based methods produce good detection accuracy in general, they often rely on a substantial volume of training data, which the ground truth map needs to be manually annotated and 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 [56]. Therefore, how to address the substantial computational load remains the major challenge within existing supervised DL-based models.

In summary, based on the analysis of each category of HCD algorithms, their advantages and disadvantages are summarized in Table 1, including traditional unsupervised methods and supervised methods.

3. Methodology

The diagram of the proposed GASSM is presented in Fig. 2, which is composed of four main modules, i.e., 1) global attention mechanism based on spectral and spatial attention; 2) feature extraction module for minimizing redundancy; 3) SSM-based Mamba block for capturing long-range dependencies; and 4) decision making of detected HCD results.

Given a pair of spatially aligned bi-temporal hypercubes $T^1 \in \Re^{W*H*B}$and $T^2 \in \Re^{W*H*B}$, where W and H denote the width and height of the spatial size, and B represents the number of spectral bands. Then, their absolute difference can be obtained as a new hypercube T^d:

$$T^d = \left|T^2 - T^1\right| \tag{1}$$

where $T^d \in \Re^{W*H*B}$. To produce more training samples, T^d is divided into 3-D overlapped patches $P_{(\alpha,\beta)} \epsilon \Re^{S*S*B}$ with a window size S (set $S = 7$ in this network); (α, β) denote the coordinates of the patch center in the spatial domain, where $\alpha \epsilon[1, W]$, $\beta\epsilon[1, H]$ and the truth label is decided by the centered pixel. In our experiments, 20 % of pixels from both the changed and unchanged regions are randomly selected for training, while the remaining are used for testing.

3.1. Global attention mechanism

The global attention mechanism (GAM) [57] integrates the spectral and spatial attention to minimize information loss and enhance dimension-interactive features. GAM leverages a combination of a multilayer perceptron (MLP) for spectral attention and incorporates a convolutional-based spatial attention submodule. The spectral attention submodule employs a two-layer MLP, structured as an encoder-decoder, which focuses on extracting the spectral information while strengthening the dependencies across spectral dimensions.

The spatial attention (*SA*) submodule enhances the spatial features by using two convolutional layers for spatial fusion and eliminates pooling to avoid information loss, thus preserving feature maps. This combined approach effectively retains more information and improves the interaction of features across dimensions, ensuring a comprehensive understanding and utilization of complex data structures. The output of GAM, denoted as F^2 can be represented as

$$F^2 = SA(Spectral\ Attention\ (Input\ patch)) \tag{2}$$

here, the output F^2 and the input patch have the same size, with $F^2 \in \Re^{S*S*B}$.

3.2. Feature extraction

In the feature extraction stage, following the global attention, the refined feature F^2 are subjected to further spectral and spatial

Table 1
Summary of the advantages and disadvantages of each category of HCD Algorithms.

		Advantages	Disadvantages
Unsupervised	Algebra-based	Computationally fast No need for labelled data	Fail to consider the spatial characteristics and sensitive to noise
	Transformation-based	Consider the spatial characteristics No need for labelled data	Damage the similarity between adjacent pixels and ignore spectral continuity
Supervised	SVM-based machine learning	Treat each pixel as an independent sample, allows for parallel processing to improve detection efficiency	Overlook the spatial features between adjacent pixels
	CNN&RNN-based deep learning	Produce good detection accuracy	Rely on a substantial volume of training data, and the ground truth map needs to be manually annotated
	Transformers-based	Can capture long-range dependencies of features	Suffer from quadratic growth of the computational complexity

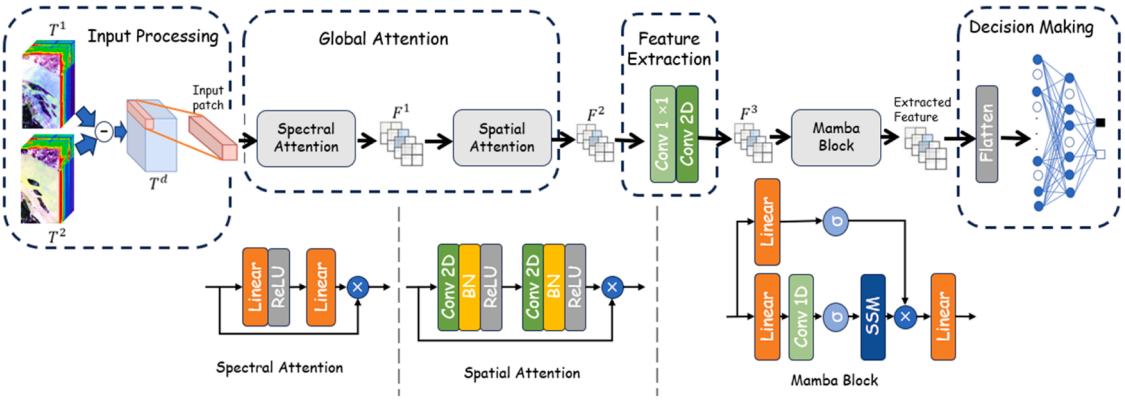

Fig. 2. The architecture of proposed network.

feature extraction through convolutional operations. This process incorporates a combination of 1×1 and 2-D convolutional layer to effectively process F^2 and generate F^3.

The 1×1 convolution layer [58] is primarily used to extract the spectral features and decrease the number of channels, thereby enhancing computational efficiency. By processing each pixel across the channels, the network can learn interactions between different channels without changing the spatial dimensions. By selecting an appropriate number of kernels (set to 128 in this network), This not only improves computational efficiency but also allows for increasing the network depth, which helps in learning more complex features. Furthermore, the 1×1 convolution introduces more nonlinearity, thereby enhancing the model's expressive power.

On the other hand, the 2-D convolution [59] captures spatial features by employing a larger convolution kernel (set to 5×5 in the network). This allows it to extract features from a wider spatial area, identifying relationships and patterns among pixels in local regions, thus generating feature maps enriched with spatial structure information. In addition, by setting fewer number of kernels (set to 64 in this network), the channels of F^3 are further reduced.

By combining 1×1 convolution and 2-D convolution, this feature extraction stage not only reduces the spatial and spectral dimensions of the features but also preserves and enhances critical spectral and spatial characteristics. As a result, the generated feature map F^3 contains both refined spectral information and essential spatial structure details, providing high-quality input for the subsequent Mamba block. By this point, the size of feature map F^3 is $F^3 \in \Re^{3*3*64}$.

3.3. SSM-based mamba block

The Mamba block integrates linear transformations, a 1-D convolutional layer and a SSM to further extract the long-term dynamic dependencies of F^3. This integration is designed to both local and global dependencies within the feature maps, enhancing the model's ability to learn more complex patterns. The detailed operations of the Mamba block are given below.

