SSTGCNs: Spectral-Spatial-Temporal Long-Range Dependencies Joint Feature Extraction with Graph Convolutional Networks for Adaptive Change Detection

Zhanyuan Chang,Yuwen Wei,Jie Lian*,Mingxiao Jin,Dong Wang,Xuyang Li

***Abstract*—Hyperspectral image change detection plays an important role in ground observation tasks, making full use of the rich spectral and spatial information in the bitemporal hyperspectral to identify subtle changes in the surface. Currently, most methods are extracting spatial -spectral features, ignoring the interaction between bitemporal images. In addition, there is a long-distance dependence between the spectrum and pixels of hyperspectral images. How to capture the dependence between long-distance bands and pixels is a problem that needs to be solved at present. To solve the above problems, this paper proposes SSTGCNs: spectral-spatial-temporal long-range dependencies joint feature extraction with graph convolutional networks. First, the network uses Spectral-Spatial-Temporal Feature Extraction Module (SSTFEM) to capture spectral-spatial-temporal features and increase the interactive features of bi-time phase images. Secondly, a long-distance dependency capture module is designed to capture the long-distance dependence between bands and pixels through the Long-Range Spectral-Spatial Dependency Module (LRSSDM). Finally, through Redundancy Suppression and Adaptive Feature Fusion (RSAF), redundant features are removed and high-information features are fused to improve the model's feature expression ability and generalization ability. To verify the effectiveness of the model, we conducted experiments on three public hyperspectral CD datasets. The experimental results showed that the OA and Kappa indicators of our model on the USA data were improved by 0.55% and 1.44% respectively compared with the latest methods. On the China dataset, the proposed method compared the latest AIWSEN, OA and Kappa to obtain suboptimal results, while the P, R and F1 scores achieved the optimal results, which was 2.59%, 0.89% and 1.14% higher than the state-of-the-art method respectively. The OA, P, R, F1 indicators on the River dataset achieve optimal results: 0.36%, 1.32%, 0.5% and 0.91% improvements respectively over the state-of-the-art methods. Explain that our change detection method works better for spectral data containing more information.**

***Index Terms*—spectral-spatial-temporal features,change detection, long-range dependencies, adaptively fused features.**

I. Introduction

With the rapid advancement of spectral imaging technology, hyperspectral imagery (HSI) can capture not only detailed spatial information of ground objects but also continuous spectral signatures across different materials [1]. Due to the rich surface information it provides, HSI has been widely applied in remote sensing research. Hyperspectral change detection (HSI-CD) is a popular Earth observation technique aimed at identifying differences in multi-temporal HSI acquired over the same geographic location. HSI-CD enables accurate identification and comprehensive analysis of surface cover changes, making it a core technology in various applications such as environmental monitoring, urban planning, and agricultural assessment [2].

In HSI-CD tasks, traditional deep learning methods primarily rely on convolutional neural networks (CNNs), with model parameters optimized via gradient descent algorithms [3]. However, the limited receptive field of CNNs imposes significant constraints on their ability to infer global contextual information, making it difficult to fully capture complex long-range dependent features in hyperspectral data [4]. Subsequent studies have demonstrated that graph convolutional networks (GCNs) extend traditional convolutional operations from regular grids to topological graph structures, equipping them with stronger capabilities for the collaborative extraction of global and local features. This advantage allows GCNs to exhibit excellent performance in the joint modeling of spatial-spectral features in hyperspectral data [5]. For instance, addressing the key issue of differential feature enhancement, Yang et al. [6] and Miao et al. [7] respectively integrated dynamic graph enhancement mechanisms with attention mechanisms. Through refined spatial-spectral feature modeling, they effectively improved the representational ability of differential features. Nevertheless, these GCN-based methods share a common limitation: they mostly focus on modeling single spatial-spectral joint features while neglecting the long-range dependencies inherent within the spectral dimension and within the spatial dimension. This oversight results in insufficient capture of change information in complex scenarios.

To tackle the challenge of modeling long-range dependencies, researchers have begun to explore hybrid network architectures by combining classical models (e.g., CNN, GCN, U-Net) with Transformer, and applying these hybrids to HSI-CD tasks. For example, Shafique

et al. [8] proposed a CNN-Transformer hybrid spatial-spectral convolutional network, which captures spatial-spectral dependencies by leveraging the local feature extraction capability of CNNs and the global dependency modeling capability of Transformers. Zhao et al. [9]captured long-distance change information through a graph-Transformer framework. Zhang et al. [10] developed a Transformer network with a siamese U-shaped structure (SwinSUNet), which processes bi-temporal hyperspectral images in parallel and extracts multi-scale features. Although the aforementioned Transformer-based hybrid architectures have successfully addressed the problem of modeling global long-range dependencies, Transformer still face two core challenges: first, their relatively weak ability to extract local fine-grained features; second, their quadratic computational complexity, which grows with the square of sequence length, making them incompatible with the high-dimensional nature of hyperspectral data. In recent years, researchers have found that introducing a selection mechanism into the state space model (SSM) enables the model to selectively propagate or filter information based on sequence features, while modeling long-term dependencies with linear computational complexity. Based on this insight, the integration of selective SSM with simplified linear transformations gives rise to the Mamba model. By analyzing the long-range dependencies of features in conjunction with sequence length, Mamba efficiently captures long-range dependent features while reducing computational complexity.

Hyperspectral images contain abundant spectral and spatial information, yet most existing studies overlook the interactive relationships in the temporal dimension. Currently, mainstream methods predominantly adopt differential feature interaction strategies to acquire spatio-temporal-spectral three-dimensional features. For example, Jian et al. [11] achieved effective modeling of spectral-spatial-temporal correlations through high-order interaction modeling; Wang et al. [12] extracted feature information from the spectral and spatial dimensions using a multi-layer spectral sequence Transformer and a spatial pixel Transformer, respectively, and subsequently fused temporal sequence information via a triple cross-attention mechanism. However, these methods all follow the workflow of "pixel subtraction first, feature extraction later"—a process that easily leads to the loss of effective detailed information in bi-temporal images, thereby impairing the accuracy of change detection. Additionally, hyperspectral images exhibit high-dimensional complexity [13], [14]; invalid spectral bands not only increase the computational burden of the model but also interfere with the accurate extraction of differential features, ultimately degrading change detection performance. To address this issue, existing studies mostly adopt redundancy removal strategies at the data layer or feature layer to reduce such interference [15]. For example, Chen et al. [16] eliminated redundant information by extracting high-level spectral conceptual features, effectively resolving the problems of band redundancy and spatial information noise in limited regions. How to further design adaptive redundancy removal strategies to enable models to dynamically select effective information based on data characteristics remains one of the key challenges to be addressed in the current field of HSI-CD.

Based on the above analysis, we observe that most existing methods primarily focus on the extraction of spatial-spectral features while neglecting the temporal correlations inherent in bi-temporal images. Furthermore, there exist dependencies between distant spectral bands and between spatially distant pixels. Capturing such long-range dependencies can enhance the global feature representation of the model and mitigate the impact of environmental variations on feature recognition. Lastly, due to the high dimensionality of the data, the extracted features inevitably contain redundancy. Effectively removing this redundancy and enabling adaptive fusion of informative features remains a significant challenge. In this work, we propose SSTGCNs: spectral-spatial-temporal long-range dependencies joint feature extraction with graph convolutional networks. Specifically, the SSTFEM module is first employed to extract joint spectral-spatial-temporal features. Then, the LRSSDM module is utilized to capture long-range dependencies. Finally, the RSAF module is introduced to remove redundant information and integrate the most informative features from both branches, thereby enhancing the model's representation and generalization capabilities. The main contributions of this paper are summarized as follows:

(1) For hyperspectral change detection, we propose a spectral-spatial-temporal Long-Range Dependency Network that integrates GCN-based joint feature extraction with adaptive change detection. In this framework, the SSTFEM module is designed to extract spectral-spatial-temporal features. Notably, instead of directly applying a subtraction operation between bi-temporal images, we combine GCN and the Temporal Interaction Fusion Module (TIFM) to extract difference information. This strategy enhances the capture of spectral-spatial-temporal features while minimizing the loss of informative content, thereby improving the interaction modeling between bi-temporal images.

