A Hyperspectral Image Change Detection Framework With Self-Supervised Contrastive Learning Pretrained Model

Xianfeng Ou, Member, IEEE, Liangzhen Liu, Member, IEEE, Shulun Tan, Guoyun Zhang, Member, IEEE, Wujing Li, Member, IEEE, and Bing Tu, Member, IEEE

Abstract—Hyperspectral images (HSIs) have high spatial resolution and spectral resolution, and using HSI as a change detection (CD) data source is crucial for detecting surface changes. However, there is a large amount of real noise in HSIs, and most deep-learning-based CD methods require a large number of ground-truth labels for training, which is difficult and expensive to label manually. To reduce the dependence of CD on ground-truth labels and weaken the interference of noise on CD in HSIs, in this article, we propose an HSI change detection framework with a self-supervised contrastive learning pretrained model (CDSCL). CDSCL consists of two parts: self-supervised contrastive learning pretrained model and CD classification network. The main contributions of this article are as follows: a data augmentation strategy based on Gaussian noise is proposed to improve the ability of the model to extract variation information from HSIs with different random Gaussian noises; based on the information bottleneck theory, a progressive feature extraction module is developed to remove redundant or irrelevant details in changing information spectrum; and a contrastive loss function based on the Pearson correlation coefficient and negative cosine correlation is designed to make the features extracted by the two branches of the siamese network close to each other. Experimental results on four real hyperspectral datasets demonstrate that the CD performance of CDSCL outperforms the most representative CD methods.

Index Terms—Change detection (CD), contrastive learning, data augmentation, hyperspectral image (HSI), pretrained model.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) have high spatial and spectral resolution and contains more information than multispectral images and RGB images [1]. In recent years, with the development of imaging technology, HSI datasets have been continuously enriched, and more and more applications based on HSIs have emerged, such as change detection (CD) [2], [3], [4], image classification [5], [6], and target detection (TD) [7]. The main task of HSI CD is to use dual-temporal HSIs to determine whether the corresponding scene [8] or pixel [9] has changed, and it is widely used in green area analysis [10], ecosystem monitoring [11], and urban scene layout [12].

Early HSI CD methods are mostly derived from multispectral CD techniques [13], e.g., change vector analysis (CVA) [14] and principal component analysis (PCA) [15]. Based on this, sequential spectral CVA is proposed and used for HSI CD [16]. Many scholars have also improved PCA, such as HSI CD method using temporal principal components [17] and HSI CD method based on sparse PCA [18]. Furthermore, the authors in [19] applied multivariate alteration detection (MAD) to HSI CD. On this basis, to better extract the multivariate change information of dual-temporal HSIs, some scholars have tried iteratively reweighted multivariate alteration detection (IR-MAD) and combined it with feature reduction [20] and initial change mask [21], respectively. The covariance structure of dual-temporal HSIs can provide partial information of linear changes between them, so an HSI CD method based on covariance equalization is proposed [22]. In addition, some scholars have proposed a subspace-based HSI CD method, the principle of which is to measure the spectral changes [9]. Traditional CD methods have high requirements on the quality of the input images, while the application of advanced image denoising methods and techniques in the field of CD is seldom and immature, which greatly limit the performance of the aforementioned methods in CD.

In recent years, some CD algorithms have been improved, and deep-learning-based HSI CD methods have become more and more popular. Improved CD algorithms, e.g., a bidirectional continuous change detection (Bi-CCD) [23] proposed by improved continuous change detection (CCD). Some classic deep learning frameworks applied to change detection are convolutional neural networks (CNN) [24], long short term memory [25], recurrent neural networks (RNN) [26], etc. A deep siamese convolutional multiple-layers RNN [27] can extract spatial-spectral features from heterogeneous images and map them to new feature spaces to mine variation information. A general end-to-end two-dimensional CNN (GETNET) [28] makes...
full use of the subpixel-level endmember abundance information combined with the cross-channel gradient information of HSIs for CD. Deeply supervised image fusion network (IFN) [29] utilizes the attention module to fuse the multilevel deep features of the original image with the image difference features, improve the boundary integrity and internal compactness of the objects in the output change map, and reconstruct the change map. A spectral-spatial convolution neural network with siamese architecture [30] obtains two spectral space vectors by extracting tensor pairs in dual-temporal HSIs and merging them into a spectral space network, and then, calculating the similarity between the vectors to obtain a CD map. End-to-end siamese CNN (SiamNet) with a spectral-spatial-wise attention mechanism (SSA-SiamNet) [31] has a spectral-spatial attention mechanism, emphasizing information-rich spectral channels and locations, which can effectively improve the performance of CD. A CNN framework with slow-fast band selection (SFBS) and feature fusion grouping (SFBS-FFGNET) [32] designs SFBS to select slow and fast changing bands in HSIs, and also designs feature fusion grouping (FFG) module to change the direction of the feature vector. A multiscale fully convolutional neural network (MFCN) [33] uses multiscale convolution kernels to extract detailed features of ground features, and also designs a loss function to solve the problem of imbalanced positive and negative samples.

In the existing HSI CD techniques, we often ignore the most essential factor limiting the performance of CD, that is, the quality of the dual-temporal HSIs itself. In the process of collecting HSI, the acquisition equipment or the environment may bring a lot of real noise to the HSI, which will greatly affect the quality of the HSI. The complex real noise in HSI can be simulated by a noise map with zero-mean Gaussian distribution [34]. Furthermore, most deep-learning-based CD methods require a large number of ground-truth labels for training, and manual annotation of labels is difficult and expensive. How to effectively reduce the negative impact of noise on the performance of HSI CD and reduce the dependence of CD on ground-truth labels are major challenges.

The CDSCL proposed in this article does not require ground-truth labels in the pretrained phase and may extract effective change information in the dual-temporal HSIs, using only a limited number of pseudolabel samples for classification training. To reduce the interference of noise and improve the performance of HSI CD, the main contributions of this article are as follows.

1) A data augmentation strategy based on Gaussian noise is designed. Gaussian noise is used to simulate real noise [34]. After adding Gaussian noise, the amount of pretrained hyperspectral data is increased, the generalization ability of the model to Gaussian noise is improved, and the influence of noise on the extraction of changing features is alleviated. In addition, the data augmentation strategy based on the Gaussian noise can enhance the ability of the model to extract the changing information from different random Gaussian noise HSI while also increasing its robustness.

2) Based on the information bottleneck (IB) theory [35], a progressive feature extraction module (PFEM) is developed. The PFEM compresses information by designing the network structure into a bottleneck shape, removes redundant or irrelevant details, and retains only the most relevant and effective features for the detection task. A progressive bottleneck structure is designed in the PFEM to further ensure the effectiveness of the retained features.

3) A contrastive loss function is constructed based on the Pearson correlation coefficient and negative cosine similarity on the siamese contrastive representation learning structure. The loss function based on the Pearson correlation coefficient makes the features of the upper and lower branches of the siamese network close to each other, so as to obtain common change information. The loss function based on the negative cosine similarity makes the deeper features of the upper and lower branches of the siamese network intersect and close, and further unifies the change information of the two branches. When the change information extracted by the two branches of the siamese network has a high correlation, it indicates that the random Gaussian noise has little effect on the change information extraction.

The remaining four sections of this article are organized as follows. Section II summarizes related work on data augmentation and contrastive learning for HSI applications. Section III elaborates the details of the proposed CDSCL. Section IV designs experiments to evaluate the performance of CDSCL and compare it with other CD methods. Finally, Section V concludes this article.

II. RELATED WORK

Data augmentation and contrastive representation learning are key components in CDSCL and have unique advantages in dual-temporal HSI CD. In this section, we present a review of related work from both data augmentation and contrastive representation learning.

