

Hyperspectral Image Change Detection Based on Gated Spectral–Spatial–Temporal Attention Network With Spectral Similarity Filtering

Haoyang Yu^{ID}, Member, IEEE, Hao Yang^{ID}, Lianru Gao^{ID}, Senior Member, IEEE, Jiaochan Hu^{ID}, Antonio Plaza^{ID}, Fellow, IEEE, and Bing Zhang^{ID}, Fellow, IEEE

Abstract—Hyperspectral imaging enables advanced change detection (CD) but struggles with extensive redundant data across spatial and spectral dimensions. This bloats model size and computational loads. To address this problem, we propose a new gated spectral–spatial–temporal attention network with spectral similarity filtering (HyGSTAN) with a lightweight yet accurate architectural design. Specifically, our HyGSTAN introduces three innovative modules: 1) spectral similarity filtering to reduce spectral redundancy via cosine similarity; 2) gated spectral–spatial attention to capture intra-image spatial features using single-head weak self-attention and gated mechanisms; and 3) gated spectral–spatial–temporal attention to extract inter-image temporal changes. Experiments on three benchmark datasets demonstrate HyGSTAN’s ability to balance accuracy, model complexity, and computational efficiency. The proposed attention mechanisms extract more discriminative information without sacrificing performance. The source code of this work will be released at <https://github.com/Welcome-to-LISA/HyGSTAN>.

Index Terms—Attention mechanism, change detection (CD), deep learning, hyperspectral images.

I. INTRODUCTION

SURVEILLANCE of natural terrestrial surface transformations and human-induced alterations over an extended period has been a critical endeavor within the realm of remote sensing [1]. This monitoring is beneficial across various sectors such as natural disaster assessment [2], climate change

Manuscript received 12 December 2023; revised 21 February 2024; accepted 1 March 2024. Date of publication 5 March 2024; date of current version 21 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42101350 and Grant 42371353; in part by the China Postdoctoral Science Foundation under Grant 2022T150080 and Grant 2023T160073; and in part by Open Research Fund of Key Laboratory of Digital Earth Science, Chinese Academy of Sciences under Grant 2022LDE008. (*Corresponding author: Lianru Gao*)

Haoyang Yu, Hao Yang, and Jiaochan Hu are with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: yuh@dlmu.edu.cn; alain@dlmu.edu.cn; hujc@dlmu.edu.cn).

Lianru Gao is with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaolr@aircas.ac.cn).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, 10003 Cáceres, Spain (e-mail: aplaza@unex.es).

Bing Zhang is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zb@radi.ac.cn).

Digital Object Identifier 10.1109/TGRS.2024.3373820

analysis [3], environmental monitoring [4], agriculture [5], and military applications. In this context, examining dual-temporal-phase images of identical geographical regions for change detection (CD) is an essential technique in regional monitoring [6]. Hyperspectral imagery, with their comprehensive and densely sampled spectral information, can potentially depict land features’ physical attributes with extraordinary precision, providing richer information than traditional multispectral imagery [7], [8], [9]. However, effectively employing these data for CD presents a significant and ongoing research challenge in remote sensing image processing [10], [11], [12].

Conventional hyperspectral image CD techniques can be broadly divided into three primary approaches. First, algebraic methods incorporate techniques stemming from image algebra, such as image differencing, image rationing, image regression, and change vector analysis (CVA) [13]. These techniques focus on employing mathematical operations to emphasize changes between dual-temporal-phase hyperspectral images [14]. Second, transform-based methods include strategies like principal component analysis (PCA) [15] and a number of multivariate CD methodologies, which aim to use mathematical transformations for extracting meaningful change information from hyperspectral datasets. Lastly, classification-based methods, such as support vector machines (SVMs) [16] and K -means clustering, have an objective to discover changes by classifying pixels into separate categories depending on their spectral characteristics. However, it is imperative to underline that these conventional methodologies harbor inherent shortcomings. A common obstacle is their inefficiency in exploiting the inherent spatial and spectral components of hyperspectral data. Coupled with sensitivity to noise interference and radiation fluctuations, this makes the process of threshold selection particularly challenging. Additionally, these techniques necessitate the manual configuration of feature extractors, which might not fully encapsulate the intricate characteristics of hyperspectral data [17].

In recent years, deep learning has achieved remarkable progress in various fields such as computer vision and natural language processing [18], [19], [20]. This advancement has brought new perspectives to address the inherent challenges in CD for hyperspectral images [21], [22], [23]. Deep learning models are renowned for their powerful feature learning capabilities and ability to autonomously extract meaningful representations, effectively capturing the complexity of spatial

and spectral elements embedded in hyperspectral data [24], [25], [26], [27], [28], [29]. This enhances the accuracy and robustness of CD, leading to wider adoption of deep learning-based strategies for hyperspectral image CD. For example, Wang et al. [30] introduced a 2-D CNN network that utilizes a hybrid affinity matrix and improved feature learning to extract change representations. Qu et al. [31] proposed a multilevel encoder-decoder attention network that incorporates convolutional layers to extract and integrate hierarchical features, a context-guided attention module to transfer spatial spectral features, and a long short-term memory to analyze temporal dependencies. Qu et al. [32] designed a dual-branch differential amplification graph convolutional network with distinct feature extraction and difference amplification to suppress similar regions. Ou et al. [33] presented a CNN framework with compact band weighting and multiscale spatial attention for pixel-level CD. Luo et al. [34] proposed a multiscale differential feature fusion network to detect subtle changes, using a two-way differential feature module and fusion module to integrate multiscale change features. Li et al. [35] first proposed a spectral reconstruction network based on multitemporal multispectral images, which can generate highly temporal-spatial-spectral resolution of hyperspectral images from the input multispectral images. They further proposed a multisensor multispectral reconstruction framework [36], which is the first attempt to comprehensively utilize the spectral response functions of multiple multispectral sensors to obtain more accurate reconstruction results.

More recently, transformer models have demonstrated exceptional performance in natural language processing, computer vision, and other tasks [37], [38], [39], [40], [41], [42], [43] owing to their self-attention mechanism and powerful modeling capabilities. This makes them well-suited for processing high-dimensional, highly correlated hyperspectral image data. Several studies have explored Transformer-based approaches for hyperspectral CD. For example, Song et al. [44] proposed a cross-temporal spectral attention network. Wang et al. [45] introduced an integrated spectral, spatial, and temporal Transformer model. Li et al. [46] adopted a cross-band 2-D self-attention network to extract spectral differences. Dong et al. [47] exploited local-global and spectral-spatial features using cascaded modules and a novel graph converter. Dong et al. [48] improved CD by modeling mixed pixels and utilizing dual-temporal correlation differences. Together, these methods showcase the potential of deep learning, especially Transformer models, for improving CD in hyperspectral images.

However, with the ongoing development of hyperspectral imaging technology [49], [50], [51], [52], [53], [54], [55], image size and resolution are constantly increasing, which also increases the demand for computing resources. At the same time, practical applications require compact, lightweight models suitable for situations with limited computing resources. Existing Transformer-based methods for hyperspectral CD require ample computational resources and extensive training time due to their reliance on the multihead self-attention mechanism. The high complexity makes these methods prone

to overfitting on limited training data. On the other hand, hyperspectral images commonly contain substantial redundant and repeated data. While past research often focuses on optimizing spectral dimension redundancy, it is noted that hyperspectral images actually contain redundancy in both the spatial and spectral domains.

Motivated by these observations, this study proposes a new gated spectral-spatial-temporal attention network with spectral similarity filtering (HyGSTAN) that addresses high-resolution hyperspectral imagery requirements without compromising classification accuracy while maintaining algorithmic efficiency. The proposed model comprises several key components. Initially, a spectral similarity filtering module (SSFM) is designed to mitigate spectral redundancy within patches. It effectively leverages cosine similarity to amalgamate similar spectral information within patches before undergoing attention computation. This is followed by the application of a lightweight gated spectral-spatial attention module (GS²AM) purposefully designed to efficiently extract spectral information with long-range dependencies. Subsequently, a gated spectral-spatial-temporal attention module (GS²TAM) is implemented to integrate time-series spectral data from dual-temporal-phase hyperspectral images. Ultimately, the inverse operation of the spectrum merging module reconstructs the merged parts' change information. The main contributions of this article can be summarized as follows.

- 1) A uniquely designed SSFM efficiently reduces spectral redundancy within patches. This component not only sustains detection accuracy but also significantly decreases the data volume entering the classifier, thereby minimizing model parameters and computational complexity while augmenting inference speed.
- 2) A new GS²AM is designed to extract autocorrelation information in single-phase spectra through a gating mechanism and a single-head weak self-attention mechanism.
- 3) A new GS²TAM is designed to extract change features by extending in the temporal dimension through GS²AM, and combining the autocorrelation information in dual-temporal spectra.

Experimental validation on three widely used hyperspectral datasets demonstrates the network's robustness and exceptional generalization performance, even with limited training samples.