(2) The LRSSDM module captures long-range dependencies at both the spectral-band and pixel levels. By modeling interactions from a spectral perspective, it extracts material-specific features, mitigates feature discrepancies introduced by environmental factors, and enriches the diversity of feature representations.

(3) To tackle the multi-dimensionality challenge, we

design an adaptive feature selection and fusion module, which differs from conventional fusion strategies by dynamically weighting key temporal features and long-range spectral dependencies through an LSTM gating mechanism. Following a two-stage feature calibration, this module effectively highlights subtle variation signals and long-distance correlations while suppressing redundant temporal noise. Compared with existing approaches, our variant consistently delivers superior results across three datasets, with the most pronounced improvement observed on the River dataset, where complex terrain and abundant spectral bands typically hinder accurate detection. These findings demonstrate that the proposed mechanism not only achieves more precise discrimination between changed and unchanged regions but also significantly improves model generalization and robustness in complex real-world scenarios.

(4) Addressing the limitations of existing methods in hyperspectral image change detection: Previous approaches have paid limited attention to feature information along the temporal dimension and often neglect long-range dependencies, leading to misclassification as changes. In this study, we propose the SSTGCNs network framework and conduct comparative experiments. The results (Table 1 and Figures 8, 9, and 10) show that, in terms of overall accuracy (OA), our method achieves 96.97 and 96.77 on the USA and River datasets, representing improvements of 0.55 and 0.36 over the current state-of-the-art methods. Moreover, on the China and River datasets, the precision (P), recall (R), and F1-score metrics are the best among all reported results. These findings validate the effectiveness of the proposed spectral-spatial-temporal feature module and long-range dependency module, providing a robust solution for hyperspectral change detection.

The remainder of this paper is organized into four sections. Section 2 presents the related work. Section 3 details the proposed network framework. Section 4 discusses the experimental design and result analysis. Finally, the conclusion of the paper is provided in Section 5.

II. RELATED WORK

A. *Spectral–Spatial–Temporal features*

Multi-temporal hyperspectral images exhibit strong distinctive features and correlations across temporal, spatial, and spectral dimensions. Temporally, unchanged regions across different time points often share similar statistical, textural, and spectral characteristics. Even in non-abrupt change regions, various features typically follow certain temporal evolution patterns. Spatially, similar to natural images and conventional remote sensing data, hyperspectral imagery also contains common local and non-local patterns such as texture, edges, and structural similarity. Spectrally, since hyperspectral images represent the reflectance intensity of the same object across multiple spectral bands, there exists a strong intrinsic correlation between adjacent bands. As a result, pixels at the same spatial location across neighboring bands tend to have highly correlated values, leading to strong inter-band spectral dependencies [17]. At present, various types of multi-dimensional correlations in change detection tasks, such as spatial and spatial-spectral correlations have been partially explored and utilized. For instance, Masud et al. [18] proposed a hybrid spectral-spatial feature extraction method, which reduces the feature space while effectively capturing meaningful spectral-spatial information. Similarly, Wang et al. [19] developed a dual-branch framework, where both superpixel and pixel-level features are extracted in parallel from bi-temporal HSIs and complement each other to obtain joint spatial-temporal attention features.Qu et al. [20] proposed a graph domain adaptive network, which extracts differential features through a graph convolution feature extractor and employs domain adaptation for unsupervised learning. Yang et al. [21] developed a sharpening-aware graph network (SAGN), which highlights change information via Laplacian sharpening graph convolution. However, these methods tend to focus on isolated aspects of the correlations—typically one or two dimensions—without fully considering the integrated modeling of multi-dimensional spatial-temporal-spectral features. In particular, there is still a notable lack of quantitative studies on the relational transfer among these multi-dimensional correlations.

Inspired by [22], we integrate Multi-GCN to develop a spectral-spatial-temporal joint feature extraction framework, which mitigates information loss during feature modeling and effectively captures spectral-spatial-temporal features. As illustrated in Fig. 1, intrinsic spectral-spatial features are first extracted using Multi-GCN. Subsequently, temporal interactions are applied to extract change-related features along the time dimension, and the autocorrelation information is then integrated to form the final feature representation.

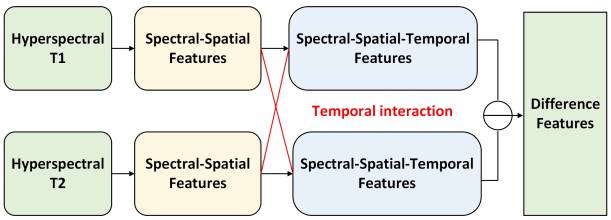


Fig. 1: Cross-Temporal Feature Interaction Strategy.

B. Mamba

Mamba is an efficient sequence modeling architecture based on linear state space models (SSM), combining the long-range dependency modeling capabilities of Transformers with the inference efficiency of RNNs. Its core advantage lies in supporting long sequence modeling with high computational efficiency, making it particularly suitable for tasks with strong temporal and spatial dimensions. Recently, many tasks have incorporated Mamba as an effective module to capture long-range dependencies while reducing computational complexity. For example, Li et al. [23] proposed a hyperspectral image (HSI) classification model based on MambaHSI, which performs long-range interaction modeling on HSI data and adaptively integrates spatial and spectral information to obtain effective interactive features. Peng et al. [24] introduced an innovative approach called FusionMamba for efficient image fusion, integrating Mamba into a U-shaped network to independently extract spatial and spectral features efficiently. Li et al. [25] presented an end-to-end deep learning network that integrates global attention (GA) and state space modeling to capture complex spatial-spectral relationships, enhancing spatial-spectral feature interaction. Wang et al. [26] proposed a dual-branch hybrid architecture that leverages band selection to enhance bidirectional Mamba for establishing long-range dependencies, combined with Chebyshev graph convolution to capture spatial similarity.From the above, it is evident that Mamba has been widely applied in remote sensing; however, its application in the HSI-CD domain remains limited. Existing methods still lack effective modeling of spatial-spectral long-range dependencies. Inspired by [23], this work incorporates Mamba for modeling long-range dependencies at both pixel and spectral levels, aiming to capture fine-grained distant correlations.

The ssm-based mamba model maps a one-dimensional function or sequence to an output $y(t)$ via a hidden state $h'(t)$, which can be expressed as a linear ordinary differential equation.

$$h'(t) = Ah(t) + Bx(t), \tag{1}$$

$$y(t) = Ch(t) + Dx(t). \tag{2}$$

Here, $x(t)$ denotes the input sequence, $h(t)$ represents the hidden state, $h'(t)$ indicates the updated hidden state, and $y(t)$ is the predicted output sequence. A is the evolution parameter, while B and C are the projection parameters.