A. Data Augmentation

Data augmentation helps improve the generalization ability and robustness of deep models. Li et al. [36] proposed a data augmentation method for pixel block pairs, which greatly increased the number of training samples for HSIs. Yang et al. [37] proposed data augmented matched subspace detector (DAMSD) and data augmented MSDinter (DAMSDI) to solve the issue of target spectral scarcity. To address insufficiently labeled samples in practical spectroscopic measurements, Mu et al. [38] proposed a conditional variational autoencoder (CVAE). Wang et al. [39] greatly improved HSI classification accuracy by performing unsupervised data augmentation on hyperspectral samples. Zhang et al. [40] proposed a spectral-spatial fractal residual convolutional neural network with data-balanced augmentation, where the data-balanced augmentation method can address the limited labeled data and class imbalance problems. To improve the generalization ability of classification models,
Acción et al. [41] proposed a superpixel-based dual-window superpixel data augmentation framework to improve the overall classification accuracy of HSI datasets. Gao et al. [42] proposed a dynamic data selection algorithm for the problem of small hyperspectral samples and unbalanced class distribution. Data augmentation is gradually being applied in the field of HSIs, but people tend to ignore the noise in HSIs. It is urgent to explore new ways of data enhancement in the field of HSI CD to better improve the performance of CD.

B. Contrastive Representation Learning

Contrastive representation learning [43] does not require ground-truth label information, and directly uses its own data as supervision information to learn the feature expression of sample data and use it for downstream detection tasks. To cope with the difficulty of HSI clustering, Hu et al. [44] proposed a new deep subspace clustering method to extract spatial-spectral features through contrastive learning. Lee et al. [45] leverage the contrastive learning framework proposed by the cross-domain CNN to learn different HSI representations with different spectral features. Li et al. [46] combined supervised and unsupervised contrastive losses using a multichannel contrastive learning strategy with multiple data transformation methods to further improve the classification accuracy and generalization ability of the network. Cai et al. [47] proposed a spectral-spatial contrastive clustering (SSCC) model, which improves the robustness of the model based on contrastive learning. Cao et al. [48] combined contrastive learning and autoencoding to improve the feature learning ability of the network. Hu et al. [49] proposed an unsupervised HSI classification framework based on a contrastive learning method and a transformer model to efficiently extract HSI features without supervision. Zhao et al. [50] achieved HSI classification with few labeled samples by introducing contrastive self-supervised learning (SSL). To reduce the negative impact of insufficient label information on the hyperspectral detection task, Hou et al. [51] proposed an HSI classification algorithm based on contrastive learning. Research on HSI CD based on contrastive learning is very scarce, and how to use contrastive learning to extract effective change information remains to be explored.

III. METHODOLOGY

The overview of the proposed CDSCL is shown in Fig. 1. CDSCL is divided into two parts: pretrained model and CD classification. First, a distance spectrum containing change information is obtained by calculating the absolute distances between dual-temporal HSIs. Second, add different random Gaussian noise to the distance spectrum. Next, the effective variation information features between pixel pairs are extracted by the PFEM. Then, the designed contrast loss function is used to make the upper and lower branch features in the siamese contrastive learning structure close to each other. Until the network is fitted, the model pretrained ends. Finally, the PFEM with parameters in the pretrained model is taken out for the downstream CD task. The whole process of CDSCL does not require ground-truth labels, and the overall network structure details are shown in Table I.

A. Data Augmentation Strategy

HSIs contains a significant amount of complicated real noise, which will greatly affect the performance of CD. However, zero-mean Gaussian distributed noise is used to simulate real noise, which is not only simple in distribution but also easy to implement [34]. After adding random Gaussian noise, when the
TABLE I
ARCHITECTURE DETAILS FOR CDSCL

| Layers          | Type                        | Filters |
|-----------------|-----------------------------|---------|
| PFEM-FC1        | Fully Connected + BN        | 128     |
| PFEM-FC2        | Activation(relu)            |         |
| PFEM-FC3        | Fully Connected + BN        | 64      |
| PFEM-FC4        | Activation(relu)            |         |
| PFEM-FC5        | Fully Connected + BN        | 32      |
| PFEM-FC6        | Activation(relu)            |         |
| PFEM-FC7        | Fully Connected + BN        | 256     |
| PFEM-FC8        | Activation(relu)            |         |
| PFEM-FC9        | Fully Connected + BN        | 128     |
| Predictor MLP   | Activation(relu)            |         |
| MLP-FC1         | Fully Connected + BN        | 128     |
| MLP-FC2         | Activation(relu)            |         |
| CD Classification |                             |         |
| CD-FC1          | Fully Connected + BN        | 10      |
| CD-FC2          | Activation(softmax)         | 2       |

self-supervised pretrained model has been fitted, the features of the upper and lower branches are highly similar, indicating that change information features in the distance spectrum are successfully extracted. The data augmentation strategy based on Gaussian noise can not only increase the amount of pretrained HSI data, but also improve the generalization ability of the model to Gaussian noise, which is beneficial to HSI CD.

Assuming that the probability density expression of a continuous random variable $X$ is defined as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$

(1)

where $\mu$ and $\sigma (\sigma > 0)$ are constants, it is defined that $X$ obeys the Gaussian distribution of parameters $\mu$ and $\sigma$, denoted as $X \sim N(\mu, \sigma^2)$. In (1), $\mu$ and $\sigma^2$ represent the mean and variance of the Gaussian distribution, respectively.

The noise comes from the dual-temporal HSIIs, but since the simple subtraction between Gaussian distributions still satisfies the Gaussian distribution, the noise in the distance spectrum still satisfies the Gaussian distribution. Given that the mean and variance of the first phase Gaussian noise are $\mu_1$ and $\sigma_1^2$, respectively, and the mean and variance of the second phase Gaussian noise are $\mu_2$ and $\sigma_2^2$, respectively, the mean and variance in the distance spectrum can be expressed as

$$\mu_d = \mu_1 - \mu_2$$

(2)

$$\sigma_d^2 = \sigma_1^2 + \sigma_2^2$$

(3)

where $\mu_d$ and $\sigma_d^2$ represent the mean and variance of the simulated noise in the distance spectrum, respectively, and are substituted into (1) as follows:

$$D(x) = \frac{1}{\sqrt{2\pi}\sigma_d} \cdot e^{-\frac{(x-\mu_d)^2}{2\sigma_d^2}}$$

(4)

where $D(x)$ represents the probability density function of simulated noise in the distance spectrum, $X$ obeys the Gaussian distribution of parameters $\mu_d$ and $\sigma_d$, that is, $X \sim N(\mu_d, \sigma_d^2)$, and $X$ is the pixel in the distance spectrum. In (2), both $\mu_1$ and $\mu_2$ are 0, so the value of $\mu_d$ is 0.

This data augmentation strategy improves the generalization ability of the model to random Gaussian noise and enhances the robustness of the model by adding Gaussian noise that satisfies the distribution of (4) to the distance spectrum. Since the proposed CDSCL takes pixel-by-pixel spectral vectors as input, we choose to add noise pixel-by-pixel to ensure that the added noise form and the input form remain uniform. Assume that the number of pixels in the distance spectrum is $n$ and the number of bands is $l$. The distance spectrum is defined as $D = \{d_1, d_2, \ldots, d_i, \ldots d_n\}$, where the length of each pixel spectral vector $d_i$ is $l$. We obtain different sets of Gaussian noise vectors $G = \{g_{i1}, g_{i2}, \ldots, g_{in}\}$ and $H = \{h_{i1}, h_{i2}, \ldots, h_{in}\}$ by randomly sampling from a sample $Z$ with a zero-mean Gaussian distribution, where both vectors $g_i$ and $h_i$ are vectors of length $l$. $G$ and $H$ are obtained as follows:

$$g_{ij} = \text{random}(Z), j \in \{1, 2, \ldots, l\}$$

(5)

$$h_{ij} = \text{random}(Z), j \in \{1, 2, \ldots, l\}$$

(6)

where random(·) is the random sampling function, and $g_{ij}$ and $h_{ij}$ are the $j$th simulated noise value in the $i$th pixel spectral vector of $G$ and $H$, respectively. Each simulated noise vector in $G$ and $H$ satisfies the zero-mean Gaussian distribution of (4). To increase the complexity of the simulated noise vector, a series of random scale coefficients $C = [c_1, c_2, \ldots, c_l]$ are generated in the range of $0 \sim 0.1$. The noise figure set $K = [k_1, k_2, \ldots, k_l]$ is obtained by the following calculation:

$$k_i = c_i \times \text{mean}(D), i \in \{1, 2, \ldots, n\}$$

(7)

where mean(·) is the mean function and $k_i$ is the $i$th coefficient in $K$. Then, the two distance spectra $D^{(1)} = [d_1^{(1)}, d_2^{(1)}, \ldots, d_i^{(1)}, \ldots, d_n^{(1)}]$ and $D^{(2)} = [d_1^{(2)}, d_2^{(2)}, \ldots, d_i^{(2)}, \ldots, d_n^{(2)}]$ after adding noise are obtained by the following calculation:

$$d_{i1}^{(1)} = d_i + k_i \times g_i, i \in \{1, 2, \ldots, n\}$$

(8)

$$d_{i2}^{(2)} = d_i + k_i \times h_i, i \in \{1, 2, \ldots, n\}$$

(9)

where $d_{i1}^{(1)}$ and $d_{i2}^{(2)}$ are the spectral vector of the $i$th pixel in the distance spectrum after adding noise, respectively, and $D^{(1)}$ and $D^{(2)}$ are the image pair after data augmentation.
B. Progressive Feature Extraction Module (PFEM)