II. METHODOLOGY

This section provides a comprehensive overview of the HyGSTAN's architectural framework and furnishes meticulous elucidations of each constituent module, as delineated in Fig. 1. The network encompasses the SSFM, the GS²AM, and the GS²TAM. Among them, SSFM includes SSFM-C and its reverse operation SSFM-R. Initially, bitemporal hyperspectral images $\mathbf{T}_1, \mathbf{T}_2 \in \mathbb{R}^{b \times h \times w}$ are provided, with b denoting the band count, and h and w indicating the height and width of the hyperspectral images. These images are subsequently partitioned into patches of dimensionality $l \times l$, hereafter designated as $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{b \times l \times l}$. Subsequently, the divided

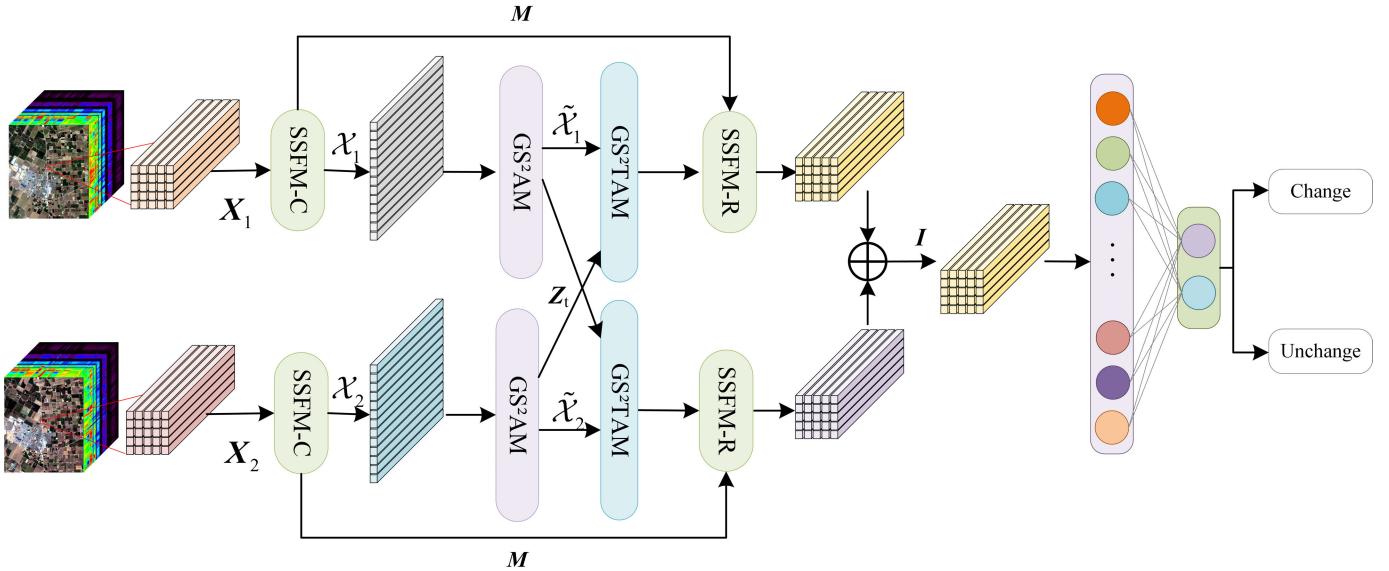


Fig. 1. Architecture of the proposed HyGSTAN for hyperspectral image CD.

images X_1 and X_2 are fed into the SSFM-C to conduct similarity feature selection, yielding features $\mathcal{X}_1, \mathcal{X}_2 \in \mathbb{R}^{b \times s}$, where s signifies the number of spectral features. Then, they are separately directed into the GS²AM for intra-image feature extraction, resulting in $\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2, \mathbf{Z}_1$ and \mathbf{Z}_2 , where \mathbf{Z} is the attention matrix. Following this stage, the two branches carry out cross-input of \mathbf{Z}_1 and \mathbf{Z}_2 into the GS²TAM to extract temporal discrepancy features between the bitemporal images. Subsequently, the obtained change features undergo position information restoration through the SSFM-R and are summed, resulting in the output denoted as $\mathbf{I} \in \mathbb{R}^{b \times l \times l}$. Finally, the change features \mathbf{I} are flattened into a 1-D vector $\mathbf{x} \in \mathbb{R}^m$, where m represents the length of \mathbf{x} with size $b \times l \times l$. Subsequently, these features undergo sequential processing through LeakyReLU, Linear, and SoftMax layers as follows:

$$\mathbf{y} = \max(\mathbf{x}, \alpha\mathbf{x}) \quad (1)$$

$$\hat{\mathbf{y}} = \text{SoftMax}(\mathbf{W}\mathbf{y} + \mathbf{b}). \quad (2)$$

Here, \mathbf{x} denotes the input, α represents a small positive slope, and \mathbf{W} and \mathbf{b} are learnable parameter matrices. The resulting $\hat{\mathbf{y}} \in [0, 1]$ represents the change probability of the patch. After obtaining the predictions of all patches, they are organized into a CD map $\mathbf{Y} \in \mathbb{R}^{h \times w}$. The inference detail of our model for HyGSTAN is shown in Algorithm 1. Subsequent sections will provide detailed explanations of each component within the network.

A. Spectral Similarity Filtering Module

As illustrated in Fig. 2, pixels within hyperspectral images often have a substantial number of neighboring pixels that exhibit high spectral similarity. The computation involving these similar neighboring pixels is frequently redundant. This redundancy poses a dual challenge: first, it imposes a computational burden on the classifier, leading to an increase in the classifier's parameter count. Second, when the proportion of

Algorithm 1 HyGSTAN's Architectural Framework for Hyperspectral Image CD

Data: Bitemporal hyperspectral images $\mathbf{T}_1, \mathbf{T}_2 \in \mathbb{R}^{b \times h \times w}$
Result: Prediction change detection map $\mathbf{Y} \in \mathbb{R}^{h \times w}$

```

1 for  $i \leftarrow 1$  to  $2$  do
2   |  $X[i] \leftarrow \text{Extract\_Patches}(\mathbf{T}[i], \text{patch\_size} = l \times l)$ ;
3 end
4 for  $i \leftarrow 1$  to  $2$  do
5   |  $\mathcal{X}[i] \leftarrow \text{SSFM-C}(X[i])$ ;
6 end
7 for  $i \leftarrow 1$  to  $2$  do
8   |  $(\tilde{\mathcal{X}}[i], \mathbf{Z}[i]) \leftarrow \text{GS}^2\text{AM}(\mathcal{X}[i])$ ;
9 end
10 for  $i \leftarrow 1$  to  $2$  do
11   | Discrepancy_Features[i]  $\leftarrow$ 
12     | |  $\text{GS}^2\text{TAM}(\tilde{\mathcal{X}}[i], \mathbf{Z}[3 - i])$ ;
12 end
13  $\mathbf{I} \leftarrow \text{SSFM-R}(\text{Discrepancy_Features}[1]) +$ 
14   | |  $\text{SSFM-R}(\text{Discrepancy_Features}[2])$ ;
14  $\mathbf{x} \leftarrow \text{Flatten}(\mathbf{I})$ ;
15  $\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{x}, \alpha)$ ;
16  $\hat{\mathbf{y}} \leftarrow \text{SoftMax}(\text{Linear}(\mathbf{y}))$ ;
17  $\mathbf{Y} \leftarrow \text{Reconstruct\_Map}(\hat{\mathbf{y}}, h, w)$ ;
```

similar neighboring pixels within the vicinity of a target pixel becomes substantial, it can adversely affect the classifier's ability to accurately discriminate the target pixel. Attempting to directly address these issues through clustering or similar methods may result in a loss of spatial information among pixels in subsequent feature layers. Given these considerations, we develop a module that leverages the cosine similarity of spectral information to perform spectral filtering operations within the patch, thereby mitigating these problems.

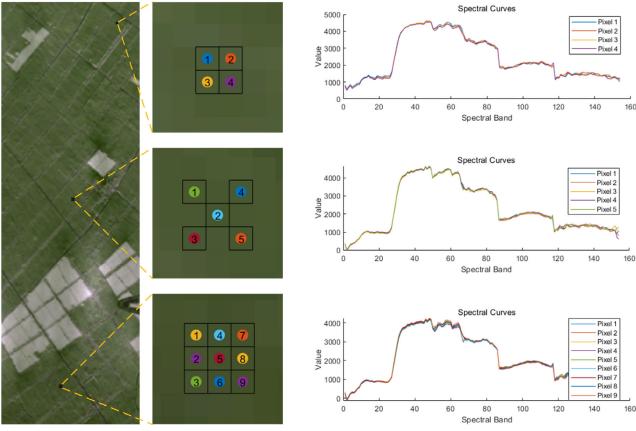


Fig. 2. Spectral similarity of neighboring pixels in hyperspectral images.

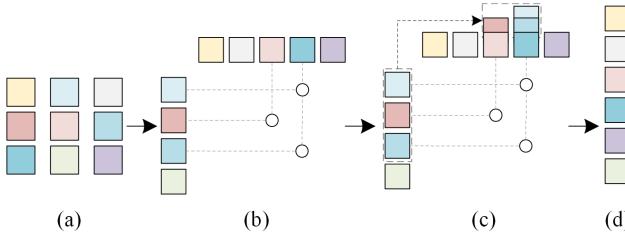


Fig. 3. Spectral similarity filtering module. (a) Original spectral vector input. (b) Extraction of index pairs for similar spectral vectors. (c) Averaging process for similar spectral pairs. (d) Resulting merged spectral vector.