The input feature maps F^3 is first processed by a linear transformation and a 1-D convolutional layer. The operations can be expressed as:

$$\widehat{F} = SiLU\left(Conv1D\left(Linear\left(F^3\right)\right)\right) \tag{3}$$

where the Sigmoid-Weight linear units (*SiLU*) [60] is a smooth, non-linear activation function, which can retain input information and improve the gradient flow during backpropagation. The 1-D convolutional layer, represented as *Conv*1*D*, applies the convolutional filters across the transformed feature map. The output of this operation, denoted as $\widehat{F}$, serves as the input for the subsequent SSM operations.

The SSM is employed to model the temporal dynamics within the transformed feature maps. The SSM consists of state update and observation equations that capture the dependencies across different time steps. The observation equation is defined as:

$$y_t = C_t * h_{t+1} \tag{4}$$

here, y_t represents the observation output at time step t, and C_t is a time-varying matrix that transforms the state vector h_{t+1} into the observation space. This transformation is crucial for linking the hidden states to the observable outputs. The state update equation is given by:

$$h_{t+1} = dA * h_t + \widehat{F} * dB \tag{5}$$

where, h_t is the state vector at time step t, and h_{t+1} is the state vector at the next time step h_{t+1}. The matrices dA and dB are dynamically adjusted state transition and input matrices, respectively, which modulate the influence of the previous state and the current input on the next state.

The matrices B_t and C_t, which are essential for the state update and observation processes, are computed as follows:

$$B_t = fc^1(\widehat{F}) \tag{6}$$

$$C_t = fc^2(\widehat{F}) \tag{7}$$

here, fc^1 and fc^2 are distinct linear transformation functions in SSM that applied to the transformed feature map $\widehat{F}$. These functions typically consist of fully connected layers that transform $\widehat{F}$ into the respective matrices B_t and C_t. These matrices play crucial roles in adjusting the state transitions and linking hidden states to observations.

The dynamic factor δ_t and the dynamic matrices dA and dB are computed as:

$$\delta_t = softplus\left(fc^1\ (\widehat{F})\right) \tag{8}$$

$$dA = \exp(\delta_t * A) \tag{9}$$

$$dB = \delta_t * B_t \tag{10}$$

where, the *softplus* function is used as an activation function to ensure that δ_t is positive, providing a smooth and continuous dynamic adjustment factor. The matrix A is the initial state transition matrix, which is modulated by δ_t to produce the dynamic state transition matrix dA. Similarly, δ_t adjusts the matrix B_t to produce the dynamic input matrix dB. These dynamic matrices allow the model to adaptively adjust the influence of past states and current inputs, enhancing its ability to capture temporal dependencies.

The Mamba block uses another set of linear transformation combined with *SiLU* activation to capture and extract the static features of F^3. By combining these static features with the dynamic features obtained from SSM, the model can better capture long-term dependencies and short-term variations in the data. This enhances the overall representational power and robustness of the model, improving its performance and accuracy. This process can be illustrated as:

$$Mamba_{out} = Linear\ \left(y_t\ * SiLU\left(Linear\left(F^3\right)\right)\right) \tag{11}$$

$Mamba_{out}$ denotes the 2D feature matrix output by the Mamba block, which is converted into a feature map with spatial characteristics, consisting of m channels, which is empirically set to 64. The extracted feature map has the same spatial size and number of channels as F^3. Finally, this extracted feature is passed to the final stage for binary decision-making.

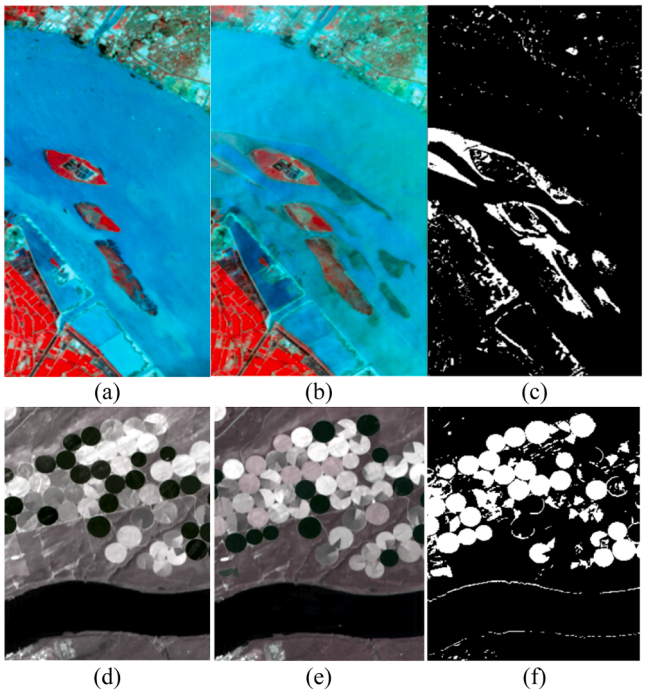

Fig. 3. Pseudo-colored and Ground truth images of the two datasets. (a) River on May 3, 2013. (b) River on December 31, 2013. (c) Ground truth of River. (d) Hermiston on May 1, 2004. (e) Hermiston on May 8, 2007. (f) Ground truth of Hermiston.

3.4. Binary decision making for HCD

In the final stage, the extracted feature is flattened and fed into a fully connected layer. This layer is responsible for the binary decision-making process, interpreting the refined features and producing the final binary classification output. By leveraging the comprehensive feature representation obtained from the preceding stages, this stage aims to make accurate and reliable predictions.

4. Experiments

4.1. Dataset description

All two datasets used in our experiment were acquired by the Hyperion sensor mounted onboard the Earth Observing-1 (EO-1) satellite, which offers up to 242 spectral bands ranging from 0.4 to 2.5 μm, with a spatial resolution of 10 m and a spectral resolution of 30 nm. All the images are spatially aligned in pairs along with the noise removal, as seen in Fig. 3. As shown in Fig. 3 (a-c), the River dataset contains 463 $\times$ 241 spatial pixels and 198 spectral bands, where the major changes are the substance in the river and the structure of the riverbank. For the Hermiston dataset given in Fig. 3 (d-f), it has 307 $\times$ 241 pixels in 154 spectral bands, where the major changes are crop growth situation and the water content of crops being affected by irrigation conditions in the farmland.

4.2. Evaluation criteria

As change detection is a binary classification task, the overall accuracy (OA), average accuracy (AA), and the Kappa coefficient (KP) were used for quantitative performance assessment. OA is the percentage of correctly classified pixels, which is defined by:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

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.