The image pairs obtained after data augmentation based on Gaussian noise contain a lot of redundant or irrelevant details. To alleviate the interference of irrelevant information on feature extraction, remove redundant information, and retain the effective change information in the distance spectrum to a large extent, based on the IB theory [35], we propose a PFEM in this article. In the PFEM, the network structure is designed into a bottleneck shape to compress the change information, reduce the amount of information passing through the bottle mouth, and suppress the passing of information irrelevant to the training task. Simultaneously, a progressive bottleneck structure is designed in the PFEM, and bottlenecks of different sizes are designed at different network depths to further extract effective change information. In the self-supervised contrastive learning pretrained model, the parameters of the PFEM of the upper and lower branches of the siamese network are shared.

The overview of the PFEM is shown in Fig. 2. First, connect the extracted spectral vectors to the input layer of the PFEM. To avoid the loss of effective information, we design a bottleneck with a slightly larger size at the front end of the PFEM to allow more information to pass through, although the information passing through the bottle mouth may still contain some redundancy. The information passed by the first bottleneck is then generalized by designing hidden layers. Next, a bottleneck layer with a smaller bottle mouth is established to further remove features irrelevant to the training task, and retain more effective features of change information. Finally, the informative features compressed by the second bottleneck layer are generalized through the hidden layer and connected to the output layer. The output of the $n$th layer in the PFEM can be expressed as

$$s_n(x) = s(\theta, x) = \text{relu}(w_ns_{n-1}(x) + b_n)$$

where $x$ is the information of the input layer, $s_n(x)$ is the output feature of the $n$th layer of the PFEM, $\theta$ represents all the parameter information of the whole module, and $w_n$ and $b_n$ represent the weight coefficient and bias vector of the $n$th layer, respectively. The expression for the final output of the PFEM in the self-supervised pretrained model are defined as

$$z_1 = \text{PFEM}(d^{(1)})$$

$$z_2 = \text{PFEM}(d^{(2)})$$

where $z_1$ and $z_2$ represent the output of the upper and lower branch PFEM modules, and $d^{(1)}$ and $d^{(2)}$ represent the spectral vector of the corresponding coordinate pixel pair in the image pair after data augmentation.

C. Design of Contrastive Loss Function

Self-supervised pretrained models are trained without ground-truth labels and learn feature representations using their own data as supervised information. After adding random Gaussian noise to the distance spectrum in Fig. 1, the pixel spectrum vectors $d^{(1)}$ and $d^{(2)}$ of the corresponding coordinates in $D^{(1)}$ and $D^{(2)}$ are less correlated. When the feature correlation extracted by the upper and lower branches of the siamese network is high, it indicates that the model extracts the common features in the image pair after data augmentation, that is, the effective change information in the distance spectrum. Furthermore, it also shows that the model effectively reduces the influence of Gaussian noise on feature extraction of change information. Therefore, we design a contrastive loss function based on the Pearson correlation coefficient and negative cosine similarity to improve the consistency of the features of the upper and lower branches of the siamese network. The purpose of the loss function based on the Pearson correlation coefficient is to improve the consistency of $z_1$ and $z_2$. The purpose of the loss function based on the negative cosine correlation is to improve the cross feature correlation between $z_1$, $z_2$, $p_1$, and $p_2$.

To obtain the cross change informative feature vectors $p_1$ and $p_2$, we design a shallow MLP predictor. In the pretrained model, the MLP predictor parameters on the upper and lower branches of the siamese network are not shared. To improve the correlation between $z_1$ and $z_2$, we design a loss function based on the Pearson correlation coefficient. The Pearson correlation coefficient is defined as

$$\text{PCCs}(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

where $\text{PCCs}(\cdot)$ is the Pearson correlation coefficient calculation function, $X$ and $Y$ are the input vectors, and $n$ is the length of the input vector. Considering that the loss value is decreasing during the network training process, the loss function is designed as

$$L_p = 1 - |\text{PCCs}(z_1, z_2)|$$
where $L_p$ is the loss function based on the Pearson correlation coefficient, and $z_1$ and $z_2$ are the input vectors. To improve the correlation between feature vectors of cross change information, we design a loss function based on the negative cosine correlation. The negative cosine correlation of the cross change informative feature vectors is calculated as follows:

$$
n_{cs}(z_1, p_2) = -\frac{z_1^T p_2}{\|z_1\|\|p_2\|}$$  \hspace{1cm} (15)

$$
n_{cs}(z_2, p_1) = -\frac{z_2^T p_1}{\|z_2\|\|p_1\|}$$  \hspace{1cm} (16)

where $n_{cs}()$ is the negative cosine correlation function, (15) is the negative cosine correlation between the cross vectors $z_1$ and $p_2$, and (16) is the negative cosine correlation between the cross vectors $z_2$ and $p_1$, $\|\cdot\|$ is $l_2$-norm. Then, the loss function is designed as

$$
L_n = \frac{1}{2}n_{cs}(z_1, p_2) + \frac{1}{2}n_{cs}(z_2, p_1)$$  \hspace{1cm} (17)

where $L_n$ is the loss function based on the negative cosine correlation. Then, the global contrastive loss function of the self-supervised pretrained model is expressed as

$$
L_c = \lambda * L_p + (1 - \lambda) * L_n$$  \hspace{1cm} (18)

where $L_c$ is the overall contrastive loss function designed, and $\lambda$ is the weight coefficient of $L_p$. When $z_1$ and $z_2$ are highly correlated, $L_p$ is close to 0. When $z_1$ and $p_2$ are highly correlated, and $z_2$ and $p_1$ are highly correlated, $L_n$ is close to $-1$.

D. HSI CD

The distance spectrum calculated by dual-temporal HSIs contains a lot of change information, and the pretrained model can eliminate the interference of noise on the feature extraction of change information, which provides an excellent foundation for the downstream HSI CD task.

The CD classification structure design is shown in Fig. 1. In the CD classification structure design, we first extract the pixel spectral vector from the distance spectrum, then connect the trained PFEM with parameters, and finally, use the softmax as the classifier to obtain the binary change detection map. In the downstream HSI CD task, we adopt a weakly supervised training method, which can be used to train the classification network by selecting a small number of pseudolabel samples from CVA predetection without requiring ground-truth labels. After self-supervised contrastive representation learning on the complete dataset, excellent results can still be achieved even with only a small number of high-confidence pseudolabels for retraining.