As shown in Fig. 3, the SSFM-C first reorganizes the input spectral patch X into a set of vectors $\mathbf{x} \in \mathbb{R}^{m \times c}$, where c is the bandwidth and $m = h \times w$. It then performs L_2 normalization to scale the spectral features to a similar range, which can be expressed as

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\sqrt{\sum_{i=1}^c \mathbf{x}_i^2}} \quad (3)$$

where $\hat{\mathbf{x}}$ represents the L_2 -normalized input. Subsequently, the spectral features are rearranged into two sets

$$\mathcal{X}_1 = \{\mathbf{x}_1, \mathbf{x}_3, \dots, \mathbf{x}_{2n-1}\} \quad (4)$$

$$\mathcal{X}_2 = \{\mathbf{x}_2, \mathbf{x}_4, \dots, \mathbf{x}_{2n}\} \quad (5)$$

where \mathcal{X}_1 and \mathcal{X}_2 are the feature sets, and $n \in [1, m]$ denotes a feature index. Here, \mathbf{x}_i and \mathbf{y}_j denote the i th and j th feature vectors, respectively. For any $\mathbf{x}_i \in \mathcal{X}_1$ and $\mathbf{y}_j \in \mathcal{X}_2$, the cosine similarity is defined as

$$\text{sim}(\mathbf{x}_i, \mathbf{y}_j) = \frac{\sum (\mathbf{x}_i \cdot \mathbf{y}_j)}{\|\mathbf{x}_i\|_2 \cdot \|\mathbf{y}_j\|_2} \quad (6)$$

where \cdot is the vector inner product and $\|\cdot\|_2$ denotes the L_2 norm of the vector. Then, the cosine similarity matrix between the two sets can be expressed as

$$\mathbf{S} = [\text{sim}(\mathbf{x}_i, \mathbf{y}_j)]_{N \times M} \quad (7)$$

where the elements in the i th row and j th column, N and M are the dimensions of the \mathcal{X}_1 and \mathcal{X}_2 , and the $\text{sim}(\mathbf{x}_i, \mathbf{y}_j)$ is the cosine similarity between vectors \mathbf{x}_i and \mathbf{y}_j . The matrix $\mathbf{S} \in \mathbb{R}^{N \times M}$ contains the cosine similarity information between

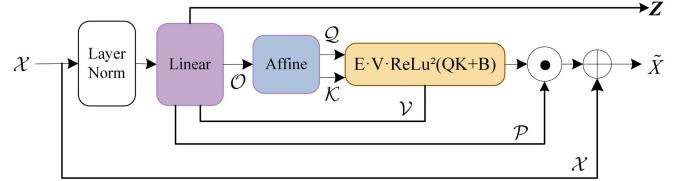


Fig. 4. Architecture of the GS²AM.

all pairs of vectors from sets \mathcal{X}_1 and \mathcal{X}_2 . We find the indices corresponding to the top p maximum values in \mathbf{S} as follows:

$$(i_k, j_k) = \arg \max_p \mathbf{S}, \quad k = 1, 2, \dots, p \quad (8)$$

where (i, j) denotes the index pair of the p elements of the matrix \mathbf{S} with the largest cosine similarity. Then, we extract the corresponding top p similar vector pairs \mathbf{x}_p and \mathbf{y}_p , such as the two connected blocks in Fig. 3(b)

$$\mathbf{x}_p = \mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_p} \subseteq \mathcal{X}_1 \quad (9)$$

$$\mathbf{y}_p = \mathbf{y}_{j_1}, \mathbf{y}_{j_2}, \dots, \mathbf{y}_{j_p} \subseteq \mathcal{X}_2. \quad (10)$$

After that, as shown in Fig. 3(c), the corresponding vector pairs are averaged and denoted as

$$\bar{\mathbf{x}}_k = \frac{\mathbf{x}_k + \mathbf{y}_k}{2}, \quad k = 1, 2, \dots, p. \quad (11)$$

Also, the indexes of the vectors involved in the merger are denoted as

$$\mathbf{M} = [(i_1, j_1), (i_2, j_2), \dots, (i_p, j_p)] \quad (12)$$

where \mathbf{M} denotes the index of the merge position for feature size reduction in SSFM-R. As shown in Fig. 3(d), finally, the remaining set of vectors is concatenated to output $\bar{\mathbf{X}}$. This process is denoted as

$$\bar{\mathbf{X}} = \mathcal{X}_1 \setminus \mathbf{x}_p \cup \mathcal{X}_2 \setminus \mathbf{y}_p \cup \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_p \quad (13)$$

where $\bar{\mathbf{X}} \in \mathbb{R}^{(m-p) \times c}$ is the final output feature that will be fed into the subsequent feature extraction module GS²AM.

B. Gated Spectral–Spatial Attention Module

As shown in Fig. 4, the GS²AM processes input features $\mathbf{X} \in \mathbb{R}^{N \times c}$, where N is the number of features, and c is the feature dimensionality. The module initially applies LayerNorm

$$\hat{\mathbf{X}} = \text{LayerNorm}(\mathbf{X}). \quad (14)$$

The normalized features $\hat{\mathbf{X}}$ then undergo a lightweight attention mechanism. First, a gating mechanism is applied

$$\mathcal{O} = g_o(\hat{\mathbf{X}} \mathbf{W}_o) \quad (15)$$

$$\mathcal{P} = g_p(\hat{\mathbf{X}} \mathbf{W}_p) \quad (16)$$

$$\mathcal{V} = g_v(\hat{\mathbf{X}} \mathbf{W}_v) \quad (17)$$

where $\mathcal{O} \in \mathbb{R}^{N \times d}$, $\mathcal{P} \in \mathbb{R}^{N \times c}$, and $\mathcal{V} \in \mathbb{R}^{N \times c}$ are the gating vectors that are produced by linear transformation. $g_o()$, $g_p()$, and $g_v()$ represent activation functions. $\mathbf{W}_o \in \mathbb{R}^{c \times d}$, $\mathbf{W}_i \in \mathbb{R}^{c \times c}$, and $\mathbf{W}_j \in \mathbb{R}^{c \times c}$ are the weight matrices. The queries \mathcal{Q} and keys \mathcal{K} are generated by an affine transformation of \mathcal{O}

$$\mathcal{Q} = \gamma_q \cdot \mathcal{O} + \beta_q \quad (18)$$

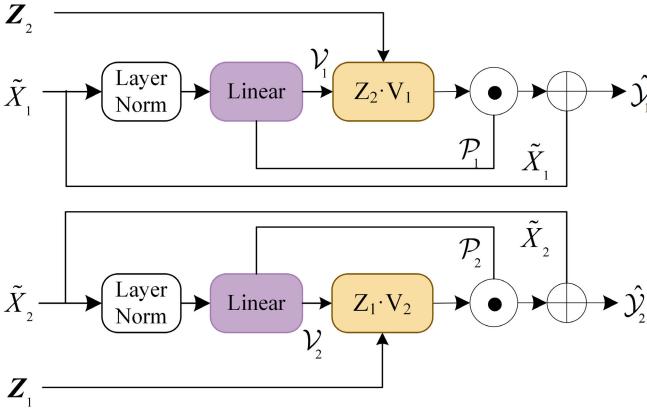


Fig. 5. Architecture of the GS²TAM.

$$\mathcal{K} = \gamma_k \cdot \mathcal{O} + \beta_k \quad (19)$$

where γ_q , γ_k , β_q , and β_k are learnable parameters. Next, the attention matrix Z is computed using \mathcal{Q} and \mathcal{K}

$$Z = \frac{1}{n} \text{relu}^2 \left(\frac{\mathcal{Q} \cdot \mathcal{K}^\top}{\sqrt{d}} + b \right). \quad (20)$$

Next, the multiplication between the attention matrix Z and \mathcal{V} [from (17)] yields the following result:

$$\hat{\mathcal{V}} = Z \cdot \mathcal{V}. \quad (21)$$

In parallel, we have

$$\mathcal{Y} = \mathcal{P} \odot \hat{\mathcal{V}} \quad (22)$$

$$\tilde{X} = \mathcal{Y} \cdot W_y + \mathcal{X}. \quad (23)$$

In summary, the GS²AM employs linear projections to generate query, key, value, and gating vectors, followed by a lightweight attention operation and gated combination. The gate vectors acting as adaptive weights, allowing the model to dynamically adjust to different features, locations, and dimensions in the input space importance. This adaptability enhances the model's ability to capture complex relationships and dependencies in the data. The combination of lightweight attention operations and gating ensures that the model not only focuses on relevant spectral features, but also selectively emphasizes or deemphasizes specific aspects. This integrated design efficiently extracts both attended and gated spectral features.

C. Gated Spectral–Spatial–Temporal Attention Module

While GS²AM focuses on extracting intrinsic spectral–spatial features within a single hyperspectral image, GS²TAM aims to effectively model relationships between bitemporal images for CD. Specifically, as shown in Fig. 5, given two input feature maps $\mathcal{X}_{t1}, \mathcal{X}_{t2} \in \mathbb{R}^{N \times d}$ from the bitemporal images, where N is the number of tokens and d is the number of channels, with t indicating different temporal capture phases of the images. Separate GS²AM modules are first applied to extract intrinsic representations for each image

$$\tilde{X}_t, Z_t = \text{GS}^2\text{AM}(\mathcal{X}_t), \quad t \in t1, t2. \quad (24)$$

The GS²AM modules transform the input into gated spectral–spatial feature embeddings $\tilde{X}_{t1}, \tilde{X}_{t2} \in \mathbb{R}^{N \times d}$, encoding intrinsic characteristics within each image via attention and gating mechanisms. Then, normalization is performed by LayerNorm

$$\hat{X}_t = \text{LayerNorm}(\tilde{X}_t), \quad t \in t1, t2. \quad (25)$$

Next, the activation function and linear variation are utilized to generate \mathcal{P}_t and \mathcal{V}_t

$$\mathcal{P}_t = g_p^t(\hat{X}_t \mathbf{W}_p^t), \quad t \in t1, t2 \quad (26)$$

$$\mathcal{V}_t = g_v^t(\hat{X}_t \mathbf{W}_v^t), \quad t \in t1, t2. \quad (27)$$

Next, cross-attention is performed between the GS²AM outputs to capture the relationships between the bitemporal images. The attention matrix Z_t is derived from the first image's GS²AM output, while the values \mathcal{V}_t come from the second image

$$\hat{\mathcal{V}}_{t1} = Z_{t2} \cdot \mathcal{V}_{t1} \quad (28)$$

$$\hat{\mathcal{V}}_{t2} = Z_{t1} \cdot \mathcal{V}_{t2} \quad (29)$$

where $\hat{\mathcal{V}}_{t1}$ and $\hat{\mathcal{V}}_{t2}$ are context vector matrices. This cross-attention module allows us to model relationships between the bitemporal images. Then, the information is filtered through the gating mechanism

$$\mathcal{Y}_t = \mathcal{P}_t \odot \hat{\mathcal{V}}_t \quad t \in t1, t2 \quad (30)$$

$$\hat{Y}_t = \mathcal{Y}_t \cdot W_y^t + \mathcal{X}_t \quad t \in t1, t2. \quad (31)$$

The gating selects informative cross-temporal features while suppressing irrelevant ones. The integrated output \hat{Y}_t contains both intrinsic and cross-temporal characteristics for CD. In summary, GS²TAM provides an effective approach to learn spectral, spatial, and temporal relationships from bitemporal images. The hybrid feature representations are tailored for CD between the image pairs.