KP is to measure the inter-rater reliability that represents the degree of similarity between the change map and the ground truth defined as follows.

$$KP = \frac{OA - PRE}{1 - PRE} \tag{13}$$

$$PRE = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + TN + FP + FN)^2} \tag{14}$$

Precision (Pre) describes the proportion of detected changed pixels that are actually changed and is defined as:

$$Pre = \frac{TP}{TP + FP} \tag{15}$$

Recall (Re) represents the proportion of actual changed pixels that have been correctly detected.

$$Re = \frac{TP}{TP + FN} \tag{16}$$

F1 score ($F1$) defines a balanced index that can be considered as harmonic mean of Pre and Re.

$$F1 = 2 * \frac{Pre * Re}{Pre + Re} \tag{17}$$

4.3. Experimental setting

In this section, we evaluate the effectiveness of the proposed method by comparing it with three start-of-the-art unsupervised methods, which include the change vector analysis (CVA) [16], principal component analysis (PCAKM) [19] and accumulated band-wise binary distancing (ABBD) [61] as well as five DL-based methods such as ML-EDAN [62], HyGSTAN [63], CASNet [64], CBANet [65] and SSTFormer [53], which represent either classical or the state-of-the-art techniques in HCD, as briefed below.

- **CVA** [16]: This method involves computing the Euclidean distance between spectral vectors of bi-temporal images. OTSU thresholding is then applied to generate a binary change map.
- **PCAKM** [19]: Principal Component Analysis is used to reduce data dimensionality and redundancy, k-means is subsequently applied to classify pixels into changed or unchanged categories.
- **ABBD** [61]: This method evaluates binary distance for each band to indicate pixel changes. The band-wise binary distance maps are accumulated into a grayscale change map, which is then processed using k-means clustering for binary classification.

- **ML-EDAN** [62]: A two-stream encoder-decoder model that integrates hierarchical features from convolutional layers of bi-temporal images. It employs a context-guided attention module for enhanced feature transfer and an LSTM subnetwork to capture temporal dependencies.
- **HyGSTAN** [63]: This self-attention-based network uses cosine similarity to reduce spectral redundancy. It employs gated spectral-spatial attention with single-head weak self-attention for spatial feature extraction and gated spectral-spatial-temporal attention for capturing temporal changes.
- **CASNet** [64]: A traditional self-attention-based method that enhances joint spatial, spectral, and temporal feature representation within each band.
- **CBANet** [65]: A 2-D self-attention-based method that combined with a cross-band feature extraction module, enhancing feature representation and discrimination.
- **SSTFormer** [53]: This method employs a joint spectral, spatial, and temporal transformer to integrate and extract relevant change features using self-attention mechanisms.

The benchmarks are established according to the specified parameters in the default settings, where DL-based methods are trained using PyTorch on an 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, KP, Pre, Re,* and *F1* are reported for comparison. In addition, in the resulting change maps comparison, false alarms (*FP*) and missing pixels (*FN*) are highlighted in red and green, respectively, while correctly detected changed pixels (*TP*) are put in white, and true negatives (*TN*) are depicted in black for ease of visual comparison.

4.4. Comparison experiments

1) *Results analysis for the River Dataset*: The quantitative assessment and extracted change maps on the River dataset for all benchmarks are shown in Table 2 and Fig. 4, respectively. Firstly, from the comparison of extracted change maps, the binary maps produced by these three unsupervised algorithms contain a significant number of false alarms or missing pixels. In contrast, all DL-based algorithms detect only a few false alarms or missing pixels. This observation is also supported by quantitative analysis that the *Pre* of these three unsupervised algorithms does not exceed 0.74, while all DL-based algorithms achieve a *Pre* of over 90 %. Similarly, the *F1* for all unsupervised algorithms are below 0.82, whereas the *F1* for all DL-based algorithms are all above 0.90. However, due to the CVA method misclassifying many unchanged pixels as changed category, it results in very few missing pixels (FN). Consequently, CVA has the highest *Re* among all benchmarks, but the lowest *F1* score, due to the highest number of false alarms (*FP*). Among all DL-based methods, ML-EDAN achieved the worst results, with an average KP of only 0.8009 and an average *F1* of 0.9016. HyGSTAN, CSANet and SSTFormer have *KP* values ranging from 0.81 to 0.84 and F1 between 0.90 and 0.92. CBANet consistently ranks second across all benchmarks, with average *KP* and *F1* of 0.8526 and 0.9296. respectively. The proposed method outperforms all others, achieving the highest scores across *OA, KP, Pre* and *F1*. Notably, it shows an improvement of approximately 0.0136 in *KP* and 0.0156 in *F1* compared to the second-best method, CBANet. Additionally, the proposed method has the highest *Pre* among all benchmarks, indicating a high accuracy in detecting *TP* with only a few false alarms. And the highest *Re* among all benchmarks, except CVA. These quantitative analysis results demonstrate the proposed method's superior capability in accurately detecting changes with minimal false alarms and missing pixels, confirming its effectiveness and reliability in HCD task.
2) *Results analysis for the Hermiston Dataset*: The extracted change maps and quantitative analysis results for all benchmarks on the Hermiston dataset are presented in Fig. 5 and Table 3, respectively. Similar to the findings on the River dataset, CVA and PCAKM exhibit a significant number of missing pixels (*FN*) in their binary maps and detect only a few false alarms (*FP*). This leads to very high *Pre* values, both exceeding 99.5 %, but their *F1* are only 0.8127 and 0.7924, respectively. Consequently, they both achieve very high precision (*Pre*) values, exceeding 99.5 %, but their *F1* scores are significantly lower at 0.8127 and 0.7924, respectively, making them the lowest among all benchmarks. ABBD stands out among the unsupervised algorithms for its superior detection accuracy, even surpassing ML-EDAN, HyGSTAN, and CSANet in the *KP*. Among all DL-based methods, HyGSTAN performs the worst, with a higher number of missing pixels compared to other DL-based algorithms, as illustrated in Fig. 5(e). CBANet and

Table 2
Quantitative assessment of different methods on the river dataset (standard deviation is only included for the supervised methods).