Cross entropy is one of the most commonly used loss functions, which can compute the difference between the trained probability distribution and the true distribution in classification tasks. The smaller the loss value, the closer the predicted result is to the expected result. The cross-entropy loss function and the softmax classifier can be combined to train the binary classification task. The cross entropy is usually defined as follows:

$$L_{ce} = -y \ln x + (1 - y) \ln (1 - x)$$  \hspace{1cm} (19)

where $y$ is the label of the pixel, the value is 0 or 1, and $x$ is the output probability of the current network. In binary classification tasks, cross-entropy and softmax classifiers often work together. The softmax classifier is defined as follows:

$$\text{softmax}(z) = \frac{e^{z_i}}{\sum_{i=1}^{K} e^{z_i}}$$  \hspace{1cm} (20)

where $\text{softmax}(z)$ is abbreviated as $s(\cdot)$. According to (19) and (20), the loss function of the binary classification task is obtained as follows:

$$L = -[y \ln(s(x)) + (1 - y) \ln(s(1 - x))]$$  \hspace{1cm} (21)

IV. EXPERIMENTS

To verify the reliability and stability of the proposed CDSCL, four real dual-temporal HSIs datasets were used for test in this section. In the experiment, we adopt several classical CD methods and three novel and representative CD methods as comparison methods to objectively demonstrate the CD performance of CDSCL. Simultaneously, we design an ablation experiment to confirm that self-supervised pretrained models play a key role in improving the HSI CD performance. Furthermore, experiments are designed in this section to reveal the CD performance of CDSCL with different proportions of the number of training samples, and are tested on four dual-temporal HSIs datasets. Finally, the experimental performance of CDSCL and all comparison methods is summarized.

A. Experiment Setting

This experiment runs on the Tensorflow platform, the GPU used is a single NVIDIA 2080Ti graphics card, and the device is required to have more than 16 G of memory. To better reflect the superiority of CDSCL, we use two classical methods based on algebraic operations and three representative new methods based on deep learning to compare with CDSCL. Two algebraic analysis methods are CVA [14] and principal component analysis change vector analysis (PCA-CVA) [52]. Three deep-learning-based methods are GETNET [28], SSA-SiamNet [31], and SFBS-FFGNET [32], where SSA-SiamNet is a supervised method and its code is reproduced from this article. The self-supervised pretrained model is the core of CDSCL, so we design CDSCL without a pretrained model as an ablation experiment. The criticality of self-supervised contrastive learning of pretrained models can be better highlighted by comparing CDSCL with CDSCL without pretrained.

The contrastive learning pretrained model in CDSCL is based on self-supervised training, while the downstream CD task is based on weakly supervised training, which does not require ground-truth labels throughout the process, avoiding manual labeling. In weakly supervised CD methods, the selection of high-confidence pseudolabels is crucial for detection performance. First, SFBS [32] is used to reduce the dimensionality of the HSI dataset, retaining the bands that are conducive to change detection. Second, binary pseudolabels are generated using CVA algebraic analysis and k-means clustering. Third, the Euclidean distance between pixel pairs of the dual-temporal HSIs are
TABLE II
CONFUSION MATRIX OF EVALUATION INDICATORS

| Truth | Predicted |
|-------|-----------|
|       | Unchanged (0) | Changed (1) |
| Unchanged (0) | TN | FN |
| Changed (1)    | FP | TP |

calculated. Fourth, select a small quantity of pixels with the largest Euclidean distance value from the changed pixels in the predetection results as high-confidence pseudochanged training samples, and select a small quantity of pixels with the smallest Euclidean distance value from the unchanged pixels in the pre-detection results as high-confidence pseudounchanged training samples. High-confidence pseudochanged training samples and pseudounchanged training samples are integrated into a training set for the CD classification network training of CDSCL, where the ratio of changed samples and unchanged samples follows the principle of 1:2 [28]. We conduct five repeated experiments for each deep learning-based method, and use the obtained mean and variance as the experimental results to intuitively reflect the performance and robustness of the method.

To comprehensively evaluate the CD performance of CDSCL and all comparison methods, a comprehensive index evaluation system was established in this experiment, including overall detection accuracy (OA), OA of changed area (OA_CHG), OA of unchanged area (OA_UN), and Kappa coefficient for consistency assessment [32]. All evaluation metrics used in the experiments are calculated as follows:

\[
\text{OA}_{\text{CHG}} = \frac{TP}{TP + FN} \tag{22}
\]

\[
\text{OA}_{\text{UN}} = \frac{TN}{TN + FP} \tag{23}
\]

\[
OA = \frac{TN + FP}{TN + FN + FP + TP} \tag{24}
\]

\[
p_e = \frac{(TN + FN)(TN + FP) + (FP + TP)(FN + TP)}{(TN + FN + FP + TP)^2} \tag{25}
\]

\[
\text{Kappa} = \frac{OA - p_e}{1 - p_e} \tag{26}
\]

Table II is the confusion matrix needed to calculate each evaluation index, where TN represents true negatives, FN represents false negatives, FP represents false positives, and TP represents true positives.

**B. Experiment on River Dataset**

The dual-temporal HSIs used for CD in the “River” dataset were collected on May 3 and December 31, 2013 from a river area in Jiangsu Province, China [28]. The sensor used for shooting is Earth Observing-1 (EO-1) Hyperion, with a spectral range of 0.4–2.5 μm, a spectral resolution of 10 nm, and a spatial resolution of 30 m. The size of the image in the dataset is 463 × 241 pixels, the source data has 242 spectral bands, and after removing some noise bands, 198 spectral bands are obtained for detection tasks. The main change in the River dataset is the difference in river channel coverage. In Fig. 3(a) and (b) are pseudocolor maps of images collected at different times, respectively, and Fig. 3(c) is the ground truth. The white areas in the ground-truth map represent changed areas, and the black areas represent unchanged areas. In experiment on the River dataset, we choose 1116 changed pseudolabel samples and 2232 unchanged pseudolabel samples as the training set, accounting for 3% of the total number of pixels.

The CD results of all experimental methods on the River dataset are shown in Table III. PCA-CVA has the highest OA_CHG, indicating that it performs well in detecting changed pixels, but its OA_UN is only higher than CVA, and its overall performance ranks fifth among all methods because of the large proportion of unchanged pixels. GETNET has the highest OA_UN, which means it is excellent at detecting unchanged pixels, but it has the worst OA_CHG, so its OA and Kappa are only higher than CVA. The OA_CHG of SSA-SiamNet is higher than that of SFBS-FFGNET, the OA_UN of SFBS-FFGNET is slightly higher than that of SSA-SiamNet, and the proportion of unchanged pixels is 91.31%, so the OA of the two methods is very close, but the Kappa of SFBS-FFGNET is slightly higher. The OA_CHG of CDSCL is significantly higher than other deep-learning-based methods, and OA_UN also reaches 98.40%, so its OA and Kappa are the first of all experimental methods. The standard deviations of the four evaluation index of CDSCL are all at very low levels, which shows that its robustness is excellent.

In ablation experiment, OA_CHG and OA_UN of CDSCL without pretrained are 77.68% and 98.14%, respectively. Compared with CDSCL without pretrained, OA_CHG is 7.43% higher and OA_UN is 0.26% higher for CDSCL, which shows that CDSCL is better than CDSCL without pretrained in detecting both changed pixels and unchanged pixels. The OA and kappa of CDSCL are higher than that of CDSCL without pretrained, which indicates that the self-supervised contrastive learning pretrained model proposed in this article can effectively improve the performance of downstream CD tasks. In addition,
CD RESULTS OF DIFFERENT METHODS ON THE RIVER DATASET

| Different methods | Evaluation index |
|-------------------|------------------|
|                  | OA_CHG | OA_UN | OA    | Kappa   |
| CVA              | 0.8612 | 0.8921 | 0.8894 | 0.5195  |
| PCA-CVA          | **0.9503** | 0.9518 | 0.9517 | 0.7477  |
| GETNET           | 0.4866 ± 0.0550 | **0.9958 ± 0.0015** | 0.9516 ± 0.0041 | 0.6111 ± 0.0470 |
| SSA-SiamNet      | 0.7898 ± 0.0557 | 0.9844 ± 0.0034 | 0.9674 ± 0.0050 | 0.7905 ± 0.0383 |
| SFBS-FFGNET      | 0.7850 ± 0.0459 | 0.9854 ± 0.0040 | 0.9680 ± 0.0017 | 0.7923 ± 0.0148 |
| CDSCL(without pre-trained) | 0.7768 ± 0.0217 | 0.9814 ± 0.0020 | 0.9636 ± 0.0029 | 0.7677 ± 0.0190 |
| CDSCL            | 0.8511 ± 0.0120 | 0.9840 ± 0.0011 | **0.9724 ± 0.0005** | **0.8278 ± 0.0038** |

The bold entities indicate the best result of the comparison methods.

the standard deviation of CDSCL is lower than that of CDSCL without pretrained, which means that the self-supervised contrastive learning pretrained model can effectively extract change features and improve the robustness of CD tasks. Although the change detection performance of CDSCL without pretrained lags behind CDSCL in all aspects, its OA performance is good, in large part because the PFEM structure suppresses redundant and irrelevant details.