D. Loss Function

An effective loss function is critical for CD model performance. The following loss function is used in our work:

$$L = -(1 - P_t)^\delta \log(P_t) + \lambda(1 - P_t)^{\delta+1} \quad (32)$$

$$P_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases} \quad (33)$$

where P_t is the predicted probability for the true label y . The parameter δ adjusts the loss weights for easy and difficult examples. The perturbation term λ alleviates overfitting and improves model robustness and generalization. Imbalanced difficult and easy samples are common in CD scenarios. By adaptively weighting samples via δ , difficult examples with lower confidence produce larger losses. This assigns higher weights to misclassified and hard-to-classify examples.

III. EXPERIMENTAL RESULTS

This section delves into the comparative experiments performed to appraise the efficiency of the proposed HyGSTAN in the context of hyperspectral CD. Initially,

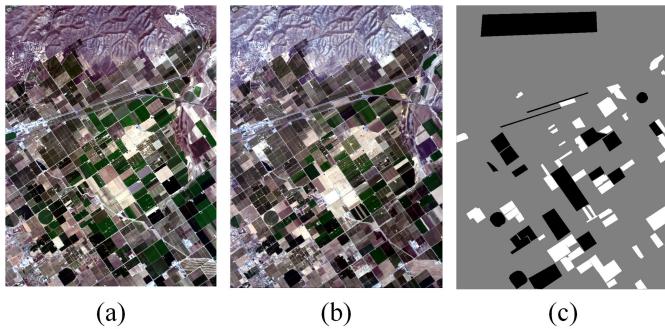


Fig. 6. Santa Barbara dataset. (a) Image acquired in 2013. (b) Image acquired in 2014. (c) Reference map.



Fig. 7. Bay area dataset. (a) Image acquired in 2013. (b) Image acquired in 2015. (c) Reference map.

Section III-A presents three frequently utilized public hyperspectral datasets. Subsequently, Section III-B elaborates on the experimental framework, which includes comparative approaches, assessment criteria, and parameter adjustments. Section III-C provides a comprehensive comparison of the proposed method's performance alongside seven distinct CD techniques utilized on the three previously discussed hyperspectral datasets. Subsequently, Section III-D delves into the scrutinizing of the effect of hyperparameters on the model's performance. To gain a deeper understanding of the performance improvements brought on by the introduction of this proposed method, an ablation study is depicted in Section III-E.

A. Datasets

1) *Santa Barbara Dataset*: As shown in Fig. 6, the Santa Barbara¹ dataset was obtained by the airborne visible/infrared imaging spectrometer (AVIRIS) over Santa Barbara, California, USA, in 2013 and 2014. Both images have a spatial resolution of 20 m with 224 spectral bands covering the 0.4–2.5 μm wavelength range. The images are 984 \times 740 pixels in size. This dataset has 52 134 labeled changed pixels and 80 418 unchanged pixels.

2) *Bay Area Dataset*: As shown in Fig. 7, the Bay Area¹ dataset was acquired by AVIRIS over Patterson, California, USA, in 2013 and 2015. It has the same number of bands and spectral range as the Santa Barbara dataset, with a spatial resolution of 20 m and an image size of 600 \times 500 pixels. This dataset has 38 425 annotated changed pixels and

¹<https://citius.usc.es/investigacion/datasets/hyperspectral-change-detection-dataset>

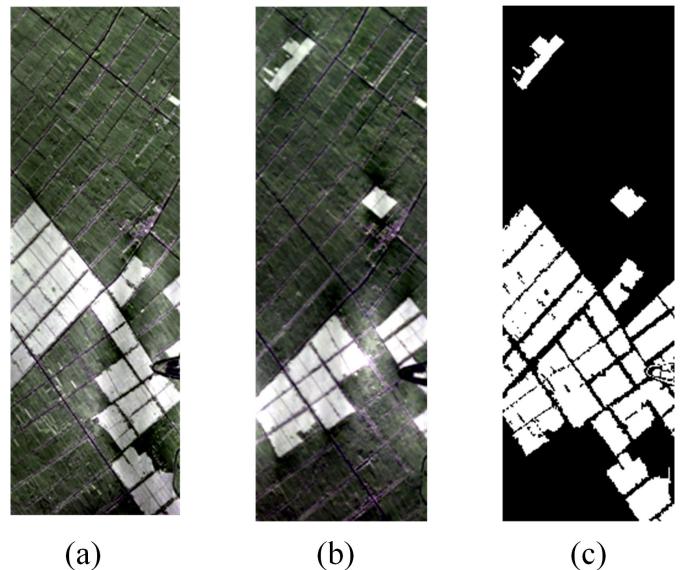


Fig. 8. Farmland dataset. (a) Image acquired on May 3, 2006. (b) Image acquired on April 23, 2007. (c) Reference map.

34 211 unchanged pixels. The three public datasets cover typical urban and agricultural areas and can well validate the effectiveness and generalization performance of the proposed hyperspectral CD method.

3) *Farmland Dataset*: As shown in Fig. 8, the Farmland² dataset was acquired by the Hyperion sensor on the Earth Observing-1 satellite over a farmland area near Yancheng City, Jiangsu Province, China. The two HSI were captured on May 3, 2006 and April 23, 2007. After denoising, each image contains 154 spectral bands with a spatial resolution of 30 m and an image size of 420 \times 140 pixels. The reference ground truth labels 18 383 changed pixels and 40 417 unchanged pixels.

B. Experimental Setting

1) *Comparison Methods and Parameter Setting*: In order to verify the effectiveness of the proposed HyGSTAN, we compare the proposed method with other CD methods, including the classic CVA, DPCA³ methods, and open source deep learning-based methods in recent years, including BIT,⁴ CSA-Net,⁵ ML-EDAN,⁶ SST-Former.⁷ The hyperparameters of these methods are consistent with those in the original papers. The methodology presented in this work was executed using the PyTorch 2.0.1 framework on an Intel Core i7-11700k CPU and an NVIDIA GeForce RTX 3060 12 GB GPU. The optimization process employed the AdamW algorithm with an initial learning rate of 0.0005. A batch size of 512 samples was used during the training process. To adjust the learning rate, a learning rate scheduler was implemented, which attenuates the learning rate by a multiplication factor of 0.9 after every one-twentieth of the total epochs. The entire training period