	OA(%)	***KP***	***Pre***	***Re***	***F***1
CVA	92.53	0.6528	0.5393	**0.9635**	0.6915
PCAKM	95.17	0.7478	0.6524	0.9506	0.7738
ABBD	96.38	0.7928	0.7377	0.9045	0.8126
ML-EDAN	96.96±0.0014	0.8009±0.0049	0.9078±0.0012	0.8961±0.0088	0.9016±0.0013
HyGSTAN	97.22±0.0012	0.8202±0.0048	0.9211±0.0109	0.9002±0.0075	0.9101±0.0024
CSANet	97.43±0.0012	0.8360±0.0049	0.9296±0.0022	0.9081±0.0030	0.9185±0.0012
CBANet	97.65±0.0036	0.8526±0.0036	0.9346±0.0105	0.9235±0.0065	0.9262±0.0008
SSTFormer	97.46±0.0015	0.8383±0.0023	0.9266±0.0216	0.9137±0.0187	0.9191±0.0011
Proposed	**97.91±0.0004**	**0.8662±0.0026**	**0.9402±0.0027**	0.9261±0.0038	**0.9330±0.0014**

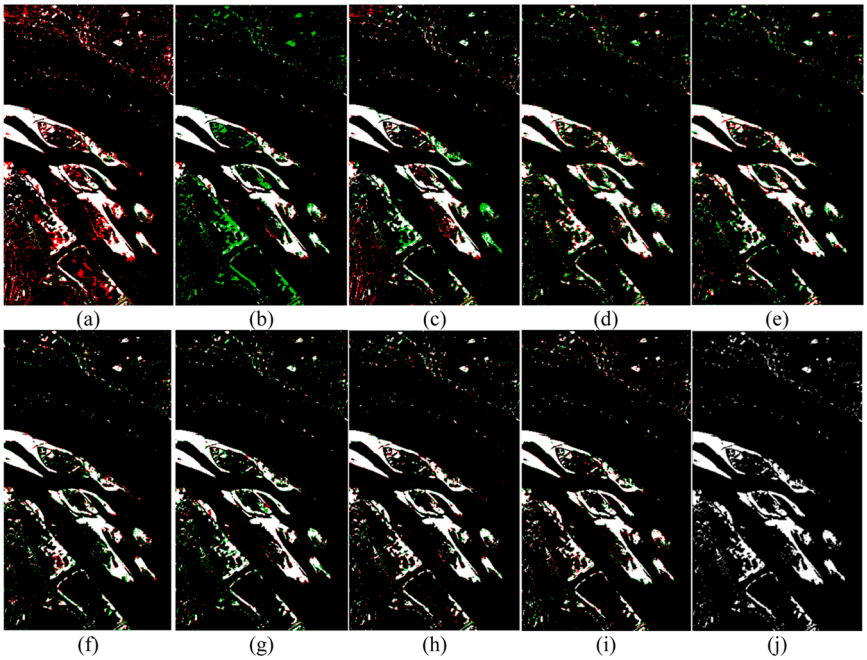

Fig. 4. Extracted change maps on the River Dataset from different methods of CVA (a), PCAKM (b), ABBD (c), ML-EDAN (d), HyGSTAN (e), CSANet (f), CBANet (g), SSTFormer (h) and proposed (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, respectively, whist white and black denotes true positive and true negative, respectively.

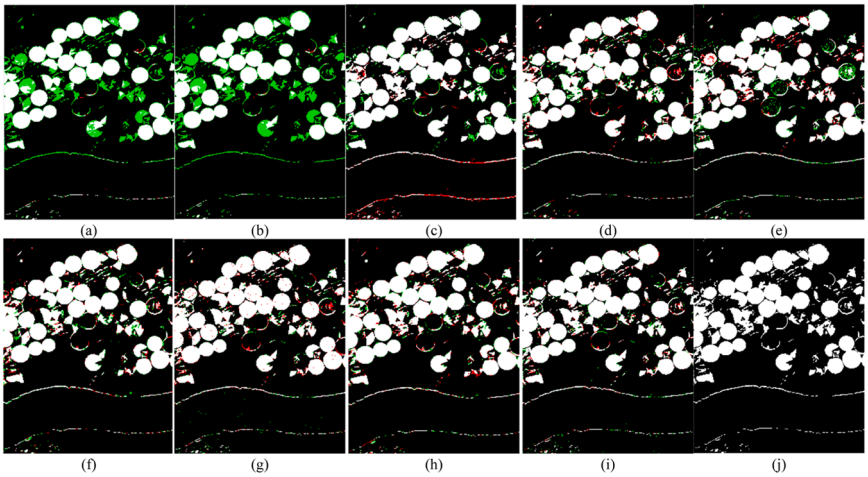

Fig. 5. Extracted change maps on the Hermiston Dataset from different methods of CVA (a), PCAKM (b), ABBD (c), ML-EDAN (d), HyGSTAN (e), CSANet (f), CBANet (g), SSTFormer (h) and proposed (i) in comparison to the Ground-truth map (j), where the false alarms and missing pixels are labelled in red and green, respectively, whist white and black denotes true positive and true negative, respectively.

SSTFormer demonstrate similar detection capabilities, both achieving average *KP* and *F1* scores around 0.92 and 0.96. However, our proposed method significantly outperforms these algorithms, achieving an *OA* of over 98 %. The average *KP* and *F1* scores of the Proposed method surpass those of the second-best algorithm by 0.0141 and 0.0071, respectively. Again, these results provide compelling evidence of the superior performance and robustness of our proposed method. Furthermore, the standard deviations for the five evaluation metrics in the quantitative analysis are all below 0.0017, the smallest among all benchmarks, indicating the stability of our proposed method.

Table 3
Quantitative assessment of different methods on the hermiston dataset (standard deviation is only included for the supervised methods).

	OA(%)	***KP***	***Pre***	***Re***	***F***1
CVA	92.87	0.7705	0.9953	0.6867	0.8127
PCAKM	92.24	0.7472	**0.9973**	0.6574	0.7924
ABBD	97.49	0.9281	0.9434	0.9453	0.9443
ML-EDAN	97.05±0.0020	0.9151±0.0058	0.9595±0.0040	0.9559±0.0054	0.9576±0.0029
HyGSTAN	96.74±0.0009	0.9054±0.0024	0.9598±0.0029	0.9461±0.0018	0.9527±0.0012
CSANet	96.87±0.0005	0.9097±0.0016	0.9585±0.0024	0.9514±0.0030	0.9596±0.0076
CBANet	97.50±0.0021	0.9285±0.0051	0.9648±0.0084	0.9640±0.0039	0.9642±0.0025
SSTFormer	97.34±0.0026	0.9234±0.0075	0.9663±0.0062	0.9563±0.0044	0.9617±0.0037
Proposed	**98.00±0.0003**	**0.9426±0.0008**	0.9725±0.0016	**0.9702±0.0012**	**0.9713±0.0004**