Fig. 4 shows CD binary maps for all experimental methods. As can be seen from Fig. 4, CVA has the worst intuitive visual effect, where large quantity of changed pixels are detected as unchanged pixels, and large quantity of unchanged pixels are detected as changed pixels. In the square box in the upper left corner, PCA-CVA detected a small number of unchanged pixels as changed pixels, while a small number of changed pixels in GETNET, and SFBS-FFGNET are detected as unchanged pixels, and the consistency between CDSCL and ground truth is the highest. In the square box in the middle, PCA-CVA is closest to the ground-truth map, and SSA-SiamNet, SFBS-FFGNET, CDSCL without pretrained, and CDSCL all have a small amount of changed pixels that are not detected. In addition, scattered areas also affect the final CD map, and CDSCL is the closest to the ground-truth map in the overall performance.

In the ablation experiment, the difference in CDSCL without pretrained and CDSCL is evident from Fig. 4. In the upper left-square box, the detection performance of CDSCL without pretrained and CDSCL is close and basically the same as the ground truth. In the square box in the middle and the rectangle in the lower right corner, CDSCL without pretrained has more changed pixels undetected than CDSCL. From the overall visual effect, CDSCL is closer to the ground-truth map, which intuitively reflects the superiority of the proposed self-supervised contrastive learning pretrained model.

C. Experiment on Farmland Dataset

The dual-temporal HSIs used for CD in the “Farmland” dataset were collected on May 3, 2006 and April 23, 2007 from farmland near Yancheng City, Jiangsu Province, China [9]. The sensor used is Earth Observing-1 (EO-1) Hyperion, with a spectral range of 0.4–2.5 μm, a spectral resolution of 10 nm, and a spatial resolution of 30 m. The size of the image in the dataset is 450 × 140 pixels, the source data have 242 bands, and after processing through the preprocessing techniques of [53] and [54], 155 bands are obtained for detection tasks. The source
TABLE IV
CD RESULTS OF DIFFERENT METHODS ON THE FARMLAND DATASET

| Different methods         | Evaluation index       |
|---------------------------|------------------------|
|                           | OA_CHG | OA_UN | OA   | Kappa   |
| CVA                       | 0.9808 | 0.9466 | 0.9566 | 0.8979 |
| PCA-CVA                   | 0.9587 | 0.9701 | 0.9668 | 0.9202 |
| GETNET                    | 0.9817 ± 0.0097 | 0.9608 ± 0.0082 | 0.9668 ± 0.0063 | 0.9213 ± 0.0148 |
| SSA-SiamNet               | 0.9514 ± 0.0090 | 0.9825 ± 0.0038 | 0.9735 ± 0.0015 | 0.9355 ± 0.0037 |
| SFBS-FFGNET               | 0.9780 ± 0.0010 | 0.9755 ± 0.0020 | 0.9762 ± 0.0012 | 0.9429 ± 0.0029 |
| CDSCL(without pre-trained)| 0.9614 ± 0.0126 | 0.9730 ± 0.0029 | 0.9697 ± 0.0016 | 0.9269 ± 0.0033 |
| CDSCL                     | 0.9788 ± 0.0028 | 0.9852 ± 0.0007 | 0.9834 ± 0.0003 | 0.9598 ± 0.0009 |

The bold entities indicate the best result of the comparison methods.

Fig. 5. Three components in the Farmland dataset. (a) Imagery acquired on May 3, 2006. (b) Imagery acquired on April 23, 2007. (c) Ground truth.

data have 242 spectral bands, and after simple preprocessing, 155 spectral bands are obtained for detection tasks. The main change in the image is the area covered by farmland. In Fig. 5(a) and (b) are pseudocolor maps of images collected at different times, respectively, and Fig. 5(c) is the ground truth. The white areas in the ground-truth map represent changed areas, and the black areas represent unchanged areas. In experiment on the Farmland dataset, we choose 840 changed pseudolabel samples and 1680 unchanged pseudolabel samples as the training set, accounting for 4% of the total number of pixels.

The CD results of all experimental methods on the Farmland dataset are shown in Table IV. GETNET has the highest OA_CHG, indicating that it can detect the changing pixels of Farmland well, but its OA is only 96.08%, and its overall performance is only the fifth best among all methods because of the large proportion of unchanged pixels. Among the deep-learning-based comparison methods, SSA-SiamNet has the highest OA_UN and GETNET has the highest OA_CHG, but since both OA_CHG and OA_UN of SFBS-FFGNET exceed 97%, SFBS-FFGNET is the best performer among the three methods. The OA_UN, OA, and Kappa of CDSCL are the best among all the experimental methods, indicating that it can not only detect the unchanged pixels in the Farmland dataset excellently, but also outperform all the comparison methods in overall CD performance. Furthermore, the standard deviations of the four indicators of CDSCL are all at a very low level, which shows that the robustness of CDSCL is excellent.

In ablation experiment, the OA_CHG and OA_UN of CDSCL without pretrained are 96.14% and 97.30%, respectively. Compared with CDSCL without pretrained, CDSCL has 1.74% higher OA_CHG and 1.22% higher OA_UN, indicating that CDSCL is stronger than CDSCL without pretrained in detecting both changed and unchanged pixels in the Farmland dataset. Both the OA and Kappa of CDSCL are larger than those of CDSCL without pretrained, which reflects the superiority of the self-supervised contrastive learning pretrained model, and also shows that the proposed pretrained model can extract the change information features excellently. In addition, the standard deviations of the four evaluation indicators of CDSCL are all lower than those of CDSCL without pretrained, which indicates that the self-supervised contrastive learning pretrained model can not only improve the performance of the downstream HSI CD task, but also improve the robustness of the HSI CD task. Although the change detection performance of CDSCL without pretrained lags behind CDSCL in all aspects, its OA reaches 96.97%, which is largely due to the suppression of redundant and irrelevant details by the PFEM structure.

The CD binary maps of all experimental methods on the Farmland dataset is shown in Fig. 6. In the rectangular area in the upper right corner, CVA, PCA-CVA, GETNET, SSA-SiamNet, and CDSCL without pretrained have many unchanged pixels detected as changed pixels, while SFBS-FFGNET and CDSCL have only a few false alarm pixels. In the square area in the middle, CVA and PCA-CVA have a small number of false alarm pixels, SSA-SiamNet, SFBS-FFGNET, CDSCL without pretrained, and CDSCL have few changed pixels that are not detected, while GETNET misses almost all the changed pixels. In the elliptical area in the lower left corner, CVA, PCA-CVA, GETNET, and CDSCL without pretrained have more unchanged pixels detected as changed pixels, and other methods have only very few false alarm pixels in this area. In the rectangular area in the lower right corner, CVA,
PCA-CVA, GETNET, SSA-SiamNet, SFBS-FFGNET, and CD-SCL without pretrained all have more false alarm pixels, while CD-SCL is almost consistent with the ground-truth map. In terms of intuitive visual effects, SSA-SiamNet and SFBS-FFGNET are at the same level, while CD-SCL is closer to the ground-truth map.

CD binary maps for ablation experiments are shown in Fig. 6(f) and (g). In the rectangular area in the upper right corner, both CD-SCL without pretrained and CD-SCL have a small number of false alarm pixels, but the number of CD-SCL without pretrained is more than that of CD-SCL. In the square area in the middle, the detection results of CD-SCL without pretrained and CD-SCL are relatively consistent, but still not perfect compared to the ground-truth map. In the oval area in the lower left corner and the rectangular area in the lower right corner, CD-SCL without pretrained has many unchanged pixels detected as changed pixels, while the detection results of CD-SCL without pretrained and CD-SCL are relatively consistent, but still not perfect compared to the ground-truth map. In conclusion, CD-SCL performed better than CD-SCL without pretrained, which demonstrates the effectiveness of the proposed self-supervised contrastive learning pretrained model.