²<https://rslab.ut.ac.ir/data>

³<https://github.com/Bobholamovic/ChangeDetectionToolbox>

⁴https://github.com/justchenhao/BIT_CD

⁵<https://github.com/SYFYNO317/CSA-Net>

⁶<https://github.com/Jiahuiqu/ML-EDAN>

⁷https://github.com/yanhengwang-heu/IEEE_TGRS_SSTFormer

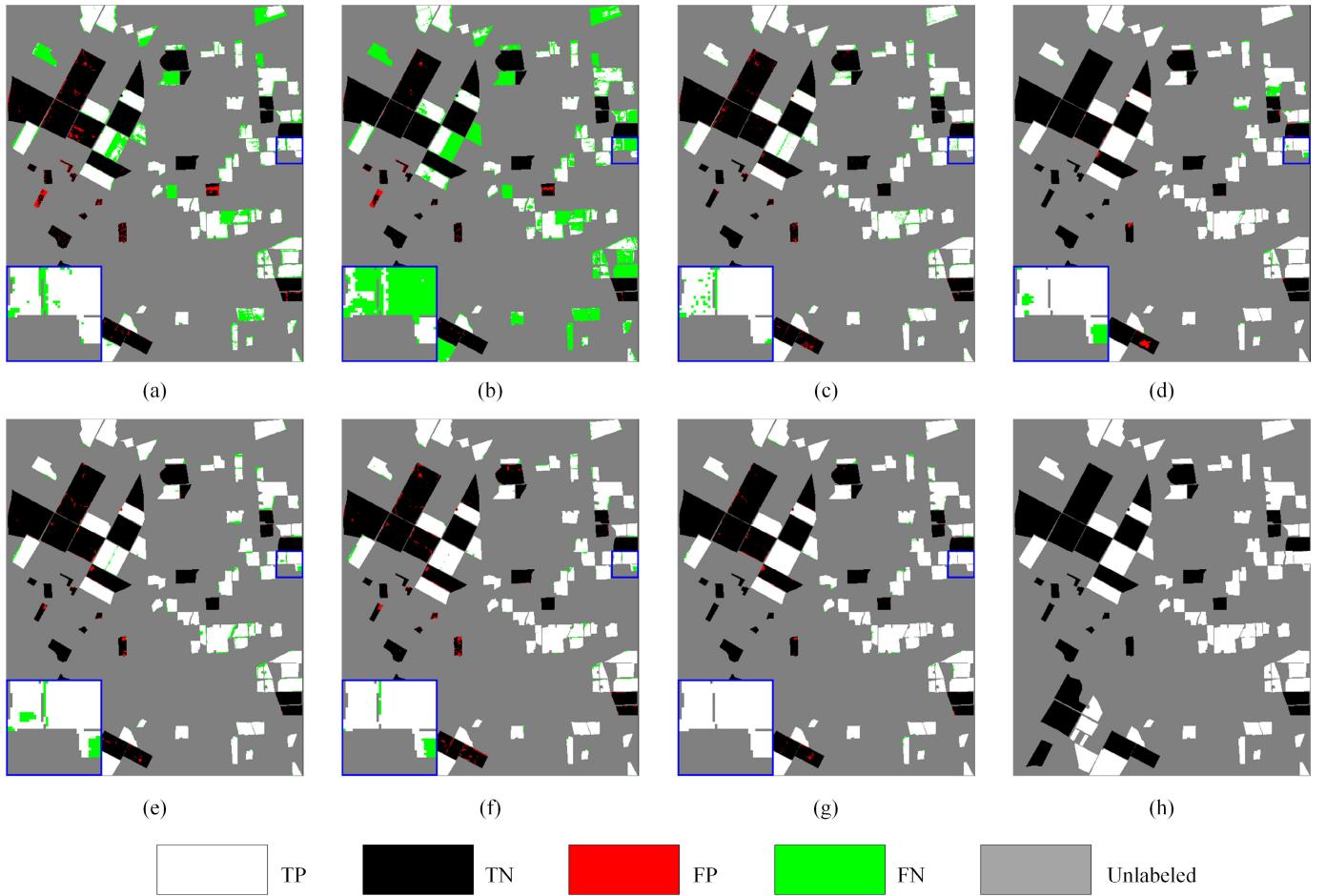


Fig. 9. CD maps for the Bay dataset. (a) CVA. (b) DPCA. (c) BIT. (d) CSA-Net. (e) ML-EDAN. (f) SST-Former. (g) Ours. (h) Ground truth.

consists of 100 epochs. The sample is set to randomly select 1% of the pixels that change and do not change.

2) *Evaluation Metrics*: To evaluate the efficacy of the different methods, commonly used quantitative metrics were employed, including overall accuracy (OA), Kappa coefficient (Kappa), F1-score (F1), Precision (Pr), and Recall (Re). These are computed as follows:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (34)$$

$$Pr = \frac{TP}{TP + FP} \quad (35)$$

$$Re = \frac{TP}{TP + FN} \quad (36)$$

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re} \quad (37)$$

$$\text{Kappa} = \frac{OA - P_c}{1 - P_c} \quad (38)$$

$$P_c = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + FP + TN + FN)^2}. \quad (39)$$

Here, the metrics utilized to evaluate the algorithm's performance encompass four intermediary indices. These are: true positives (TPs), denoting the count of accurately detected changed pixels; true negatives (TNs), representing the number of correctly identified unchanged pixels; false positives (FPs),

accounting for the instances of false-alarm pixels; and false negatives (FNs), indicating the amount of missed changed pixels. In addition to these metrics, floating-point operations per second (FLOPs) quantify the number of FLOPs required during the model's inference. The total number of parameters reflects the model's complexity, influencing its memory requirements and training duration.

C. Comparison Experiments

1) *Results Analysis for Bay Area Dataset*: Fig. 9 shows the CD results for the Bay Area dataset. In order to facilitate visual comparison, we set TP to white, TN to black, FP and FN to red and green, respectively, and the lower left corner is a local enlarged image. Table I provides a comprehensive benchmark of CD accuracy metrics for several state-of-the-art techniques on the Bay Area dataset. HyGSTAN outperforms all other techniques with all five metrics, demonstrating the superiority of HyGSTAN in this application. In addition, methods based on deep learning are generally superior to traditional methods such as CVA and DPCA in terms of indicators, reflecting the advantages of deep learning technology over traditional methods in hyperspectral image CD tasks under certain conditions.

Among deep learning technologies, HyGSTAN achieves 98.59% OA. Compared with the second highest CSA-Net, it is 1.07% higher, indicating that the prediction map of HyGSTAN is closer to the ground truth. This is further substantiated by

TABLE I
CD RESULTS FOR THE BAY AREA DATASET

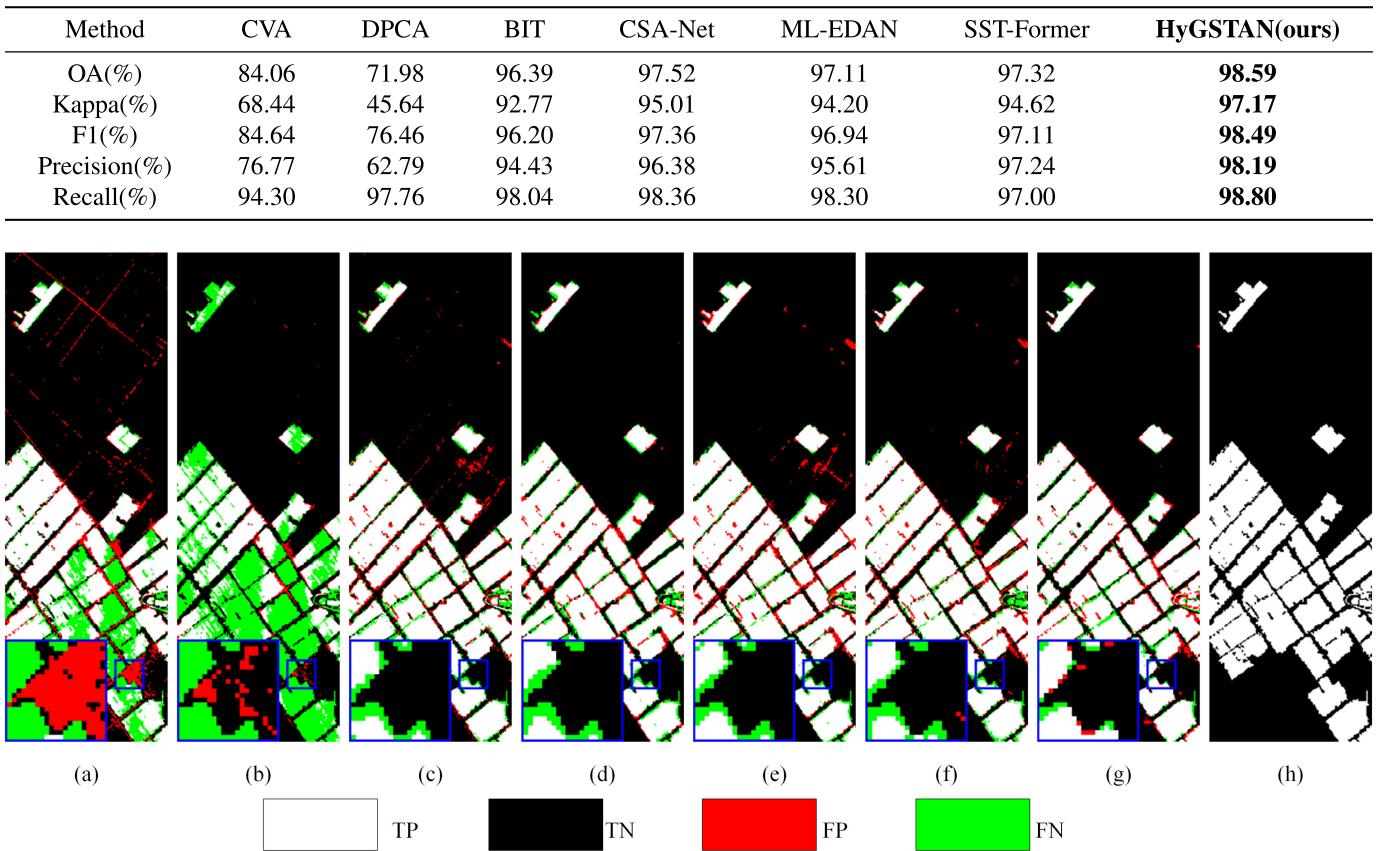


Fig. 10. CD maps for the Farmland dataset. (a) CVA. (b) DPCA. (c) BIT. (d) CSA-Net. (e) ML-EDAN. (f) SST-Former. (g) Ours. (h) Ground truth.

TABLE II
CD RESULTS FOR THE FARMLAND DATASET

Method	CVA	DPCA	BIT	CSA-Net	ML-EDAN	SST-Former	HyGSTAN
OA(%)	87.49	81.00	94.07	94.61	94.08	94.38	95.11
Kappa(%)	69.98	48.09	86.26	87.41	86.43	87.06	88.72
F1(%)	78.82	58.25	90.58	91.32	90.79	91.17	92.30
Precision(%)	83.71	93.06	90.06	91.97	88.36	89.64	90.95
Recall(%)	74.47	42.39	91.10	90.67	93.35	92.75	93.67

the visual observation of the images, particularly Fig. 9(g). Analysis of the Kappa places consistency over chance, further validating HyGSTAN's capabilities. Kappa is 97.17%, 2.06% higher than CSA-Net, which reflects HyGSTAN's robustness to change recognition and has better observation consistency and reliability. Judging from the F1-score, HyGSTAN is 1.13% higher than the second-highest CSA-Net, achieving a better balance between FPs and FNs. In terms of accuracy, HyGSTAN is 0.95% higher than the second highest SST-Former. HyGSTAN has fewer FPs. From the image point of view, there are relatively fewer red areas in Fig. 9(g), which shows an advantage in identifying changed pixels. Compared to CSA-Net, HyGSTAN exhibits a 0.44% increase in OA, with a lower missed detection rate as evidenced by the reduced

green area. This is further demonstrated in the enlarged image, where DPCA has the most substantial missed detections while Fig. 9(g) has the least.