4.5. Ablation study

1) ***Model Efficiency***: Table 4 provides a comprehensive comparison of the number of parameters, FLOPs, and the training time of single epoch for all DL-based benchmarks on the River dataset, which contains 198 bands, with a batch size set to 1. From Table 3, it is evident that our proposed algorithm ranks third in terms of parameter count, it boasts a significantly lower parameter count compared to ML-EDAN and SSTFormer. This reduction in parameters translates to lower computational resource requirements. Moreover, despite the lower parameter count, our method does not compromise on performance. It demonstrates outstanding detection capabilities, achieving high accuracy and robustness. This exceptional detection performance, coupled with the reduced computational load, highlights the strength of our approach. It strikes a perfect balance between efficiency and effectiveness, making it an optimal choice for scenarios where both resource constraints and high performance are critical.
2) ***Patch size:*** Experiments were conducted using five patch sizes: {5 × 5, 7 × 7, 9 × 9, 11 × 11, 13 × 13}, the variations of *KP* are illustrated in Fig. 6(a). It can be observed that the average of *KP* for test accuracy initially increased and then declined, with both metrics peaking at a patch size of 7 × 7. Moreover, it is evident that as the patch size increases, the training time progressively lengthens. Observing these trends, it becomes clear that the patch size of 7 × 7 can balances performance and efficiency.
3) ***Kernel Number of Feature Extracted Module***: The kernel number of the feature extraction module determines the number of channels in the feature map F^3, which subsequently serves as the input for the Mamba block. To investigate the impact of different kernel numbers, experiments were conducted with the following values: {8, 16, 32, 64, 128}. The *KP* curves of two datasets are depicted in Fig. 6(b) reveals that setting the kernel number to 64 yields the optimal performance across both datasets.
4) ***Number of Mamba$_{out}$ channels***: The number of *Mamba*$_{out}$ channels determine the number of output channels in the Mamba block, which subsequently affects the size of the flattened dimensions in the subsequent fully connected layer. To investigate the impact of different *Mamba*$_{out}$ channel numbers, experiments were conducted using the following values: {8, 16, 32, 64, 128}. The *KP* curves for two datasets are depicted in the Fig. 6(c). The results reveal that the *KP* values for the Hermiston dataset remain consistently high, ranging from 0.94 to 0.95, regardless of the number of *Mamba*$_{out}$ channels. This indicates that the Hermiston dataset maintains high performance with minimal impact from varying the number of *Mamba*$_{out}$channels. In contrast, the *KP* values for the River dataset are ranging from 0.85 to 0.87. However, within this range, the optimal performance is observed when the number of *Mamba*$_{out}$ channels is set to 64. Thus, for both datasets, a *Mamba*$_{out}$ channel number of 64 is recommended to achieve the best model performance.
5) ***Key Stage Analysis***: This section primarily analyzes the role of each module within the model. By separately removing the three main modules (spectral attention module, spatial attention module, and feature extraction module), we evaluate their impact on the model's performance. Fig. 7 presents the quantitative assessment results on the River and Hermiston datasets. Although removing certain modules can reduce the model's parameters and FLOPs, the overall performance analysis indicates that removing any module negatively affects the model's detection accuracy. The feature extraction module has the most significant impact on both datasets, particularly on the Hermiston dataset, where its removal leads to a substantial performance drop. Removing the spatial attention module also results in notable declines across various metrics, especially the *KP* value, highlighting the importance of spatial attention in enhancing performance on both datasets. The spectral attention module, when removed, causes slight decreases in performance metrics, particularly for the Hermiston dataset, indicating its helpful role in improving model performance.

Table 4
Parameters of different DL-based methods.

	Parameters (M)	**FLOPs** (M)	**Training Time** (s/per epoch)
ML-EDAN	88.53	569.94	131.36
HyGSTAN	0.03	0.83	2.26
CSANet	2.45	144.33	29.53
CBANet	0.45	11.05	18.49
SSTFormer	2.53	443.28	44.19
Proposed	1.63	136.71	24.95

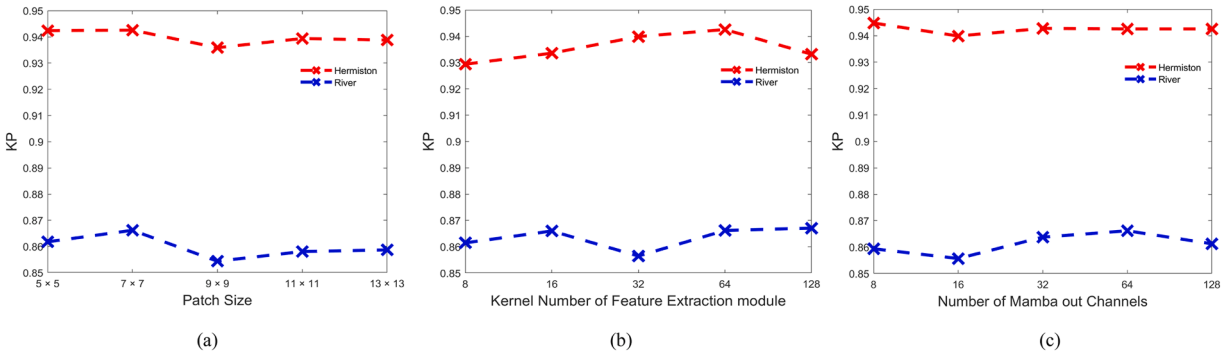

Fig. 6. Experiments and results of proposed method under different parameter settings on two datasets, including (a) results with different patch size, (b) results with different kernel number of feature extraction module, (c) results with different number of Mamba out channels.

The proposed combination of these modules achieves optimal performance, ensuring that the model performs exceptionally well across different datasets.

4.6. *Further discussion*

The proposed method demonstrates significant advantage in terms of higher detection accuracy compared to state-of-the art benchmarks across two datasets. Firstly, traditional attention mechanisms typically extract information within a local scope, whereas the global attention mechanism can fuse information over the whole image. This global perspective ensures that even subtle changes are precisely detected, significantly reducing false alarms and missed detections. By comprehensively considering spatial and spectral features, our method achieves a notable improvement in detection accuracy compared to other methods.

The comparison with unsupervised algorithms, such as CVA, PCAKM and ABBD, while computationally efficient and relatively simple to implement, often struggle with false alarms or missing pixels, leading to lower *Pre* and *F1* score, this is primarily dure to their limited capability to exploit spatial context and noise sensitively. PCAKM and ABBD, although better than CVA in certain metrics, still fall short in comparison to DL-based methods. These findings reinforce the limitations of traditional unsupervised methods in handling the complexity of change detection tasks, particularly in diverse and intricate datasets like River and Hermiston. In contrast, DL-based methods exhibit remarkable capabilities in extracting and integrating relevant features, achieving higher *Pre, Re* and *F1* scores. ML-EDAN integrates hierarchical features from convolutional layers and employs a context-guided attention module and an LSTM sub-network for temporal dependencies. Despite its advanced architecture, ML-EDAN's performance is hindered by the complexity of the model, which may lead to overfitting and increased computational burden. In addition, the integration of hierarchical features may not fully capture the global spatial and spectral context. Although HyGSTAN has minimal parameters, its single-head weak self-attention mechanism seems insufficient for complex change detection tasks that require capturing global dependencies. CASNet enhances joint spatial, spectral, and temporal feature representation using a traditional self-attention method. While it effectively captures detailed features within each band, CASNet may struggle with integrating these features into a coherent global context, resulting in lower accuracy. CBANet enhances feature representation and discrimination by introducing 2-D self-attention. Despite its strong performance, CBANet may not fully leverage long-range dependencies relationships, which are crucial for precise change detection. Transformer-based SSTFormer network also leverages extensive linear transformations and self-attention mechanisms, making it powerful in capturing complex features, it only extracts static features and does not account for long-term feature dependencies. While its detection accuracy is somewhat better than other benchmarks, both the model's efficiency and detection accuracy fall short compared to our proposed method, which is based on dynamic feature extraction.