In the process of mapping the input sequence $x(t)$ to the output space $y(t)$,the state transitions required by the state space model (SSM) are discrete. The discrete output is obtained from sampled values based on the time step of the input. Specifically, the module introduces a temporal parameter Δ and employs the zero-order hold (ZOH) discretization scheme to convert the continuous parameters A and B into their discrete counterparts $\bar{A}$ and $\bar{B}$, which are defined as follows:

$$\bar{A} = \exp(\Delta A), \tag{3}$$

$$\bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - 1) \cdot \Delta B, \tag{4}$$

$$\bar{C} = C, \tag{5}$$

$$h_k = \bar{A}h_{k-1} + \bar{B}x_k, \tag{6}$$

$$y_k = \bar{C}h_k + \overline{D}x_k. \tag{7}$$

Here, $\bar{A}$ and $\bar{B}$ are the discrete matrices, h_{k-1} denotes the hidden state at the previous time step, and h_k represents the hidden state at the current time step.

III. Methodology

Fig. 2 illustrates the overall architecture of the proposed spectral-spatial-temporal feature extraction and adaptive long-range dependency modeling network. Specifically, the SSTFEM module is employed to extract spectral-spatial-temporal features from bi-temporal images independently, capturing interactive multi-dimensional representations. Subsequently, the LRSSDM module is utilized to model long-range dependencies across both pixels and spectral bands. This is followed by an adaptive filtering and fusion module that integrates the long-range dependencies with the spectral-spatial-temporal features in an adaptive manner, filtering out redundant information and generating more discriminative difference features for subsequent binary change detection.

A. Spectral-Spatial-Temporal Feature Extraction Module

Most existing methods for hyperspectral feature extraction primarily focus on spectral–spatial features, while overlooking the temporal interactions between features. In addition, many approaches perform pixel-wise subtraction before feature extraction, which can lead to the loss of critical information. To address these issues, as shown in Fig. 3, we design the SSTFEM module for spectral-spatial-temporal feature extraction. This module effectively preserves key information and enhances the representational capacity of the extracted features. First, the bi-temporal hyperspectral images are segmented into super-pixels to obtain region-aggregated graph structures H_1 and H_2, which can be formally defined as follows:

$$G_{sparse}(A, H) = f_{coo-matrix}(G(A, H)). \tag{8}$$

Here,$f_{coo-matrix}$ denotes the coo-matrix function, which converts the super-pixel graph structure into an

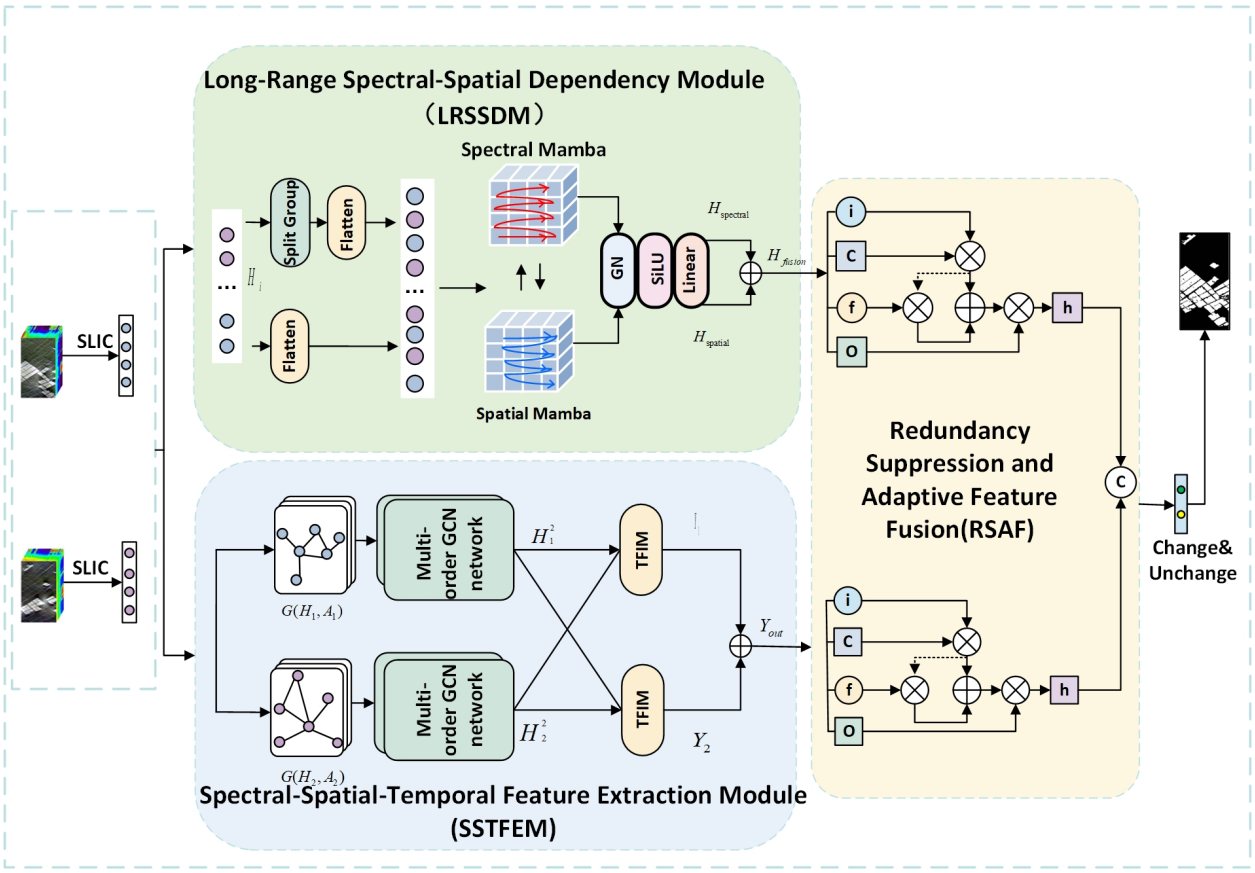


Fig. 2: Overall architecture. It includes Spectral-Spatial-Temporal Feature Extraction Module,Long-Range Spectral-Spatial Dependency Module, Redundancy Suppression and Adaptive Feature Fusion, and unchanged feature suppression.

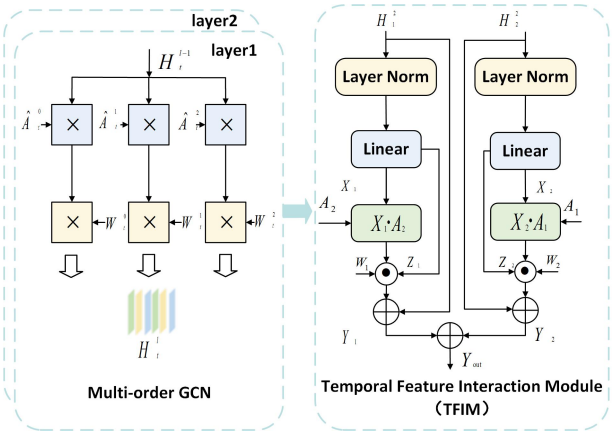


Fig. 3: Multi-order GCN network and cross-temporal information interaction module in spectral-spatial-temporal feature extraction.

adjacency matrix and a feature matrix with informative properties.