2) *Results Analysis for Farmland Dataset:* As shown in Table II, five indicators of different methods on the Farmland dataset are provided, and Fig. 10 is a visual result on the Farmland dataset. On this dataset, HyGSTAN again achieves the highest accuracy of 95.11% OA and 88.72% Kappa, outperforming all other methods. Compared to the second-best performing CSA-Net, HyGSTAN improves OA by 0.5% and Kappa by 1.31%. Observing the image, it can be seen that the results in Fig. 10(g) are better than those provided by other methods in processing the boundary in the enlarged area in the lower leftmost corner. In terms of accuracy, DPCA reaches

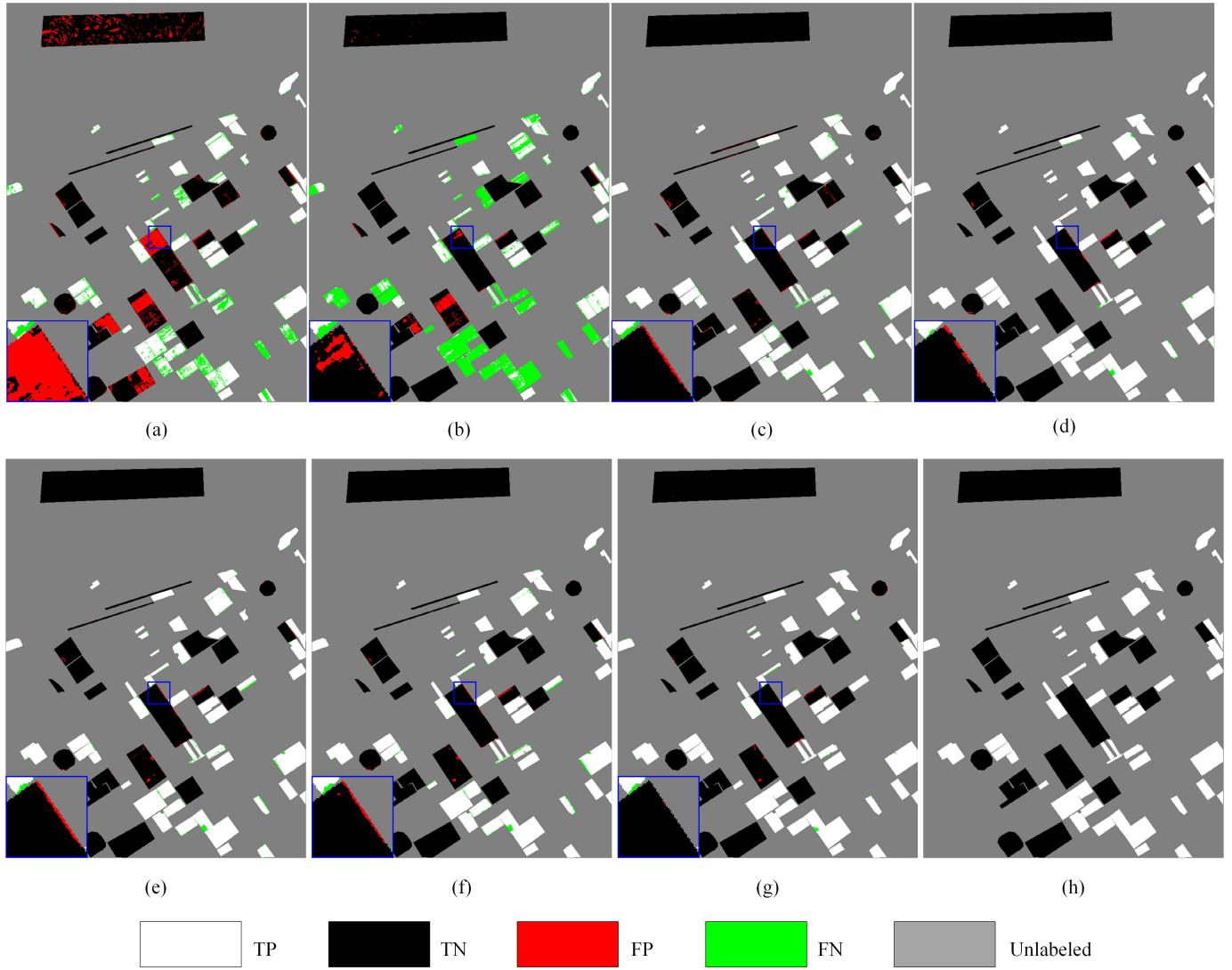


Fig. 11. CD maps for the Santa Barbara dataset. (a) CVA. (b) DPCA. (c) BIT. (d) CSA-Net. (e) ML-EDAN. (f) SST-Former. (g) Ours. (h) Ground truth.

93.06%, but its recall rate is only 42.39%. This is because DPCA misses most of the changing areas. It can be seen from Fig. 10(b) that there are many missed changed areas.

3) *Results Analysis for Santa Barbara Dataset:* Fig. 11 and Table III present visual and quantitative results obtained from experiments on the Santa Barbara dataset. An examination of the resulting images reveals considerable red and green regions in the outputs of both CVA and DPCA methods. The prevalence of red indicates that CVA exhibits a propensity for FP errors, while the green regions suggest that DPCA is prone to FN results. Among the deep learning techniques, relatively high accuracy is achieved overall. Scrutinizing the enlarged image sections in the lower leftmost corners of sub-figures (c), (d), (e), and (f), one can discern that subfigure (g) corresponds to the proposed HyGSTAN method. The minimal red area in (g) shows that HyGSTAN avoids the FP errors that manifest prominently in its counterparts. This improved boundary delineation can be attributed to the gated attention mechanism in HyGSTAN, which enhances contextual feature learning both spectrally and spatially. By capturing relevant

intra-image characteristics, HyGSTAN effectively minimizes false detections. The quantitative results in Table III corroborate these visual inferences, with HyGSTAN achieving the highest OA, Kappa, and F1-score. The consistent superiority of HyGSTAN across both visual and numeric assessments on the Barbara dataset verifies its effectiveness for hyperspectral image CD.

4) *Analysis of Model Efficiency of Deep Learning Methods:* Table IV compares the deep learning-based methods in terms of FLOPs, the number of parameters, and the time per single epoch. HyGSTAN achieves the lowest FLOP count at 4.24 MB, representing only 3% of the performance of the second-best CSA-Net method. Similarly, HyGSTAN boasts the smallest model size, with parameters totaling a mere 0.18 MB, compared to other methods which range from 14 to 623× larger. Lastly, HyGSTAN also demonstrates superior computational efficiency, with a training time of 0.0681 seconds per single epoch, which is 3-5× faster than other methods. Overall, the results in Table IV underscore that the proposed HyGSTAN method outperforms other methods across the

TABLE III
CD RESULTS FOR THE SANTA BARBARA DATASET

Method	CVA	DPCA	BIT	CSA-Net	ML-EDAN	SST-Former	HyGSTAN
OA(%)	83.25	75.66	97.72	99.05	98.46	98.30	99.29
Kappa(%)	65.19	44.02	95.21	98.01	96.76	96.43	98.51
F1(%)	85.98	82.76	98.13	99.22	98.72	98.61	99.42
Precision(%)	87.33	72.55	97.60	99.15	98.31	98.11	99.37
Recall(%)	84.66	96.32	98.66	99.28	99.17	99.11	99.46