In summary, the superior performance of our proposed GASSM is attributed to its advanced feature extraction and state space model, which effectively captures dynamic, spatial, and spectral characteristics. This comprehensive approach ensures high accuracy in distinguishing the characteristics of changes, making our method a reliable and efficient solution for HCD tasks.

5. Conclusion

In this paper, a novel end-to-end DL-based network, GASSM, is proposed for HCD task. The detection accuracy of GASSM surpasses existing state-ot-the-art benchmarks and demonstrates robustness across various testing scenarios. Its innovation architecture includes a global attention and a dedicated feature extraction module, effectively enhancing the spatial-spectral feature interactions and reducing feature redundancy. The Mamba block incorporates a state space model to adeptly capture long-range dependencies, highlighting the network's ability to comprehensively model global spatial-spectral features. Additionally, GASSM has lower computational costs compared to most deep learning methods, particularly those based on transformers.

Due to the necessity of manually labeled training sets, our GASSM model still faces a common limitation inherent to all supervised learning algorithms. To address this challenge, our next step is to leverage the powerful feature representation capabilities of GASSM by utilizing it as the backbone and incorporating pseudo ground truth generated by unsupervised algorithms. Specifically, we will

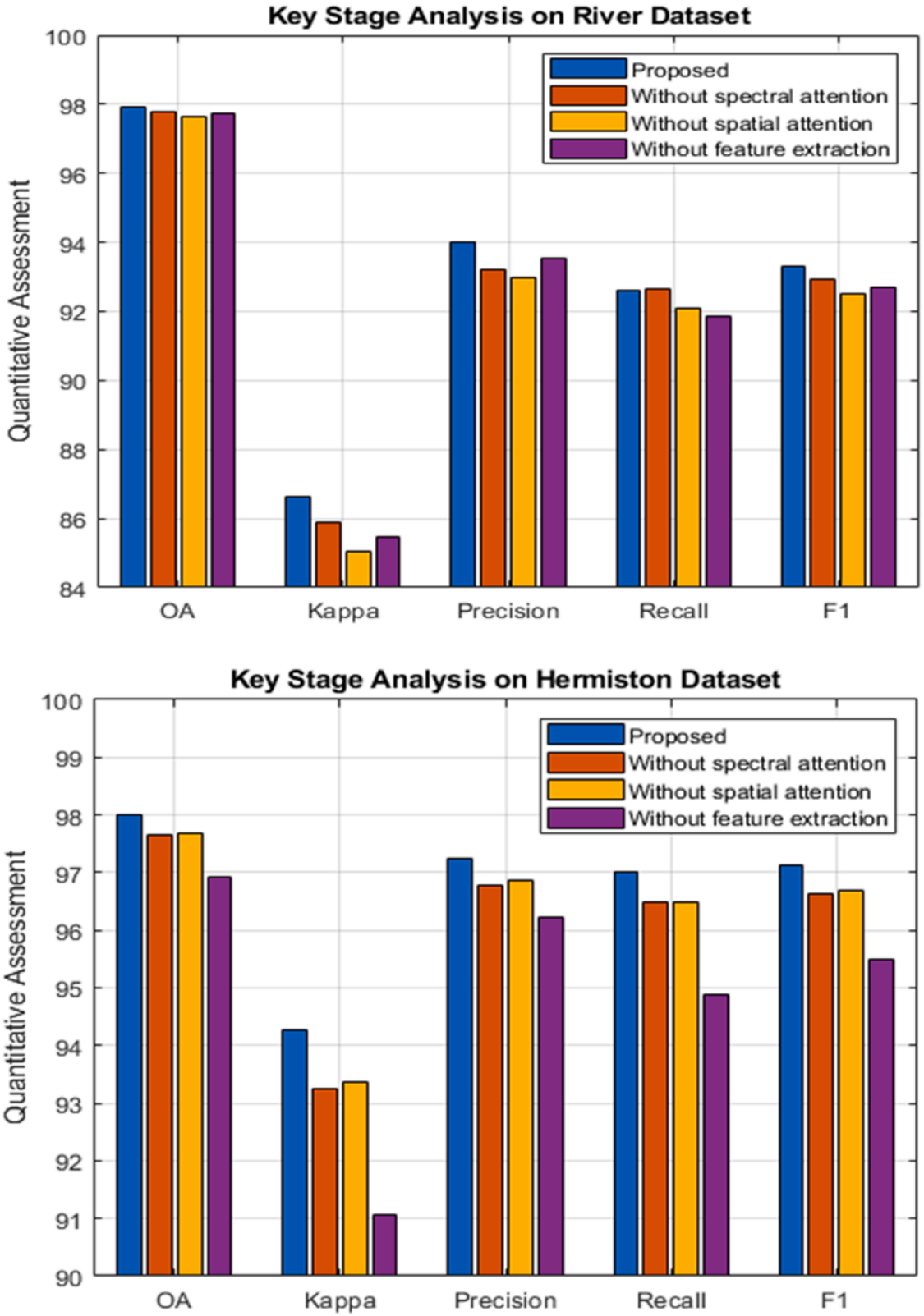

Fig. 7. Key stage analysis of proposed network on the River and Hermiston datasets.

experiment with different combinations of unsupervised algorithms and enhance the accuracy of the pseudo ground truth through a majority voting mechanism. Our strategy is to implement a self-supervised learning approach that does not require any manually annotated training samples, using new techniques such as cross-modal change detection [66], semi-supervised learning [67], channel-wise attention [68] and quality assessment in wavelet-domain [69].