As illustrated in Fig. 3, intrinsic spectral-spatial features are extracted using a Multi-order GCN, whose improved multi-order design enhances feature representation across different orders. Specifically, we design a two-layer multi-order GCN module, where each graph convolutional network (GCN) integrates features from the first-order neighborhood with those from the P-order neighborhood, thereby aggregating information across multiple neighborhood orders.In Layer 1, the multi-order GCN comprises three vanilla GC branches, each performing graph convolution operations. Concretely, for the first branch, the feature representation is obtained by left-multiplying the feature matrix with the adjacency matrix, followed by right-multiplication with the weight matrix. The feature representations from the three vanilla GC branches are then concatenated to form the fused features of the Layer 1 multi-order GCN. Layer 2 has the same structure as Layer 1 and performs a second graph convolution on the fused features from Layer 1 to further refine them.Through this process, feature extraction and fusion are performed across different adjacency orders, enabling the model to comprehensively capture both

local and distant structural information of graph nodes, thereby enhancing the feature representation capability. Each vanilla GC is defined as follows:

$$H_q^l = \| j \in \sigma_p(\hat{A}_q^j H_q^{l-1} W_{q,j}^{l-1}). \tag{9}$$

Here, P denotes the neighborhood order, while $||\,(.)$and $\sigma(\cdot)$ represent column-wise concatenation and the activation function ReLU, respectively. H_q^{l-1} and H_q^l denote the input and output of the l-th multi-order GC layer in the q-th branch, respectively. $W_{q,j}^{(l-1)}$ denotes the weight matrix of the j-th vanilla GC layer. $\hat{A}_q^j$ represents the j-th power of the adjacency matrix $\hat{A}_q$ the q-th branch($A_q^j = (A_q)^j$), the self-connected symmetric normalized adjacency matrix $\widehat{A}_q$ for the q-th branch.

The preceding multi-order graph convolutional layers extract the inherent spectral–spatial features H_q^l from T_1 and T_2. Inspired by [27], we devise the TFIM module, as illustrated in Fig. 3, with the objective of establishing temporal interactions between the bi-temporal hyperspectral images. Specifically, the extracted features are first subjected to normalization, followed by the application of an activation function and a linear transformation to generate X_t and Z_t. Subsequently, a gated attention mechanism is employed to derive the attention matrix A_q of H_q^l. The attention matrix A_q is then cross-multiplied with the information matrix to obtain the cross-temporal interaction matrix V_t. Finally, through a gated filtering mechanism, cross-temporal features enriched with discriminative information are selectively preserved, resulting in the output feature Y_t.

More specifically, H_q^l is first normalized and then transformed by an activation function and a linear operation to produce X_t and Z_t, where W_x and W_z denote the learnable weight matrices.

$$\widehat{H}_q^l = LayerNorm(H_q^l), \tag{10}$$

$$X_t = Linear(ReLU(\widehat{H}_q^l \cdot W_x)), \tag{11}$$

$$Z_t = Linear(\mathrm{ReLU}\left(\widehat{H}_q^l \cdot W_z\right)). \tag{12}$$

The attention matrix A_q is then obtained using the following formulation, where W_o denotes the associated weight matrix, and Q and K represent the query and key in the attention mechanism, respectively. θ_{q}, μ_{q}, θ_{k}, and μ_{k} are learnable parameters.

$$Q = \theta_q \cdot (Linear(ReLU(\widehat{H}_q^l \cdot W_o))) + \mu_q, \tag{13}$$

$$K = \theta_{\mathbf{k}} \cdot \left(Linear\left(ReLU(\widehat{H}_q^l \cdot W_o)\right)\right) + \mu_{\mathbf{k}}, \tag{14}$$

$$A_q = \frac{1}{n} relu^2\left(\frac{Q \cdot K^T}{\sqrt{d}} + b\right). \tag{15}$$

Subsequently, the linearized feature matrix X_t is cross-multiplied with the attention matrix A_q to model cross-temporal interaction relationships:

$$V_1 = X_1 \cdot A_2, \tag{16}$$

$$V_2 = X_2 \cdot A_1. \tag{17}$$

V_1 and V_2 are cross-temporal interaction feature matrices. This cross-attention module facilitates the modeling of interactions between bi-temporal images, thereby obtaining spectral–spatial–temporal features. A gated mechanism is then employed to filter the information, selectively preserving cross-temporal features enriched with discriminative differences while suppressing irrelevant ones. The integrated output Y_t encompasses both intrinsic change detection features and cross-temporal features, while Y_{out} represents the overall spectral–spatial–temporal features.

$$Y_t = (Z_t \odot V_t) \cdot W_t + \widehat{H}_q^l, \tag{18}$$

$$Y_{out} = Y_1 \oplus Y_1. \tag{19}$$

This module provides an effective approach for learning spectral, spatial, and temporal relationships from sequential images. The multidimensional feature representations enhance the model's expressive capacity and improve the accuracy of change detection.

B. Long-Range Spectral-Spatial Dependency Module

In hyperspectral images, complex long-range dependencies widely exist both between spectral bands and among spatial pixels. These dependencies intrinsically reflect the underlying patterns of surface object variations. However, most existing feature extraction methods rely on shallow architectures, which are insufficient for capturing discriminative and salient change features, thereby limiting the accuracy of change detection tasks. Therefore, effectively modeling long-range dependencies in hyperspectral data is of significant research importance. Although Transformer architectures have demonstrated outstanding capabilities in modeling such long-range interactions, their quadratic computational complexity imposes substantial resource burdens when applied to large-scale hyperspectral data. To address these issues, as illustrated in Fig. 4, we propose a dual-branch spectral–pixel long-range dependency modeling structure based on Mamba. The Mamba block integrates linear transformation, one-dimensional convolution, and a state space model (SSM) to capture long-range dependencies across pixels and spectra, thereby enhancing the model's capability to learn from complex scenes. The detailed operations for spectral–pixel long-range dependency extraction are as follows:

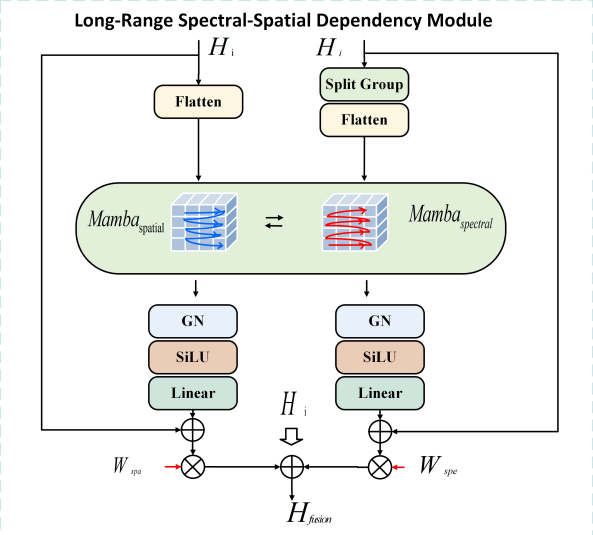


Fig. 4: Illustration of Long-Range Spectral-Spatial Dependency Module(LRSSDM).