D. Experiment on Hermiston Dataset

The dual-temporal HSIs used for CD in the “Hermiston” dataset were collected in 2004 and 2007 in the urban area of Hermiston, OR, USA [55]. The sensor used for shooting is Earth Observing-1 (EO-1) Hyperion, with a spectral range of 0.4–2.5 μm, a spectral resolution of 10 nm, and a spatial resolution of 30 m. The size of the image in the dataset is 390×200 pixels, and the source data have 242 spectral bands, all of which are used for detection tasks. The main type of change in the image is the area of the city. In Fig. 7, (a) and (b) are pseudocolor maps of images collected at different times, respectively, and (c) is the ground truth. The white areas in the ground-truth map represent changed areas, and the black areas represent unchanged areas. In experiment on the Hermiston dataset, we choose 1560 changed pseudolabel samples and 3120 unchanged pseudolabel samples as the training set, accounting for 6% of the total number of pixels.

The CD results of all experimental methods on the Hermiston dataset are shown in Table V. CD-SCL without pretrained has the highest OA_CHG, but it has the lowest OA_UN, so its OA and Kappa perform poorly. The OA_UN of CVA is the highest, but its OA_CHG is the lowest, so its OA is the worst. Among the three deep learning-based comparison methods, GETNET has the worst overall performance, as its OA_CHG is only 71.59%, resulting in Kappa only 80.41%. Among all the comparison methods, SFBS-FFGNET and CD-SCL performed the best and are close, and CD-SCL has a slight lead compared to SFBS-FFGNET. The standard deviations of three of the four evaluation metrics of CD-SCL are the lowest among all comparison methods, which shows that CD-SCL has excellent robustness.

In ablation experiment, OA_CHG and OA_UN of CD-SCL without pretrained are 93.85% and 96.45%, respectively. Compared with CD-SCL without pretrained, OA_CHG is 1.83% lower and OA_UN is 3.09% higher for CD-SCL. In the Hermiston dataset, the proportion of unchanged pixels reaches 87.20%, so CD-SCL is far ahead of CD-SCL without pretrained in both OA and Kappa. A comprehensive comparison of CD-SCL and CD-SCL without pretrained can clearly demonstrate the advantages of the proposed self-supervised contrastive learning pretrained model. Furthermore, the standard deviations of the
Table V
CD RESULTS OF DIFFERENT METHODS ON THE HERMISTON DATASET

| Different methods      | Evaluation index               |
|------------------------|--------------------------------|
|                        | OA_CHG    | OA_UN    | OA      | Kappa    |
| CVA                    | 0.5316    | 0.9990   | 0.9392  | 0.6607   |
| PCA-CVA                | 0.9319    | 0.9856   | 0.9787  | 0.9060   |
| GETNET                 | 0.7159 ± 0.0769 | 0.9972 ± 0.0010 | 0.9612 ± 0.0095 | 0.8041 ± 0.0538 |
| SSA-SiamNet            | 0.8618 ± 0.0926 | 0.9957 ± 0.0019 | 0.9785 ± 0.0104 | 0.8970 ± 0.0567 |
| SFBS-FFGNET            | 0.8962 ± 0.0033 | 0.9964 ± 0.0013 | 0.9836 ± 0.0009 | 0.9239 ± 0.0018 |
| CDSCL (without pretrained) | **0.9385 ± 0.0116** | 0.9645 ± 0.0240 | 0.9611 ± 0.0195 | 0.8384 ± 0.0550 |
| CDSCL                  | 0.9202 ± 0.0069 | 0.9954 ± 0.0011 | **0.9858 ± 0.0003** | **0.9349 ± 0.0006** |

The bold entities indicate the best result of the comparison methods.

The four evaluation index of CDSCL are significantly lower than those of CDSCL without pretrained, which intuitively proves that the proposed pretrained model can improve the robustness of the HSI CD task.

CD binary maps for all experimental methods are shown in Fig. 8. In the triangle area in the upper left corner, CVA has a large quantity of changed pixels that are not detected, and GETNET and SSA-SiamNet also miss a small quantity of changed pixels, while PCA-CVA, SFBS-FFGNET, and CDSCL perform well in this area. Besides, the CDSCL without pretrained has large quantity of false alarm pixels in the upper left triangle area. In the oval area in the center, CVA and GETNET have detected large quantity of changed pixels as unchanged pixels, and PCA-CVA and CDSCL without pretrained have many false alarm pixels. In the rectangular area in the lower-right corner, CVA and GETNET have a small number of changed pixels that are not detected, CDSCL without pretrained has large quantity of unchanged pixels detected as changed pixels, and only CDSCL has a small number of false alarm pixels, other methods performed well. From the overall visual effect, the CD maps of CDSCL are closer to the ground-truth map, which reflects the superiority of the self-supervised contrastive learning pretrained model proposed in this article.

Fig. 8. Binary CD maps of all experimental methods on the Hermiston dataset. (a) CVA. (b) PCA-CVA. (c) GETNET. (d) SSA-SiamNet. (e) SFBS-FFGNET. (f) CDSCL (without pretrained). (g) CDSCL. (h) Ground-truth map.

E. Experiment on Bay Area Dataset

The dual-temporal HSIs used for CD in the “Bay Area” dataset were collected in 2013 and 2015 in Paterson, CA, USA [56]. The sensor used for shooting is AVIRIS, with a spectral range of 0.4–2.5 μm, a spectral resolution of 10 nm, and a spatial resolution of 20 m. The size of the image in the dataset is 600×500 pixels, and the source data have 224 spectral bands, all of which are used for detection tasks. In the Bay Area dataset, 34 211 unchanged pixels and 39 270 changed pixels are labeled, and the rest of the pixels belong to unlabeled unknown pixels.

In Fig. 9, (a) and (b) are pseudocolor maps of images collected at different times, respectively, and (c) is the ground truth. The white areas in the ground-truth map represent changed areas,
TABLE VI
CD RESULTS OF DIFFERENT METHODS ON THE BAY AREA DATASET

| Different methods   | Evaluation index      |
|---------------------|-----------------------|
|                     | OA_CHG  | OA_UN  | OA   | Kappa   |
| CVA                 | 0.8155  | 0.9189 | 0.8637 | 0.7283  |
| PCA-CVA             | 0.8503  | 0.8873 | 0.8675 | 0.7348  |
| GETNET              | 0.9692 ± 0.0246       | 0.9419 ± 0.0476 | 0.9565 ± 0.0142 | 0.9124 ± 0.0289 |
| SSA-SiamNet         | 0.9752 ± 0.0083       | 0.9927 ± 0.0036 | 0.9833 ± 0.0047 | 0.9666 ± 0.0093 |
| SFBS-FFGNET         | 0.9916 ± 0.0027       | 0.9670 ± 0.0035 | 0.9801 ± 0.0012 | 0.9600 ± 0.0023 |
| CDSCL(without pre-trained) | 0.9837 ± 0.0011       | 0.9588 ± 0.0024 | 0.9721 ± 0.0017 | 0.9439 ± 0.0033 |
| CDSCL               | 0.9873 ± 0.0011       | 0.9915 ± 0.0007 | 0.9892 ± 0.0005 | 0.9783 ± 0.0010 |

The bold entities indicate the best result of the comparison methods.

The black areas represent unchanged areas, and the gray areas represent unlabeled unknown areas. In experiment on the Bay Area dataset, we choose 1225 changed pseudolabel samples and 2 250 unchanged pseudolabel samples as the training set, accounting for 5% of the total number of pixels.

The CD results of all experimental methods on the Bay Area dataset are shown in Table VI. SFBS-FFGNET has the highest OA_CHG, indicating that it performs best in detecting changed pixels in the Bay Area dataset, but its OA_UN is only 96.70%, so it ranks third in the overall detection performance. SSA-SiamNet has the highest OA_UN, and its OA_CHG also reaches 97.52%, indicating that it performs well in detecting changed and unchanged pixels in the Bay Area dataset. From Table VI, it can be seen that the two methods based on algebraic analysis, CVA and PCA-CVA, perform poorly, which is caused by the complexity of the Bay Area dataset. Among all CD experimental methods, the OA and Kappa of CDSCL are the highest, and the robustness of CDSCL is excellent from the standard deviation.