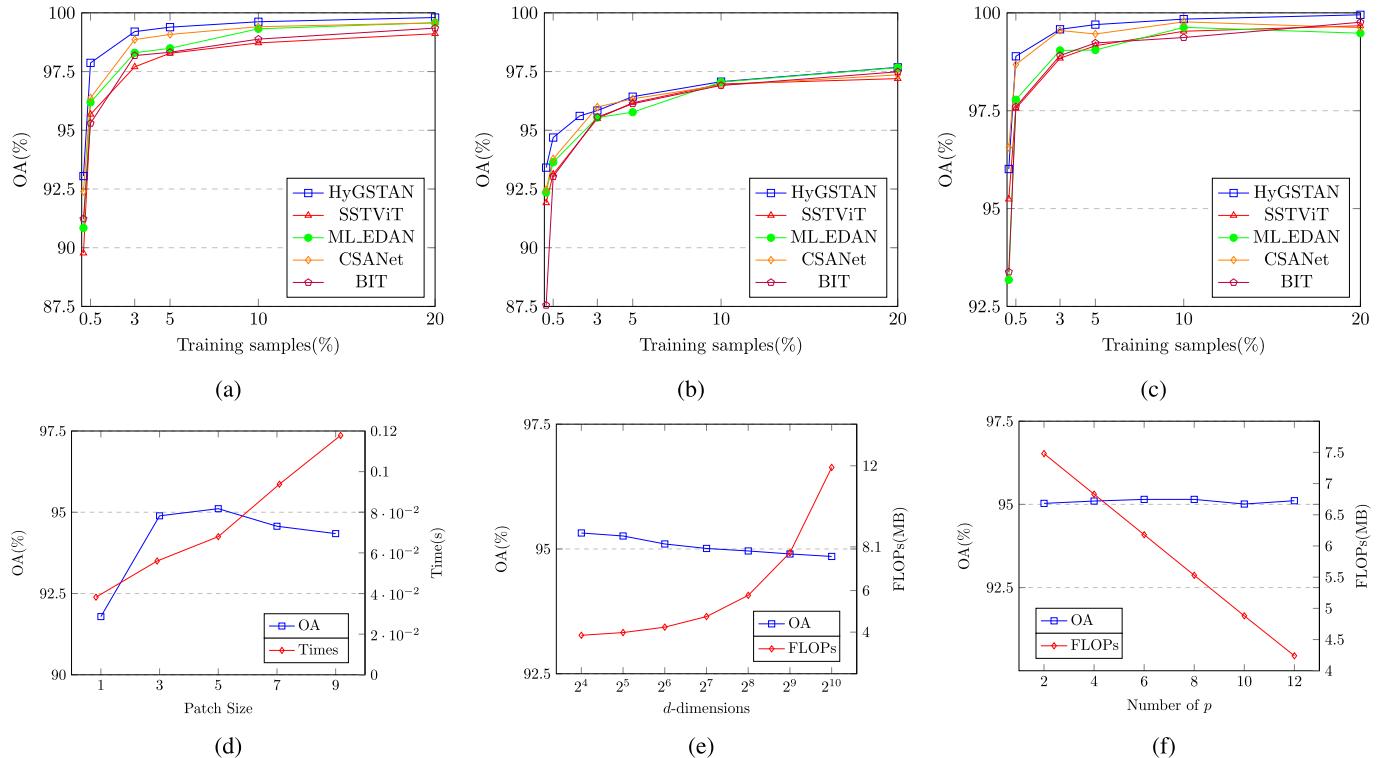


Fig. 12. Experiments and results of HyGSTAN under different parameter settings, including (a) results with different training samples on the BayArea dataset, (b) results with different training samples on the Farmland dataset, (c) results with different training samples on the Santa Barbara dataset, (d) OA and single epoch training time under different patch sizes, (e) OA and FLOPs with different values of d , and (f) OA and FLOPs with different values of p .

TABLE IV
COMPARISON OF FLOPs, PARAMETERS, AND PROCESSING
TIMES OF DIFFERENT METHODS

Method	FLOPs(MB)	Parameters(MB)	Times(s)
BIT	17.76	12.42	0.3939
CSA-Net	140.23	40.42	0.5434
ML-EDAN	274.78	112.29	1.5732
SST-Former	114.16	2.50	0.3762
HyGSTAN	4.24	0.18	0.0681

three evaluation metrics of FLOPs, parameters, and inference time.

D. Effect of Hyperparameters

To gain a deeper understanding of the performance of HyGSTAN, we conduct a comprehensive study on the hyperparameters of HyGSTAN, which includes the size of the

training set, PatchSize, the size of the attention matrix d , and the number of merged spectra p .

1) *Training Set Ratio*: Experiments were conducted on three datasets for five deep learning-based methods using eight training set proportions: {0.1%, 0.5%, 1%, 2%, 3%, 5%, 10%, 20%}. Fig. 12(a), (b), and (c) presents the results, demonstrating that the OA increases with more training samples. Notably, HyGSTAN consistently achieves the highest precision across all proportions.

2) *Patch Sizes*: Experiments were conducted using five patch sizes: {1, 3, 5, 7, 9} as shown in Fig. 12(c), training time gradually increased with larger patch sizes, while training accuracy initially improved and then declined, peaking at a patch size of 5.

3) *Dimension d* : The dimension d is associated with the size of the attention matrix from (20). To determine the impact of d , experiments were conducted with $d = \{16, 32, 64, 128, 256, 512, 1024\}$. Fig. 12(e) shows that while the OA did not improve with larger d , FLOPs rapidly increased.

TABLE V
ABLATION EXPERIMENT RESULTS (SSFM)

	Base	Without-SSFM
OA (%)	95.11	94.86
Kappa (%)	88.72	87.95
F1 (%)	92.3	91.66
Precision (%)	90.95	93.03
Recall (%)	93.67	90.33
FLOPs (MB)	4.24	8.12
Parameters (MB)	0.18	0.18
Times (s)	0.068	0.0953

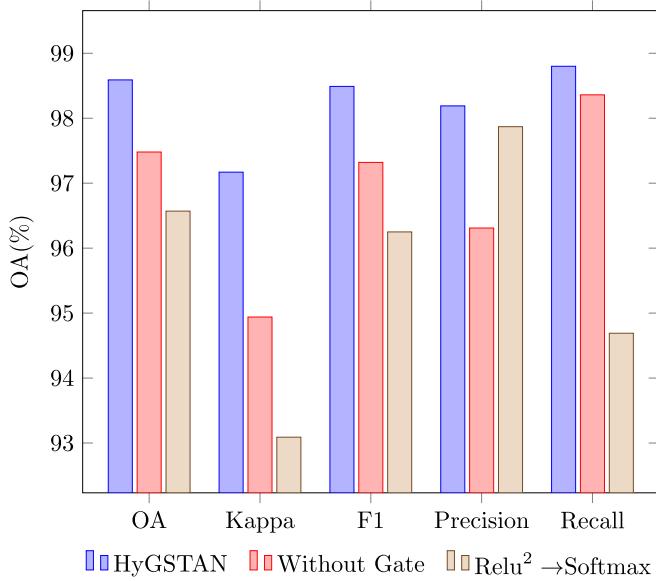


Fig. 13. HyGSTAN ablation experiments and results under different settings on the Bay Area dataset.

4) *Combined Quantity*: The parameter p represents the number of similar pixels merged in the SSFM module from (8). Experiments were conducted with p values of {2, 4, 6, 8, 10, 12}, as shown in Fig. 12(f). The results demonstrate that the OA remained relatively stable as p increased, while FLOPs decreased linearly with higher values of p .

E. Ablation Experiment

To thoroughly examine the effectiveness of individual modules, this article conducted three ablation experiments as follows.

1) *SSFM Ablation*: The SSFM module was removed from the HyGSTAN architecture, denoted as HyGSTAN-S. As shown in Table V, compared to HyGSTAN-S, HyGSTAN exhibited slightly higher OA, Kappa, and F1 metrics, along with a 3.34% increase in Recall, reducing missed detections. Additionally, HyGSTAN's computational load, measured in FLOPs, was only 52.22% of HyGSTAN-S, with a 28.31% reduction in training time.

2) *Gating Vector \mathcal{P}* : Excluding the gating vector \mathcal{P} from the original network led to significant declines in all performance metrics, as shown in Fig. 13. This indicates the detrimental impact on model performance of removing the gating mechanism.

3) *Relu² Ablation*: Replacing Relu² with Softmax in the HyGSTAN architecture resulted in substantial decreases across all performance metrics, as depicted in Fig. 13.

IV. CONCLUSION

This article proposes a new HyGSTAN for hyperspectral image CD in remote sensing applications. The incorporation of an SSFM, a GS²AM, and a gated spectral–spatial–temporal attention module enables the proposed HyGSTAN to effectively minimize redundancy, capture relevant intra-image features, and extract inter-image change features. Evaluations on benchmark datasets demonstrate excellent generalization ability, making HyGSTAN suitable for real applications. The model achieves an excellent balance between accuracy, complexity, and computational efficiency. Experiments also show that spectral similarity filtering helps classifier discrimination while reducing computing needs. Furthermore, the gated attention mechanism extracts more effective information without the calculation burden of multihead attention. Future work will extend HyGSTAN to long-term and multiphase CD applications. We will also explore image-level approaches to avoid redundant patch-based processing.

REFERENCES

- [1] P. Shanmugapriya, S. Rathika, T. Ramesh, and P. Janaki, “Applications of remote sensing in agriculture—A review,” *Int. J. Current Microbiol. Appl. Sci.*, vol. 8, no. 1, pp. 2270–2283, 2019.
- [2] W. K. Smith et al., “Remote sensing of dryland ecosystem structure and function: Progress, challenges, and opportunities,” *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111401.
- [3] S. Liu, M. Chi, Y. Zou, A. Samat, J. A. Benediktsson, and A. Plaza, “Oil spill detection via multitemporal optical remote sensing images: A change detection perspective,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 324–328, Mar. 2017.
- [4] M. Guo, J. Li, C. Sheng, J. Xu, and L. Wu, “A review of wetland remote sensing,” *Sensors*, vol. 17, no. 4, p. 777, 2017.
- [5] S. Thomas et al., “Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective,” *J. Plant Diseases Protection*, vol. 125, no. 1, pp. 5–20, Feb. 2018.
- [6] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, “A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges,” *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
- [7] B. Zhang, X. Sun, L. Gao, and L. Yang, “Endmember extraction of hyperspectral remote sensing images based on the ant colony optimization (ACO) algorithm,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2635–2646, Jul. 2011.
- [8] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, “Deep learning for hyperspectral image classification: An overview,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [9] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, “Adaptive Markov random field approach for classification of hyperspectral imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.
- [10] Z. Hou, W. Li, L. Li, R. Tao, and Q. Du, “Hyperspectral change detection based on multiple morphological profiles,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5507312.
- [11] M. Hu, C. Wu, and L. Zhang, “HyperNet: Self-supervised hyperspectral spatial-spectral feature understanding network for hyperspectral change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5543017.
- [12] M. Gong et al., “A spectral and spatial attention network for change detection in hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521614.
- [13] W. A. Malila, “Change vector analysis: An approach for detecting forest changes with Landsat,” in *Proc. LARS Symposia*, 1980, pp. 326–336.
- [14] D. Lu, P. Mausel, E. Brondizio, and E. Moran, “Change detection techniques,” *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Apr. 2004.