The proposed model consists of two branches. The left branch is designed to capture long-range dependencies in the superpixel space. Specifically, the spatial information H_i is first flattened, after which Mamba performs selective horizontal and vertical scanning to capture the potential dependencies among distant spatial pixels. These dependencies are repeatedly scanned and preserved, followed by group normalization (GN). Finally, an activation function and a linear transformation are applied to obtain the change information H_{spa}. The corresponding formulation is expressed as follows:

$$\widehat{H}_i = Flatten(H_i), \quad (20)$$

$$H_{spa} = Linear(SiLU(GN(\text{Mamba}(\widehat{H}_i)))). \quad (21)$$

The right branch is dedicated to capturing long-range dependencies in the spectral domain. Unlike the superpixel spatial dependency modeling, this process begins with band grouping (Split Group). In our experiments, the numbers of groups for the three datasets are set to 5, 7, and 9, respectively. After grouping, the same preprocessing operations are applied, and the spectral Mamba is utilized to model intra-group long-range dependencies, where each intra-group dependency characterizes the properties of the corresponding group. Subsequently, inter-group long-range dependencies are extracted through two rounds of selective storage of spectral dependencies. This yields the intrinsic spectral change features. Here, SG denotes the Split Group operation,and the detailed formulation is expressed as follows:

$$\widehat{H}_i = \text{Flatten}(\text{SG}(H_i), \quad (22)$$

$$H_{spe} = Linear(SiLU(GN(\text{Mamba}(H_i))))). \quad (23)$$

Finally, the spatial long-range dependency features and spectral long-range dependency features are fused to achieve the modeling of long-range dependency features for hyperspectral data.

$$H_{fusion} = H_i + W_{spa} \times H_{spa} + W_{spe} \times H_{spe}. \quad (24)$$

The super-pixels are first flattened, followed by long-range spectral and spatial dependency modeling using the Mamba module. A linear activation function is then applied to generate the output features. Subsequently, effective fusion of spectral and spatial features is performed to obtain the final long-range dependency representation.

C. Redundancy Suppression and Adaptive Feature Fusion

Due to the high dimensionality of hyperspectral data, a significant amount of redundant features may affect the accuracy of change detection. To mitigate this issue, we design an Adaptive Fusion Module, which filters out irrelevant information and integrates complementary features from both the SSTFEM and LRSSDM modules. This fusion strategy effectively emphasizes key features, enabling more accurate discrimination between changed and unchanged regions. This module leverages an LSTM-based approach to capture the temporal characteristics inherent in bi-temporal hyperspectral imagery. It primarily consists of an input gate i_t,cell state c_t, forget gate f_t, and output gate o_t. he input gate i_t is defined as follows:

$$i_t = \sigma(W_i \cdot [h_t, x_t] + b_i). \quad (25)$$

Here, $\sigma(\cdot)$ denotes the sigmoid activation function, and W_i is the weight matrix. h_t and x_t epresent the previously retained useful information (i.e., the hidden state) and the current input to be memorized, respectively. These two components are combined to form a new memory representation. b_i is the bias vector.

The forget gate f_t removes irrelevant information from the cell state, W_f denotes the weight matrix associated with the forget gate, and b_f is the corresponding bias vector. defined as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \quad (26)$$

The cell state C_t selectively removes outdated information from the previous state and integrates new information. It is specifically defined as follows:

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_f), \quad (27)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \quad (28)$$

Here, $\widetilde{C}_t$ denotes the candidate cell state. The function $tanh(\cdot)$ is the hyperbolic tangent activation, and C_{t-1} represents the previous cell state. This process continuously updates the stored important information, ensuring that the final state retains critical features of changed and unchanged regions.

The output gate o_t controls the amount of information from the cell state C_t to be exposed, and can be defined as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{29}$$

$$h_t = o_t * tanh(C_t). \tag{30}$$

Here,h_t represents the final output. Through the Adaptive Fusion Module, spectral-spatial-temporal features and long-range dependencies are effectively integrated while redundant information is filtered out. This process enables more accurate identification of changed and unchanged regions.

D. Loss Function

Hyperspectral change detection is essentially a binary classification task. In this study, the cross-entropy loss function is adopted and is defined as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log \widehat{y}_i + (1-y_i)\log(1-\widehat{y}_i)). \tag{31}$$

Here, N denotes the number of training samples, y_i represents the ground truth label of the i-th pixel, and $\hat{y}_i$ indicates the predicted label for the i-th sample

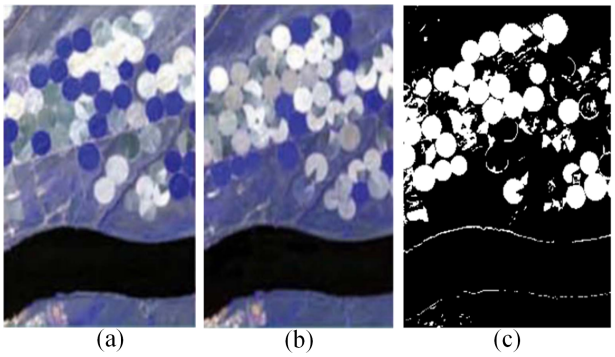
IV. EXPERIMENTS

A. Datasets

To evaluate the performance of the proposed network, we employ several widely used hyperspectral image (HSI) change detection datasets, including the USA Dataset, China Dataset, and River Dataset.

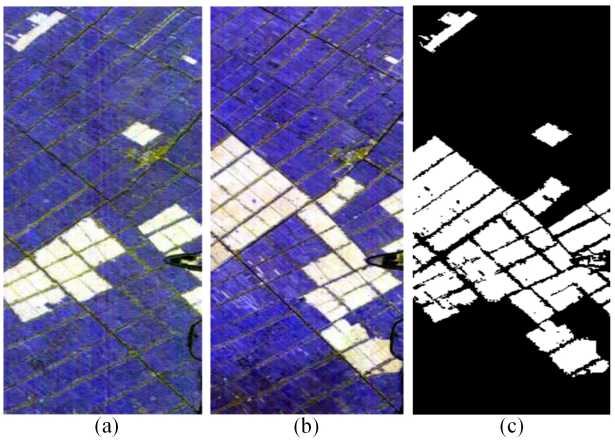
(1)USA Dataset: As illustrated in Fig. 5, the USA Dataset was collected over irrigated farmland in Hermiston, Umatilla County, Oregon, USA. Two hyperspectral images were acquired using the EO-1 Hyperion sensor on May 1, 2004, and May 8, 2007, respectively. With a spatial resolution of 50 meters, the dataset spans a spectral range from approximately $0.4\,\mu$m to $2.5\,\mu$m , comprising 154 spectral bands. Each image has spatial dimensions of 307×241 pixels.

(2)China Dataset: As shown in Fig. 6, the China Dataset was acquired over a farmland area in Yancheng, Jiangsu Province, China. Two hyperspectral images were captured using the EO-1 Hyperion sensor on May 3, 2006, and April 23, 2007, respectively. With a spatial resolution of 30 meters, the dataset spans a spectral range from approximately $0.4\,\mu$m to $2.5\,\mu$m, covering



(a) (b) (c)

Fig. 5: USA dataset. (a) T1 (2004). (b) T2 (2007). (c) GT.



(a) (b) (c)

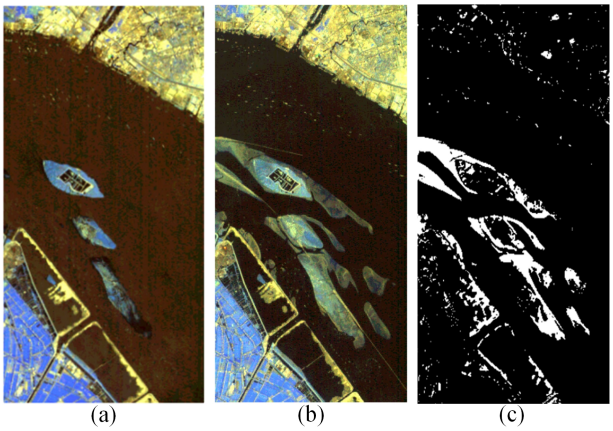
Fig. 6: China dataset. (a) T1 (2006). (b) T2 (2007). (c) GT.