CRediT authorship contribution statement

Yinhe Li: Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Jinchang Ren:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Hang Fu:** Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Genyun Sun:** Writing – review & editing, Visualization, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Liu, D. Marinelli, L. Bruzzone, F. Bovolo, A review of change detection in multitemporal hyperspectral images: current techniques, applications, and challenges, IEEe Geosci. Remote Sens. Mag. 7 (2) (2019) 140–158.
[2] Y. Yan, J. Ren, Q. Liu, H. Zhao, H. Sun, J. Zabalza, PCA-domain fused singular spectral analysis for fast and noise-robust spectral–spatial feature mining in hyperspectral classification, IEEE Geosci. Remote Sens. Lett. 20 (2021) 1–5.
[3] Y. Yan, J. Ren, H. Sun, R. Williams, Nondestructive quantitative measurement for precision quality control in additive manufacturing using hyperspectral imagery and machine learning, IEEe Trans. Industr. Inform. (2024).
[4] H. Fu, Z. Ling, G. Sun, J. Ren, A. Zhang, L. Zhang, X. Jia, HyperDehazing: a hyperspectral image dehazing benchmark dataset and a deep learning model for haze removal, ISPRS J. Photogramm. Remote Sens. 218 (2024) 663–677.
[5] G. Sun, X. Zhang, X. Jia, J. Ren, A. Zhang, Y. Yao, H. Zhao, Deep fusion of localized spectral features and multi-scale spatial features for effective classification of hyperspectral images, Int. J. Appl. Earth Observ. Geoinf. 91 (2020) 102157.
[6] A. Ertürk, A. Plaza, Informative change detection by unmixing for hyperspectral images, IEEE Geosci. Remote Sens. Lett. 12 (6) (2015) 1252–1256.
[7] B. Zhang, D. Dai, J. Huang, J. Zhou, Q. Gui, F. Dai, Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: a review, Crit. Rev. Food Sci. Nutr. 58 (12) (2018) 2099–2118.
[8] N. Farmonov, K. Amankulova, J. Szatmári, A. Sharifi, D. Abbasi-Moghadam, S.M.M. Nejad, L. Mucsi, Crop type classification by DESIS hyperspectral imagery and machine learning algorithms, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 16 (2023) 1576–1588.
[9] X. Wang, S. Wang, Y. Ding, Y. Li, W. Wu, Y. Rong, W. Kong, J. Huang, S. Li and H. Yang, "State space model for new-generation network alternative to transformers: a survey," *arXiv Preprint arXiv:2404.09516,* 2024.
[10] A. Gu and T. Dao, "Mamba: linear-time sequence modeling with selective state spaces," *arXiv Preprint arXiv:2312.00752,* 2023.
[11] X. Ma, X. Zhang, M. Pun, RS 3 Mamba: visual state space model for remote sensing image semantic segmentation, IEEE Geosci. Remote Sens. Lett. (2024).
[12] K. Chen, B. Chen, C. Liu, W. Li, Z. Zou, Z. Shi, Rsmamba: remote sensing image classification with state space model, IEEE Geosci. Remote Sens. Lett. (2024).
[13] H. Fu, G. Sun, Y. Li, J. Ren, A. Zhang, C. Jing and P. Ghamisi, "HDMba: hyperspectral remote sensing imagery dehazing with state space model," *arXiv Preprint arXiv:2406.05700,* 2024.
[14] P. Ma, J. Ren, Z. Gao, Y. Li, R. Chen, Hyperspectral imagery quality assessment and band reconstruction using the prophet model, CAAI. Trans. Intell. Technol. (2024).
[15] P. Du, S. Liu, P. Gamba, K. Tan, J. Xia, Fusion of difference images for change detection over urban areas, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 5 (4) (2012) 1076–1086.
[16] F. Bovolo, L. Bruzzone, A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain, IEEe Trans. Geosci. Remote Sens. 45 (1) (2006) 218–236.
[17] F. Bovolo, S. Marchesi, L. Bruzzone, A framework for automatic and unsupervised detection of multiple changes in multitemporal images, IEEe Trans. Geosci. Remote Sens. 50 (6) (2011) 2196–2212.
[18] S. Kuching, The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis, J. Comput. Sci. 3 (6) (2007) 419–423.
[19] T. Celik, Unsupervised change detection in satellite images using principal component analysis and $ k $-means clustering, IEEE Geosci. Remote Sens. Lett. 6 (4) (2009) 772–776.
[20] A.A. Nielsen, K. Conradsen, J.J. Simpson, Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: new approaches to change detection studies, Remote Sens. Environ. 64 (1) (1998) 1–19.
[21] A.A. Nielsen, The regularized iteratively reweighted MAD method for change detection in multi-and hyperspectral data, IEEe Trans. Image Process. 16 (2) (2007) 463–478.
[22] C. Wu, B. Du, L. Zhang, Slow feature analysis for hyperspectral change detection. 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2014, pp. 1–4.
[23] V. Wargnier-Dauchelle, T. Grenier, F. Durand-Dubief, F. Cotton, M. Sdika, Constrained non-negative networks for a more explainable and interpretable classification. Medical Imaging with Deep Learning, 2024.
[24] J. Shi, Z. Zhang, C. Tan, X. Liu, Y. Lei, Unsupervised multiple change detection in remote sensing images via generative representation learning network, IEEE Geosci. Remote Sens. Lett. 19 (2021) 1–5.
[25] L. Bruzzone, D.F. Prieto, Automatic analysis of the difference image for unsupervised change detection, IEEe Trans. Geosci. Remote Sens. 38 (3) (2000) 1171–1182.
[26] H. Chen, S. Ye, D. Zhang, L. Areshkina, S. Ablameyko, Change detection based on difference image and energy moments in remote sensing image monitoring, Pattern Recognit. Image Anal. 28 (2018) 273–281.
[27] S. Liu, Q. Du, X. Tong, A. Samat, H. Pan, X. Ma, Band selection-based dimensionality reduction for change detection in multi-temporal hyperspectral images, Remote Sens. 9 (10) (2017) 1008.
[28] Z. Lv, M. Zhang, W. Sun, J.A. Benediktsson, T. Lei, N. Falco, Spatial-contextual information utilization framework for land cover change detection with hyperspectral remote sensed images, IEEe Trans. Geosci. Remote Sens. 61 (2023) 1–11.
[29] A. Ertürk, M. Iordache, A. Plaza, Sparse unmixing-based change detection for multitemporal hyperspectral images, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 9 (2) (2015) 708–719.