In ablation experiment, the OA_CHG and OA_UN of CDSCL without pretrained are 98.37% and 95.88%, respectively. Compared with CDSCL without pretrained, CDSCL has 0.36% higher OA_CHG and 3.27% higher OA_UN, indicating that CDSCL is stronger than CDSCL without pretrained in detecting changed and unchanged pixels in the Bay Area dataset. Both the OA and Kappa of CDSCL are larger than those of CDSCL without pretrained, which reflects that after the self-supervised contrastive learning pretrained model is trained, the obtained PFEM with parameters can help the downstream CD task to extract change information features. In addition, the standard deviations of the four evaluation metrics of CDSCL are all at extremely low levels, which can also indicate that self-supervised contrastive learning pretrained models can bring stability to the CD task. Although the change detection performance of CDSCL without pretrained is inferior to CDSCL in all aspects, its OA reaches 97.21%, which is largely due to the suppression of redundant and irrelevant details by the PFEM structure.

The CD binary maps of all experimental methods on the Bay Area dataset is shown in Fig. 10. In the upper right rectangular area, CVA, PCA-CVA, GETNET, and SSA-SiamNet all have some changed pixels detected as unchanged pixels, while the rest of the methods are detected well. In the circular area in the upper left corner, CVA, PCA-CVA, and GETNET have large quantity of unchanged pixels detected as changed pixels, while the rest of the other methods have a small quantity of false alarm pixels in this area. In the middle rectangular area, CVA, PCA-CVA, and GETNET have some changed pixels detected as unchanged pixels, and the remaining other methods perform well but not perfect in this area. In the oval area in the lower left corner, CVA, PCA-CVA, and GETNET have some changed pixels detected as unchanged pixels, and the remaining other methods perform well but not perfect in this area. In the oval area of SFBS-FFGNET and CDSCL without pretrained all have a large quantity of unchanged pixels detected as changed pixels, and SSA-SiamNet and CDSCL also have a small quantity of false alarm pixels in this area. Most of the unchanged pixels in the lower left oval area of SFBS-FFGNET and CDSCL without pretrained are detected as changed pixels, indicating that their detection accuracy of unchanged pixels is not high, which corresponds to the data in Table VI. In terms of intuitive effects, the detection effects of SSA-SiamNet and CDSCL are both excellent. In detail, CDSCL is more consistent with the ground-truth map.

CD binary maps for ablation experiment are shown in Fig. 10(f) and (g). In the rectangular area in the upper right corner, the circular area in the upper left corner, and the rectangular area in the middle, the detection effects of CDSCL without pretrained and CDSCL are excellent. In the oval area in the lower left corner, CDSCL without pretrained almost detected more
than half of unchanged pixels as changed pixels, and CDSCL has only a small number of false-alarm pixels in this area. Although the overall detection performance of CDSCL without pretrained is good, it is largely inferior to CDSCL in detecting unchanged pixels of the Bay Area dataset. Overall, CDSCL outperforms CDSCL without pretrained in all aspects of detection, which demonstrates the effectiveness of the proposed self-supervised contrastive learning pretrained model.

F. Hyperparameter Analysis

In the CDSCL proposed in this article, many parameters are involved, such as the weight distribution between $L_p$ and $L_n$ in (18), batch size, regularization parameter $r$, and the size of the bottleneck layer in the PFEM. In order to determine the best combination of parameters for CDSCL, we designed experiments to study the effects of these parameters on change detection by using the control variable method.

For the global contrastive loss function in (18), we set the sum of the weight coefficients of $L_p$ and $L_n$ to 1. When the weight of $L_p$ is $\lambda$, the weight of $L_n$ is $1 - \lambda$. Under the framework of CDSCL, keeping other parameters unchanged, we gradually increase $\lambda$ from 0 to 1, increasing by 0.1 each time, and the obtained OA curves are shown in Fig. 11(a). When $\lambda$ is 0 or 1, that is, when a single loss function works, the value of OA is the lowest. Although the OA curves of River, Farmland, and Hermiston have small fluctuations, their OA at $\lambda = 0.5$ is still the highest. From Fig. 11(a), it can be concluded that when the weight of $L_p$ is 0.5, CDSCL has the best detection performance on the four datasets, and the weight of $L_n$ is also 0.5 at this time, indicating that $L_p$ and $L_n$ are equally important.

For the hyperparameter batch size, we defined five sizes {32, 64, 128, 256, 512}. Under the framework of CDSCL, other parameters are fixed and the batch size is changed each time, and the OA curves obtained are shown in Fig. 11(b). The curves of the river and the bay area rise rapidly between batch sizes of 32 and 64, and then, decline slowly after reaching a peak. The curves of Farmland and Hermiston are flat at batch size 32 and 64, and then, start to decline rapidly. From Fig. 11(b), it can be concluded that when the batch size is 64, the OA curves of CDSCL on the four datasets all reach the apex, so 64 is the optimal batch size for CDSCL.

For the regularization parameter $r$, we define eight values {0.1, 0.001, 0.0001, 1e-05, 1e-06, 1e-07, 1e-08}. Under the CDSCL framework, other parameters are fixed, the value of $r$ changes every time, and the obtained OA curves are shown in Fig. 11(c). The OA curves of River, Hermiston, and Bay Area rise gently until the value of $r$ is 0.001, then fall sharply. The OA curve of Farmland rises faster until the value of $r$ is 0.001, after which the volatility falls. From Fig. 11(c), it can be concluded that when the value of $r$ is 0.001, the OA curves of CDSCL on the four datasets all reach the highest point, so $r=0.001$ is the best regularization parameter for CDSCL.

For the setting of the size of the bottleneck layers of the PFEM, based on the principle that the size ratio of the first bottleneck layer and the second bottleneck layer is 2:1, experiments with different combinations are carried out on four datasets. Under the CDSCL framework, keeping other parameters unchanged and only changing the size of the bottleneck layer in the PFEM, the obtained OA curves are shown in Fig. 11(d). The OA curves of the four datasets are all the lowest at the ratio of 16:8, and then, decrease slowly after reaching the highest point at 32:16. When
the ratio of the two bottleneck layers of the PFEM is 16:8, the lowest OA of CDSCL on all datasets indicates that the setting of the bottleneck layer is too small, thus suppressing a large amount of effective change information. The slow decline of the OA curves of the four datasets indicates that as the size of the bottle mouth increases, some redundant or irrelevant information passes through the bottle mouth, which cannot achieve the effect of compressing the amount of information. From Fig. 11(d), it can be concluded that under the CDSCL framework, 32 and 16 are the best combination of the two bottleneck layer sizes of the PFEM.

G. Experiment on Training Sample Ratios for CDSCL

In this article, our proposed CDSCL does not require ground-truth labels in the CD process, and only needs a small number of high-confidence pseudolabels for the classification network training. To examine the CD performance of CDSCL in the case of fewer proportions of pseudolabeled training samples, we conduct experiments on CDSCL with different proportions of training samples on four real HSI datasets. The relationship between the proportion of pseudolabeled training samples and OA on the four HSI datasets is shown in Fig. 12. In experiments on the River dataset, taking 3% of high-confidence pseudolabels for training peaks the OA for CD. The reason for the further decline of the River curve is that its CVA predetection accuracy is not high. As the number of selected training samples increases, the confidence of pseudolabels decreases, which directly affects the CD performance. The reason why the curve of River decreases further is that its CVA predetection accuracy is not high, resulting in a larger percentage of selected pseudolabels and a lower confidence level, which directly affects the performance of CD. In experiments on the Farmland dataset, the CDSCL curve reaches a inflection point when the percentage of high-confidence pseudolabeled training samples is 4%. In experiments on the Hermiston dataset, the CDSCL curve reaches an inflection point
TABLE VII
COMPUTATIONAL COSTS OF DIFFERENT CD METHODS ON FOUR HSI DATASETS