155 spectral bands in total. Each image has spatial dimensions of 450×140 pixels.

(3)River Dataset: As depicted in Fig. 7, the River Dataset was collected over a river area near Jiangsu Province, China. Two hyperspectral images were acquired by the EO-1 Hyperion sensor in May 2013 and December 31, 2013, respectively. The dataset features a spatial resolution of 30 meters and covers a spectral range from approximately $0.4\,\mu$m to $2.5\,\mu$m, consisting of 198 spectral bands in total. Each image has spatial dimensions of 463×241 pixels.

B. Evaluation Metrics

To comprehensively evaluate the performance of the proposed method, five widely used metrics are adopted: Overall Accuracy (OA), Kappa Coefficient (K), Precision (P), Recall (R), and F1-Score (F1).OA measures the proportion of correctly detected change and non-change pixels over the entire dataset. The Kappa coefficient quantifies the agreement between the predicted results



(a) (b) (c)

Fig. 7: River dataset. (a) T1 (2013). (b) T2 (2013). (c) GT.

and the ground truth, while accounting for the possibility of agreement occurring by chance. Precision reflects the proportion of true positive predictions among all positive predictions, indicating the model's reliability in predicting the change class. Recall denotes the proportion of true positive samples correctly identified from all actual positive samples, reflecting the model's sensitivity.F1-score, defined as the harmonic mean of precision and recall, provides a balanced evaluation by considering both false positives and false negatives. Higher values of all five metrics indicate better performance of the proposed change detection method. The formal definitions of these metrics are given as follows:

$$OA = \frac{TP + TN}{TP + TN + FN + FP}, \tag{32}$$

$$Kappa = \frac{OA - p_e}{1 - p_e}, \tag{33}$$

$$P = \frac{TP}{TP + FP}, \tag{34}$$

$$R = \frac{TP}{TP + FN}, \tag{35}$$

$$F1 = 2 \times \frac{P \times R}{P + R}, \tag{36}$$

$$p_e = \frac{(TP + FP)(TP + FN) + (TN + FP)(TN + FN)}{(TP + FP + TN + FN)^2}. \tag{37}$$

Here, TP (True Positives) refers to the number of pixels that are correctly identified as changed regions, while TN (True Negatives) denotes the number of pixels correctly identified as unchanged regions. FP (False Positives) represents the number of pixels incorrectly classified as changed, and FN (False Negatives) refers to the number of pixels incorrectly classified as unchanged.

C. Compared Methods and Experimental Setup

To verify the effectiveness of the proposed Spatio-Spectral-Temporal Long-Range Dependency Network with GCN-based Joint Feature Extraction and Adaptive Change Detection, we conducted comparative experiments with several recent state-of-the-art hyperspectral change detection methods, including CDFormer [13], SST-Former [12], DCENet [28], DIEFEN [29], AIWSEN [30], and MGCN [22]. Among these methods, CDFormer and SST-Former employ the Transformer architecture to extract interactive temporal features for change detection. DCENet, DIEFEN, and AIWSEN focus on enhancing differential features to extract key change information for binary change detection tasks. MGCN applies graph convolution to capture local differential information within superpixels for subsequent change detection.

In our experiments, 5% of labeled samples were randomly selected for training. The learning rate was set to 0.0001, with a weight decay of 0.0001, and the models were trained for 1000 epochs in total. The multi-scale GCN layers were configured in a lightweight structure with channel dimensions of 64, 64, and 8, respectively. For the China, USA, and River datasets, the number of spectral groups was set to 5, 7, and 9, respectively. All methods were implemented using PyTorch and trained/tested with Mamba acceleration on an NVIDIA vGPU with 32 GB memory.

D. Experimental Results

Table 1 presents the performance comparison between the proposed method and several recent baseline algorithms on the USA, China, and River datasets. The evaluation metrics include Overall Accuracy (OA), Kappa Coefficient (Kappa), Precision, Recall, and F1-Score.Figure 8, Figure 9, and Figure 10 illustrate the visual comparison results for the USA, China, and River datasets, respectively.

(1)USA Dataset Results Analysis: As shown in Fig. 8, DCENet, DIEFEN, and MGCN exhibit missed detections and false alarms within the yellow boxes, and they are less sensitive to edge changes. This is likely because these methods rely on difference operations between bi-temporal data before feature extraction, which tends to lose information. Moreover, hyperspectral data contain multi-dimensional spectral bands, and some attention-based methods struggle to capture spectral differences effectively.CD-Former and SST-Former show some false positives in both yellow and red boxes. Although Transformers excel at extracting global information and modeling long-range dependencies, their ability to capture

TABLE I: Performance Comparison of Baseline Methods and the Proposed Method on USA, China, and River Datasets

Dataset	Methods	OA(%)	κ(%)	P(%)	R(%)	F1(%)
USA	CD-Former	93.64	81.69	86.33	85.27	85.79
	SST-Former	95.49	86.96	89.93	91.97	83.61
	DCENet	94.62	83.89	93.40	81.93	87.29
	DIEFEN	95.58	87.53	90.41	90.22	92.31
	AIWSEN	96.42	89.74	92.05	92.10	92.03
	MGCN(base)	95.11	86.01	95.12	**95.11**	**95.11**
	Ours	**96.97**	**91.18**	**95.66**	91.28	93.13
China	CD-Former	94.38	86.55	88.54	92.62	90.54
	SST-Former	95.21	88.43	92.96	95.04	93.88
	DCENet	96.56	91.56	95.72	92.27	93.97
	DIEFEN	96.86	92.48	94.70	92.77	96.74
	AIWSEN	**97.53**	**94.03**	95.77	95.12	96.45
	MGCN(base)	96.62	92.19	95.70	96.52	96.09
	Ours	97.01	92.78	**98.36**	**97.41**	**97.88**
River	CD-Former	94.27	82.76	87.99	80.71	84.19
	SST-Former	93.88	64.69	76.47	83.88	82.16
	DCENet	95.32	**90.24**	94.07	88.61	91.53
	DIEFEN	96.06	74.27	76.41	79.67	73.49
	AIWSEN	96.41	75.59	77.54	84.10	72.10
	MGCN(base)	95.08	67.43	96.84	97.80	97.32
	Ours	**96.77**	79.48	**98.16**	**98.30**	**98.23**

fine-grained local features is relatively weak, leading to local misclassifications. In contrast, our proposed method closely matches the ground truth within the red and yellow boxes, accurately identifying subtle and edge changes. This improvement stems from the incorporation of the LRSSM module, which extracts spectral features of change information and thus can accurately identify changes from the spectral dimension in complex scenes.

As shown in Table I, compared with other state-of-the-art methods, our approach achieves the best performance in terms of OA, Kappa, and Precision, with values of 96.97, 91.18, and 95.66, respectively. These results outperform MGCN by 1.86%, 5.17%, and 0.54%. This demonstrates that our model exhibits greater stability and achieves superior performance in extracting discriminative information from bi-temporal hyperspectral data.