[30] Q. Guo, J. Zhang, Y. Zhang, Multitemporal hyperspectral images change detection based on joint unmixing and information coguidance strategy, IEEe Trans. Geosci. Remote Sens. 59 (11) (2021) 9633–9645.
[31] Z. Hou, W. Li, R. Tao, Q. Du, Three-order tucker decomposition and reconstruction detector for unsupervised hyperspectral change detection, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 14 (2021) 6194–6205.
[32] D. Marinelli, F. Bovolo, L. Bruzzone, A novel change detection method for multitemporal hyperspectral images based on binary hyperspectral change vectors, IEEe Trans. Geosci. Remote Sens. 57 (7) (2019) 4913–4928.
[33] Y. Sun, L. Lei, D. Guan, G. Kuang, L. Liu, Graph signal processing for heterogeneous change detection, IEEe Trans. Geosci. Remote Sens. 60 (2022) 1–23.
[34] S.T. Seydi, M. Hasanlou, A new structure for binary and multiple hyperspectral change detection based on spectral unmixing and convolutional neural network, Measurement 186 (2021) 110137.
[35] A. Shafique, S.T. Seydi, T. Alipour-Fard, G. Cao, D. Yang, SSViT-HCD: a spatial–spectral convolutional vision transformer for hyperspectral change detection, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 16 (2023) 6487–6504.
[36] H.F. Tolie, J. Ren, E. Elyan, DICAM: deep inception and channel-wise attention modules for underwater image enhancement, Neurocomputing. 584 (2024) 127585.
[37] S. Suthaharan, S. Suthaharan, Support vector machine. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, 2016, pp. 207–235.
[38] S.J. Rigatti, Random forest, J. Insur. Med. 47 (1) (2017) 31–39.
[39] X. Yin, C. Yang, W. Pei, H. Hao, Shallow classification or deep learning: an experimental study, in: 2014 22nd International Conference on Pattern Recognition, 2014, pp. 1904–1909.
[40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[41] V. Sze, Y. Chen, T. Yang, J.S. Emer, Efficient processing of deep neural networks: a tutorial and survey, Proc. IEEE 105 (12) (2017) 2295–2329.
[42] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, Recent advances in convolutional neural networks, Pattern. Recognit. 77 (2018) 354–377.
[43] L. Medsker, L.C. Jain, Recurrent Neural Networks: Design and Applications, CRC press, 1999.
[44] T. Zhan, B. Song, Y. Xu, M. Wan, X. Wang, G. Yang, Z. Wu, SSCNN-S: a spectral-spatial convolution neural network with Siamese architecture for change detection, Remote Sens. 13 (5) (2021) 895.
[45] Q. Wang, Z. Yuan, Q. Du, X. Li, GETNET: a general end-to-end 2-D CNN framework for hyperspectral image change detection, IEEe Trans. Geosci. Remote Sens. 57 (1) (2018) 3–13.
[46] Y. Lin, S. Li, L. Fang, P. Ghamisi, Multispectral change detection with bilinear convolutional neural networks, IEEE Geosci. Remote Sens. Lett. 17 (10) (2019) 1757–1761.
[47] A. Song, J. Choi, Y. Han, Y. Kim, Change detection in hyperspectral images using recurrent 3D fully convolutional networks, Remote Sens. 10 (11) (2018) 1827.
[48] C. Shi, Z. Zhang, W. Zhang, C. Zhang, Q. Xu, Learning multiscale temporal–spatial–spectral features via a multipath convolutional LSTM neural network for change detection with hyperspectral images, IEEe Trans. Geosci. Remote Sens. 60 (2022) 1–16.
[49] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, J. Ren, Multiscale diff-changed feature fusion network for hyperspectral image change detection, IEEe Trans. Geosci. Remote Sens. 61 (2023) 1–13.
[50] G. Xie, J. Ren, S. Marshall, H. Zhao, R. Li, R. Chen, Self-attention enhanced deep residual network for spatial image steganalysis, Digit. Signal Process. (2023) 104063.
[51] Y. Yang, J. Qu, S. Xiao, W. Dong, Y. Li, Q. Du, A deep multiscale pyramid network enhanced with spatial–spectral residual attention for hyperspectral image change detection, IEEe Trans. Geosci. Remote Sens. 60 (2022) 1–13.
[52] R. Song, W. Ni, W. Cheng, X. Wang, CSANet: cross-temporal interaction symmetric attention network for hyperspectral image change detection, IEEE Geosci. Remote Sens. Lett. (2022).
[53] Y. Wang, D. Hong, J. Sha, L. Gao, L. Liu, Y. Zhang, X. Rong, Spectral–spatial–temporal transformers for hyperspectral image change detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.
[54] H. Yu, H. Yang, L. Gao, J. Hu, A. Plaza, B. Zhang, Hyperspectral Image Change Detection Based on Gated Spectral–Spatial–Temporal Attention Network with Spectral Similarity Filtering, IEEe Trans. Geosci. Remote Sens. (2024).
[55] Y. Ji, W. Sun, Y. Wang, Z. Lv, G. Yang, Y. Zhan, C. Li, Domain adaptive and interactive differential attention network for remote sensing image change detection, IEEE Trans. Geosci. Remote Sens. (2024).
[56] W. Sun, Q. Du, Hyperspectral band selection: a review, IEEe Geosci. Remote Sens. Mag. 7 (2) (2019) 118–139.
[57] Y. Liu, Z. Shao and N. Hoffmann, "Global attention mechanism: retain information to enhance channel-spatial interactions," *arXiv Preprint arXiv:2112.05561,* 2021.
[58] S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, M. Gabbouj, 1-D convolutional neural networks for signal processing applications, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8360–8364.
[59] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, HybridSN: exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 17 (2) (2019) 277–281.
[60] S. Elfwing, E. Uchibe, K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, Neural Networks 107 (2018) 3–11.
[61] Y. Li, J. Ren, Y. Yan, P. Ma, M. Assaad, Z. Gao, ABBD: accumulated band-wise binary distancing for unsupervised parameter-free hyperspectral change detection, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. (2024).
[62] J. Qu, S. Hou, W. Dong, Y. Li, W. Xie, A multilevel encoder–decoder attention network for change detection in hyperspectral images, IEEe Trans. Geosci. Remote Sens. 60 (2021) 1–13.
[63] H. Yu, H. Yang, L. Gao, J. Hu, A. Plaza, B. Zhang, Hyperspectral image change detection based on gated spectral–spatial–temporal attention network with spectral similarity filtering, IEEe Trans. Geosci. Remote Sens. (2024).
[64] R. Song, W. Ni, W. Cheng, X. Wang, CSANet: cross-temporal interaction symmetric attention network for hyperspectral image change detection, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.
[65] Y. Li, J. Ren, Y. Yan, Q. Liu, P. Ma, A. Petrovski, H. Sun, CBANet: an end-to-end cross band 2-D attention network for hyperspectral change detection in remote sensing, IEEe Trans. Geosci. Remote Sens. (2023).
[66] E. Zhang, et al., ICSF: integrating inter-modal and cross-modal learning framework for self-supervised heterogeneous change detection, IEEE Trans. Geoscience and Remote Sensing 63 (2024) 5501516.
[67] Z. Fang, et al., Dual teacher: improving the reliability of pseudo labels for semi-supervised oriented object detection, IEEE Trans. Geoscience and Remote Sensing 63 (2024) 5602515.
[68] H.F. Tolie, et al., DICAM: Deep inception and channel-wise attention modules for underwater image enhancement, Neurocomputing 584 (2024) 127585.
[69] H.F. Tolie, et al., Blind sonar image quality assessment via machine learning: Leveraging micro-and macro-scale texture and contour features in the wavelet domain, Engineering Applications of Artificial Intelligence 141 (2025) 109730.