- [15] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008.
- [16] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
- [17] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.
- [18] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
- [19] C. Tao, J. Qi, M. Guo, Q. Zhu, and H. Li, "Self-supervised remote sensing feature learning: Learning paradigms, challenges, and future works," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5610426.
- [20] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.
- [21] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5518615.
- [22] H. Yu, Z. Xu, K. Zheng, D. Hong, H. Yang, and M. Song, "MSTNet: A multilevel spectral-spatial transformer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5532513.
- [23] W. Rao, L. Gao, Y. Qu, X. Sun, B. Zhang, and J. Chanussot, "Siamese transformer network for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526419.
- [24] H. Yang, H. Yu, K. Zheng, J. Hu, T. Tao, and Q. Zhang, "Hyperspectral image classification based on interactive transformer and CNN with multilevel feature fusion network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5507905.
- [25] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2020.
- [26] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, Aug. 2020.
- [27] K. Zheng et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2020.
- [28] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2022.
- [29] J. W. P. Sun, H. Li, W. Li, X. Meng, C. Ge, and Q. Du, "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 10–43, Jun. 2021.
- [30] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2018.
- [31] J. Qu, S. Hou, W. Dong, Y. Li, and W. Xie, "A multilevel encoder-decoder attention network for change detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518113.
- [32] J. Qu, Y. Xu, W. Dong, Y. Li, and Q. Du, "Dual-branch difference amplification graph convolutional network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5519912.
- [33] X. Ou, L. Liu, B. Tu, L. Qing, G. Zhang, and Z. Liang, "CBW-MSSANet: A CNN framework with compact band weighting and multiscale spatial attention for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507618.
- [34] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502713.
- [35] T. Li, T. Liu, Y. Wang, X. Li, and Y. Gu, "Spectral reconstruction network from multispectral images to hyperspectral images: A multitemporal case," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535016.
- [36] T. Li, T. Liu, X. Li, Y. Gu, Y. Wang, and Y. Chen, "Multi-sensor multispectral reconstruction framework based on projection and reconstruction," *Sci. China Inf. Sci.*, vol. 67, no. 3, Mar. 2024, Art. no. 132303.
- [37] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Proces. Syst. (NeurIPS)*, vol. 30, 2017, pp. 1–8.
- [38] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Apr. 2020.
- [39] K. Han et al., "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2022.
- [40] W. Hu et al., "Transformer quality in linear time," in *Proc. 39th Int. Conf. Mach. Learn.*, vol. 162, 2022, pp. 9099–9117.
- [41] D. Bolya, C.-Y. Fu, X. Dai, P. Zhang, C. Feichtenhofer, and J. Hoffman, "Token merging: Your ViT but Faster," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022, pp. 1–8.
- [42] X. Chen, B. Yan, J. Zhu, D. Wang, X. Yang, and H. Lu, "Transformer tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8126–8135.
- [43] Y. Huang et al., "Two-branch attention adversarial domain adaptation network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5540813.
- [44] R. Song, W. Ni, W. Cheng, and X. Wang, "CSANet: Cross-temporal interaction symmetric attention network for hyperspectral image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [45] Y. Wang et al., "Spectral–spatial–temporal transformers for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536814.
- [46] Y. Li et al., "CBA-Net: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513011.
- [47] W. Dong, Y. Yang, J. Qu, S. Xiao, and Y. Li, "Local information-enhanced graph-transformer for hyperspectral image change detection with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509814.
- [48] W. Dong et al., "Abundance matrix correlation analysis network based on hierarchical multihead self-cross-hybrid attention for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501513.
- [49] X. Wang, Q. Hu, Y. Cheng, and J. Ma, "Hyperspectral image super-resolution meets deep learning: A survey and perspective," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 8, pp. 1668–1691, Aug. 2023.
- [50] C. Deng, X. Luo, and W. Wang, "Multiple frame splicing and degradation learning for hyperspectral imagery super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8389–8401, 2022.
- [51] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1082–1096, 2020.
- [52] J. Li, K. Zheng, J. Yao, L. Gao, and D. Hong, "Deep unsupervised blind hyperspectral and multispectral data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [53] J. Li, K. Zheng, Z. Li, L. Gao, and X. Jia, "X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5518317.
- [54] J. Li et al., "Deep learning in multimodal remote sensing data fusion: A comprehensive review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102926.
- [55] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.



Haoyang Yu (Member, IEEE) received the B.S. degree in information and computing science from Northeastern University, Shenyang, China, in 2013, and the Ph.D. degree in cartography and geographic information system from the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing, China, in 2019.

He is currently an Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. His research interests include models and algorithms for hyperspectral image processing, analysis, and applications.



Hao Yang received the B.S. degree from Liaoning Petrochemical University, Fushun, China, in 2021. He is currently pursuing the master's degree with Dalian Maritime University, Dalian, China, under the supervision of Dr. H. Yu.

His research interests include hyperspectral image processing and deep learning.



Lianru Gao (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, CAS. He also has been a Visiting Scholar with the University of Extremadura, Cáceres, Spain, in 2014, and the Mississippi State University (MSU), Starkville, MS, USA, in 2016. In last ten years, he was the PI of ten scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2018–2020, 2022–2025, 2024–2028), and by the National Key Research and Development Program of China (2021–2025). He has published more than 240 peer-reviewed papers, and there are more than 150 journal papers included by Science Citation Index (SCI). He was the coauthor of three academic books including *"Hyperspectral Image Information Extraction."* He obtained 30 National Invention Patents in China. His research interests include hyperspectral image processing and information extraction.

Dr. Gao is a fellow of the Institution of Engineering and Technology. He was supported by the National Science Foundation for Distinguished Young Scholars of China in 2023, and was awarded the Outstanding Science and Technology Achievement Prize of the CAS in 2016, and won the Second Prize of The State Scientific and Technological Progress Award in 2018. He received the 2021 Outstanding Paper Award at the IEEE Workshop on Hyperspectral Image Processing: Evolution in Remote Sensing (WHISPERS). He is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IET Image Processing.



Jiaochan Hu received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2012, the M.S. degree in surveying and mapping engineering from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2015, and the Ph.D. degree in cartography and geographic information system from the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), in 2019.

She is currently an Associate Professor with the College of Environmental Sciences and Engineering, Dalian Maritime University, Dalian, China. Her research interests include optical remote-sensed data processing, analysis, and applications.



Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is a Full Professor and the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has reviewed more than 500 manuscripts for more than 50 different journals. He has authored more than 800 publications in this field, including 418 JCR journal papers, 25 book chapters, and 330 peer-reviewed conference proceeding papers. He has guest edited 17 special issues on hyperspectral remote sensing for different journals. His main research interests comprise hyperspectral data processing and parallel computing of remote sensing data.

25 book chapters, and 330 peer-reviewed conference proceeding papers. He has guest edited 17 special issues on hyperspectral remote sensing for different journals. His main research interests comprise hyperspectral data processing and parallel computing of remote sensing data.

Prof. Plaza is a fellow of IEEE “for contributions to hyperspectral data processing and parallel computing of Earth observation data,” a fellow of the Asia-Pacific Artificial Intelligence Association (AAIA), and an elected member of Academia Europaea, The Academy of Europe. He was also a member of the steering committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He was a recipient of the recognition of Best Reviewers of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010, for which he served as Associate Editor in 2007–2012. He was a recipient of the Best Column Award of IEEE Signal Processing Magazine in 2015, and the 2013 Best Paper Award of the JSTARS journal, and the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*. He received best paper awards at the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, the IEEE International Conference on Space Technology, and the IEEE Symposium on Signal Processing and Information Technology. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and a President of the Spanish Chapter of IEEE GRSS in 2012–2016. He is currently serving as the Chair for Publications Awards Committee of IEEE GRSS and as a Vice-Chair of the Fellow Evaluations Committee of IEEE GRSS. He is also an Associate Editor of IEEE ACCESS (receiving the recognition of Outstanding Associate Editor for the journal in 2017) and was also member of the Editorial Board of IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and *IEEE Geoscience and Remote Sensing Magazine* in 2013. He served as the Editor-in-Chief for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal for five years (2013–2017) and IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS (2019–2020). He has been included in the Highly Cited Researchers List from Clarivate Analytics from 2018 to 2022. Additional information: <http://sites.google.com/view/antonioplaza>.



Bing Zhang (Fellow, IEEE) received the B.S. degree in geography from Peking University, Beijing, China, in 1991, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 1994 and 2003, respectively.

He is a Full Professor and the Deputy Director of the Aerospace Information Research Institute, CAS, where he has been leading lots of key scientific projects in the area of hyperspectral remote sensing for more than 25 years. He has developed five soft-

ware systems in image processing and applications. His creative achievements were rewarded ten important prizes from the Chinese government, and special government allowances of the Chinese State Council. He has authored more than 400 publications, including more than 230 journal articles. He has edited ten books/contributed book chapters on hyperspectral image processing and subsequent applications. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas.

Dr. Zhang received the National Science Foundation for Distinguished Young Scholars of China in 2013, and the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of Awards for the CAS Scholars. He has been serving as a Technical Committee Member of IEEE Workshop on Hyperspectral Image and Signal Processing, since 2011, the President of Hyperspectral Remote Sensing Committee of China National Committee of International Society for Digital Earth since 2012, and has been the Standing Director of Chinese Society of Space Research (CSSR) since 2016. He is the Student Paper Competition Committee Member in International Geoscience and Remote Sensing Symposium (IGARSS) from 2015 to 2021. He is serving as an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.