(2)China Dataset Results Analysis: The surface changes in the China dataset are relatively regular, but subtle variations remain difficult to detect. As shown in Fig. 9, SST-Former, CD-Former, and DCENet exhibit false positives within the red boxes, misclassifying unchanged areas as changed, and show limited accu-

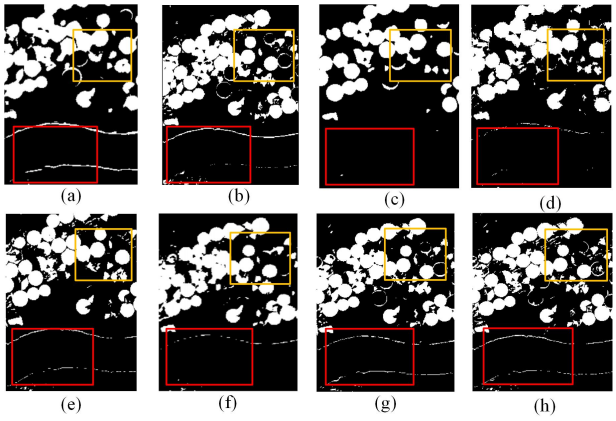


Fig. 8: Change Detection Map of the USA Dataset. (a) CD-Former (b) SST-Former (c) DCENet (d) DIEFEN (e) AIWSEN (f) MGCN (g) Ours (h) Ground-truth

racy in delineating the boundaries between changed and unchanged regions. Overall, DIEFEN, AIWSEN, and MGCN struggle with identifying complex small regions, often misclassifying changed areas as unchanged.In contrast, our proposed method successfully detects subtle changes in complex regions highlighted by the red and yellow boxes. This improvement is attributed to our strategy of extracting features before computing difference information, which better preserves subtle variations that might otherwise be lost through direct subtraction.

TABLE II: Ablation Experiments on USA, China, and River

Datasets	TFIM	LRSSDM	RSAF	OA(%)	Kappa(%)
USA	✗	✗	✗	95.11	86.01
	✓	✗	✗	95.86	88.72
	✓	✓	✗	96.51	89.72
	✓	✓	✓	**96.97**	**91.18**
China	✗	✗	✗	96.62	92.19
	✓	✗	✗	96.82	92.75
	✓	✓	✗	97.01	92.70
	✓	✓	✓	**97.07**	**92.87**
River	✗	✗	✗	95.08	67.43
	✓	✗	✗	95.56	72.10
	✓	✓	✗	96.55	78.45
	✓	✓	✓	**96.77**	**79.48**

As shown in Table I, compared with the GCN-based MGCN method, our approach achieves improvements of 0.39% in OA, 0.59% in Kappa, 2.66% in Precision, 0.89% in Recall, and 1.79% in F1 score. In comparison with Transformer-based methods, our model incorporates Mamba for long-range dependency modeling, thereby constructing global change information relationships. Notably, our method not only reduces computational resource consumption but also surpasses Transformer-based approaches across all metrics, including OA, Kappa, Precision, Recall, and F1. These results collectively demonstrate that our method possesses distinct advantages in change detection tasks.

(3)River Dataset Results Analysis: Fig. 10 illustrates the change detection maps produced by different algorithms. It is observed that SST-Former, CD-Former, DCENet, and DIEFEN exhibit missed detections or false alarms within the yellow boxes, either failing to identify changes or misclassifying unchanged pixels as changed. Within the red boxes, most algorithms fail to accurately detect subtle changes. In comparison, our proposed method shows the closest identification of change information to the ground truth within both red and yellow boxed areas.

Table I presents the quantitative results of all methods on the River dataset. Our proposed method achieves the best performance across four metrics, namely Overall Accuracy (OA), Precision (P), Recall (R), and F1-score. Specifically, its OA value outperforms that of the state-of-the-art AIWSEN by 0.36%. Meanwhile, in terms of the Kappa coefficient, Precision (P), Recall (R), and

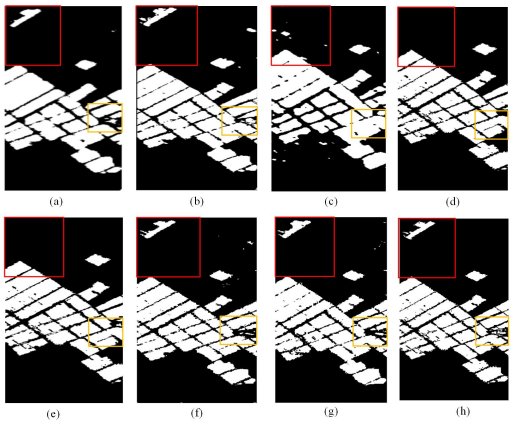


Fig. 9: Change Detection Map of the China Dataset. (a) CD-Former (b) SST-Former (c) DCENet (d) DIEFEN (e) AIWSEN (f) MGCN (g) Ours (h) Ground-truth

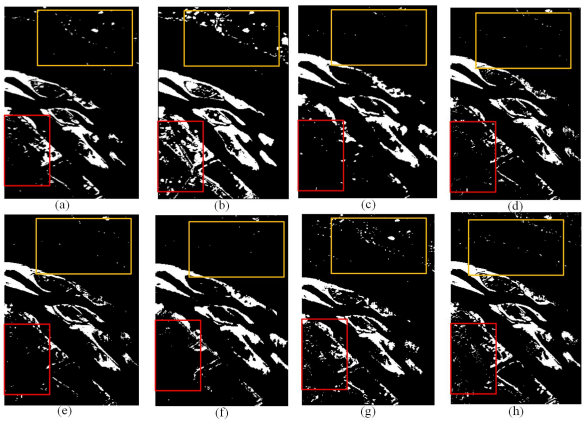


Fig. 10: Change Detection Map of the River Dataset. (a) CD-Former (b) SST-Former (c) DCENet (d) DIEFEN (e) AIWSEN (f) MGCN (g) Ours (h) Ground-truth

F1-score, our method surpasses MGCN by 12.05%, 1.32%, 0.50%, and 0.91% respectively.The underlying reasons for this superiority are twofold: first, for datasets with rich spectral information, our model is capable of capturing more spectral details. Second, to mitigate information loss, we perform bitemporal feature extraction separately prior to conducting cross-temporal information interaction. This design not only highlights change-related information more prominently but also significantly enhances the accuracy of change detection.

E. Ablation Study

The proposed GCN-based spatio-temporal-spectral long-range dependency network with adaptive change

TABLE III: Model Parameters and Training/Testing Time Comparison Across Methods

Method	CDFormer	SSTFormer	DCENet	DIEFEN	AIWSEN	MGCN	Ours
Parameter (M)	0.7	2.50	35.30	8.08	12.23	0.1	1.58
Train time	536.41	224.56	114.48	48.65	205.80	48.40	162.02
Test time	14.94	12.53	6.16	2.33	4.53	0.06	0.09

detection consists of three key modules: TFIM, LRSSDM, and RSAF. To evaluate their effectiveness, we designed four variants: the first variant excludes all three modules; the second includes TFIM, introducing cross-temporal feature interaction; the third adds TFIM and LRSSDM, which captures long-range dependencies in both spectral and spatial dimensions to enhance feature representation and improve change detection accuracy; the fourth variant incorporates all three modules (TFIM + LRSSDM + RSAF), where RSAF adaptively fuses crucial change information while suppressing unchanged features to produce the final change representation.

Experiments on the three datasets demonstrate the effectiveness of these modules. As shown in Table 2, adding TFIM significantly improves OA and Kappa across all datasets, validating the importance of cross-temporal interaction for enhanced feature representation. With the inclusion of LRSSDM, OA increases by 0.71%, 0.19%, and 0.99% on USA, China, and River datasets, respectively. This improvement is attributed to the Mamba module's ability to capture long-range spectral and spatial dependencies, thereby emphasizing global change information. Finally, RSAF further contributes by effectively fusing key change cues, resulting in improved detection accuracy.

F. Model Efficiency

As shown in Table 3, the number of parameters and the training/testing time vary across different models. It is evident that DCENet, AIWSEN, and DIEFEN possess relatively larger numbers of parameters, primarily due to their modular stacking strategies designed to achieve superior performance. However, this architectural design inevitably leads to increased memory consumption. In contrast, transformer-based approaches tend to exhibit higher computational complexity, resulting in longer training times. Compared with these methods, the proposed approach strikes a balance between model size and computational cost, demonstrating a favorable trade-off between efficiency and performance.

V. CONCLUSION

In this paper, we propose a SSTGCNs: spectral-spatial-temporal long-range dependencies joint feature extraction with graph convolutional networks, which jointly utilizes GCN-based feature extraction and adaptive change detection. The method performs information interaction across spectral, spatial, and temporal dimensions to effectively capture change information. It consists of three main modules: spectral-spatial-temporal joint feature extraction, pixel-spectral long-range dependency modeling, and adaptive redundancy suppression with feature fusion. By combining multi-order graph convolution with the TFIM module, inherent spectral-spatial features are extracted and cross-temporal information interaction is performed to obtain spectral-spatial-temporal features. Then, global difference information is captured through MAMBA-based spectral-pixel long-range dependency modeling, enhancing the model's feature representation capability. Finally, an adaptive feature fusion and redundancy suppression module is designed, which iteratively updates to suppress unchanged information while enhancing change information, thereby improving detection accuracy. Experimental results on three public datasets demonstrate that our method outperforms state-of-the-art approaches, especially in data with rich spectral information and scenarios with subtle and complex changes, where the proposed model achieves superior change detection performance.

REFERENCES

[1] Xiande Wu, Paolo Gamba, Jie Feng, Ronghua Shang, Xiangrong Zhang, and Licheng Jiao. A multitask framework for hyperspectral change detection and band reweighting with unbalanced contrastive learning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024.

[2] Changjiang Shi, Zhijie Zhang, Wanchang Zhang, Chuanrong Zhang, and Qiang Xu. Learning multiscale temporal–spatial–spectral features via a multipath convolutional lstm neural network for change detection with hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022.

[3] Jiahui Qu, Wenqian Dong, Qian Du, Yufei Yang, Yunshuang Xu, and Yunsong Li. Cyclic consistency constrained multiview graph matching network for unsupervised heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025.

[4] Bing Yang, Shirui Pan, Weiwei Sun, Ziyang Guo, Zhijing Ye, and Jiangtao Peng. Fcfda: Fine–coarse–fine progressive graph framework with distribution alignment for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[5] Rajat Singh, Sanchit Bathla, and Priyanka Meel. State-of-the-art applications of graph convolutional neural networks. In *Proceedings of 6th International Conference on Recent Trends in Computing: ICRTC 2020*, pages 107–115. Springer, 2021.

[6] Bin Yang, Xinwei Cheng, Wei Chen, and Xin Ye. A graph-based hyperspectral change detection framework using difference augmentation and progressive reconstruction with limited labels. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[7] Rui Miao, Yuxiang Zhang, Yanni Dong, and Bo Du. Cross-task meta-learning network with graph-enhanced attention module for hyperspectral change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

[8] Ayesha Shafique, Seyd Teymoor Seydi, Tayeb Alipour-Fard, Guo Cao, and Di Yang. Ssvit-hcd: A spatial–spectral convolutional vision transformer for hyperspectral change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:6487–6504, 2023.

[9] Xiaoyang Zhao, Siyao Li, Tingting Geng, and Xianghai Wang. Gtranscd: Graph transformer-guided multitemporal information united framework for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024.

[10] Cui Zhang, Liejun Wang, Shuli Cheng, and Yongming Li. Swinsunet: Pure transformer network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[11] Ping Jian, Yimin Ou, and Keming Chen. Hypergraph self-supervised learning-based joint spectral-spatial-temporal feature representation for hyperspectral image change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:741–756, 2025.

[12] Yanheng Wang, Danfeng Hong, Jianjun Sha, Lianru Gao, Lian Liu, Yonggang Zhang, and Xianhui Rong. Spectral–spatial–temporal transformers for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[13] Jigang Ding, Xiaorun Li, and Liaoying Zhao. Cdformer: A hyperspectral image change detection method based on transformer encoders. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[14] Xu Tang, Tianxiang Zhang, Jingjing Ma, Xiangrong Zhang, Fang Liu, and Licheng Jiao. Wnet: W-shaped hierarchical network for remote-sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023.

[15] Xianghai Wang, Keyun Zhao, Xiaoyang Zhao, and Siyao Li. Tritf: A triplet transformer framework based on parents and brother attention for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023.

[16] Yaxiong Chen, Zhipeng Zhang, Le Dong, Shengwu Xiong, and Xiaoqiang Lu. A joint saliency temporal–spatial–spectral information network for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

[17] Lazhar Khelifi and Max Mignotte. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *IEEE Access*, 8:126385–126400, 2020.

[18] Masud Ibn Afjal, Md. Nazrul Islam Mondal, and Md. Al Mamun. Spectral-spatial feature extraction for change detection in bi-temporal hyperspectral images. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pages 1–7, 2024.

[19] Xianghai Wang, Keyun Zhao, Xiaoyang Zhao, and Siyao Li. Csdbf: Dual-branch framework based on temporal–spatial joint graph attention with complement strategy for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022.

[20] Jiahui Qu, Jingyu Zhao, Wenqian Dong, Song Xiao, Yunsong Li, and Qian Du. Feature mutual representation-based graph domain adaptive network for unsupervised hyperspectral change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[21] Bing Yang, Weiwei Sun, and Jiangtao Peng. Sagn: Sharpening-aware graph network for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–12, 2024.

[22] Yuxiang Zhang, Rui Miao, Yanni Dong, and Bo Du. Multiorder graph convolutional network with channel attention for hyperspectral change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:1523–1534, 2024.

[23] Yapeng Li, Yong Luo, Lefei Zhang, Zengmao Wang, and Bo Du. Mambahsi: Spatial–spectral mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024.

[24] Siran Peng, Xiangyu Zhu, Haoyu Deng, Liang-Jian Deng, and Zhen Lei. Fusionmamba: Efficient remote sensing image fusion with state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024.

[25] Yinhe Li, Jinchang Ren, Hang Fu, and Genyun Sun. Gassm: Global attention and state space model based end-to-end hyperspectral change detection. *Journal of the Franklin Institute*, 362(3):107424, 2025.

[26] Hao Wang, Peixian Zhuang, Xiaochen Zhang, and Jiangyun Li. Dbmgnet: A dual-branch mamba-gcn network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–17, 2025.

[27] Haoyang Yu, Hao Yang, Lianru Gao, Jiaochan Hu, Antonio Plaza, and Bing Zhang. Hyperspectral image change detection based on gated spectral–spatial–temporal attention network with spectral similarity filtering. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024.

[28] Fulin Luo, Tianyuan Zhou, Jiamin Liu, Tan Guo, Xiuwen Gong, and Xinbo Gao. Dcenet: Diff-feature contrast enhancement network for semi-supervised hyperspectral change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[29] Lanxin Wu, Lei Ma, Jiangtao Peng, Bing Yang, Weiwei Sun, Junhao Wang, and Xinyu Luo. Diefen: Differential information-enhanced feature exchange network for hyperspectral change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–12, 2024.

[30] Lanxin Wu, Jiangtao Peng, Bing Yang, Weiwei Sun, and Zhijing Ye. Aiwsen: Adaptive information weighting and synchronized enhancement network for hyperspectral change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–12, 2025.