| Dataset | Method                  | Training Time(s) | Testing Time(s) | Parameters |
|---------|-------------------------|------------------|-----------------|------------|
| River   | GETNET                  | 1642.44          | 154.90          | 154.18M    |
|         | SSA-SiamNet             | 182.91           | 9.55            | 107.12K    |
|         | SFBS-FFGNET             | 1265.94          | 40.61           | 11.58M     |
|         | CSSCL(without pre-trained) | 96.04           | 2.47            | 258.75K    |
|         | CDSCL                   | 377.10           | 2.58            | 391.62K    |
| Farmland| GETNET                  | 1120.15          | 47.84           | 93.42M     |
|         | SSA-SiamNet             | 164.98           | 5.22            | 97.83K     |
|         | SFBS-FFGNET             | 872.51           | 18.55           | 8.77M      |
|         | CSSCL(without pre-trained) | 108.26          | 1.45            | 253.25K    |
|         | CDSCL                   | 233.37           | 1.45            | 386.11K    |
| Hermiston| GETNET                  | 3268.57          | 161.02          | 232.04M    |
|         | SSA-SiamNet             | 278.24           | 6.18            | 116.62K    |
|         | SFBS-FFGNET             | 2454.01          | 28.77           | 11.58M     |
|         | CSSCL(without pre-trained) | 133.33          | 1.78            | 264.38K    |
|         | CDSCL                   | 288.67           | 1.86            | 397.25K    |
| Bay Area| GETNET                  | 1304.78          | 66.25           | 199.27M    |
|         | SSA-SiamNet             | 108.97           | 4.47            | 112.73K    |
|         | SFBS-FFGNET             | 913.52           | 27.65           | 11.58M     |
|         | CSSCL(without pre-trained) | 64.40           | 1.63            | 262.08K    |
|         | CDSCL                   | 259.33           | 1.75            | 394.95K    |

when the percentage of high-confidence pseudolabel training samples is 6%, that is, 6% of the pseudolabel training samples can make CDSCL achieve the best CD performance. In the experiments on the Bay Area dataset, when 5% of the high-confidence pseudolabeled training samples are taken for downstream change detection network training, the OA peaks and the curve becomes stable.

H. Computational Cost Analysis

To compare the computational cost of CDSCL and other deep learning-based methods more intuitively, we fix and unify the hyperparameters, and then, record the training time, testing time, and total parameters of these methods. Table VII shows the computational cost of all deep-learning-based CD methods on the four HSI datasets, where the number of training samples is set in accordance with Sections IV-B–IV-E, respectively. The training time is related to the number of training samples, the number of bands of the HSI dataset, and the complexity of the model. The test time is related to the total number of pixels in the HSI dataset, the number of bands, and the complexity of the model. The amount of parameters is only related to the number of bands in the HSI dataset and the complexity of the model. So, the best way to compare is between different methods on a single dataset. In Table VII, the training time, test time, and parameter amount of GETNET and SFBS-FFGNET are more than an order of magnitude ahead of other methods, because both methods use the difference matrix method, which increases the computational cost. The computational cost of CDSCL without pretrained, SSA-SiamNet, CDSCL is in the same order of magnitude. The operation cost of CDSCL without pretrained is the lowest, and the operation cost of CDSCL is slightly higher, which is caused by the inclusion of pretrained structure in CDSCL. In the River dataset, the reason for the higher training time of CDSCL is that the total number of pixels in this dataset is large, while the preraining process of CDSCL is performed pixel by pixel, so the training time is longer. From the global perspective of Table VII, the computational cost of CDSCL is acceptable among all deep-learning-based contrasting methods.

I. Experiment Summary

In this article, River, Farmland, Hermiston, and Bay Area are used as the HSI datasets for the experiments. The two algebraic analysis methods we chose in the experiment are CVA and PCA-CVA. Three comparison methods of CD based on deep learning are GETNET, SSA-SiamNet, and SFBS-FFGNET. To more intuitively reflect the advantages of the proposed self-supervised contrastive learning pretrained model, we design an ablation comparison experiment method CDSCL without pretrained. Fig. 13 shows the comparison chart of the OA of CD for all experimental methods on four different datasets. In the River dataset, CDSCL performed the best, CVA performed the worst, and CDSCL without pretrained performed at an intermediate level. In the Farmland dataset, all contrasting methods performed well, the best is SFBS-FFGNET, but still lower than CDSCL, and the contrast of ablation experiments is obvious. In the Hermiston dataset, PCA-CVA, SSA-SiamNet, and SFBS-FFGNET perform well in CD, where the OA of SFBS-FFGNET is basically the same as that of CDSCL, and CVA is the worst performer. In the Bay Area dataset, CVA and PCA-CVA are the worst performers, which is related to the overly complex scene of this dataset. SSA-SiamNet, SFBS-FFGNET, and CDSCL without pretrained all perform well on the Bay Area dataset, but CDSCL is still far ahead. To sum up, the CD performance of CDSCL on the four real hyperspectral datasets is the best among all comparison methods, and greatly exceeds that of CDSCL without pretrained, reflecting the superiority of the self-supervised comparative learning pretrained model.

V. Conclusion

In this article, a novel CD framework named CDSCL is proposed. In CDSCL, to improve the generalization ability of the model to noise, we design a data augmentation strategy based
on Gaussian noise. To remove redundant or irrelevant details in spectral information, we design the PFEM based on the IB theory. To make the upper and lower branch features of the siamese network closer, we design a contrastive loss function based on the Pearson correlation coefficient and negative cosine correlation. The complete CDSCL includes a self-supervised learning pretrained model and a CD classification network, where the classification network is based on a weakly supervised training approach. The CD framework we proposed in this article has experimented on four real HSI datasets. Experimental data show that CDSCL has the best OA, Kappa, and is the most robust among all deep learning-based methods. Furthermore, in this article, we design an ablation experiment to demonstrate the importance of self-supervised contrastive learning of pretrained models in CDSCL.

REFERENCES

[1] Y. Liu, C. Chen, Z. Han, L. Ding, and Y. Liu, “High-resolution remote sensing image retrieval based on classification-similarity networks and double fusion,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 1119–1133, Mar. 2020.

[2] 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.

[3] G. Jianya, S. Haigang, M. Guorui, and Z. Qingmin, “A review of multi-temporal remote sensing data change detection algorithms,” Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci., vol. 37, no. B7, pp. 757–762, 2008.

[4] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, “Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 5, pp. 2733–2748, May 2016.

[5] B. Tu, C. Zhou, J. Peng, G. Zhang, and Y. Peng, “Feature extraction via joint adaptive structure density for hyperspectral imagery classification,” IEEE Trans. Instrum. Meas., vol. 70, Jan. 2021, Art. no. 5009616.

[6] B. Tu, X. Liao, C. Zhou, S. Chen, and W. He, “Feature extraction using multitask superpixel auxiliary learning for hyperspectral classification,” IEEE Trans. Instrum. Meas., vol. 70, Oct. 2021, Art. no. 5018216.

[7] B. Tu, Y. Chen, S.-Y. Cao, H.-L. Shen, and J. Li, “Three-channel infrared imaging for object detection in haze,” IEEE Trans. Instrum. Meas., vol. 71, Apr. 2022, Art. no. 5008513.

[8] Y. Wang, B. Du, L. Ru, C. Wu, and H. Luo, “Scene change detection VIA deep convolution canonical correlation analysis network,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2019, pp. 198–201.

[9] C. Wu, B. Du, and L. Zhang, “A subspace-based change detection method for hyperspectral images,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 6, no. 2, pp. 815–830, Apr. 2013.

[10] N. M. Noor, A. Abdullah, and M. N. H. Manzahia, “Land cover change detection analysis on urban green area loss using GIS and remote sensing techniques,” Plan. Malaysia, vol. 11, no. 3, pp. 125–138, 2013.

[11] M. Mir, S. Maleki, and V. Rahdari, “Application of landscape ecology in spatio-temporal change detection of arid regions, case study: Sistan plain,” Iranian J. Appl. Ecol., vol. 10, no. 2, pp. 67–81, 2021.

[12] J. Xiao et al., “Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing,” Landscape Urban Plan., vol. 75, no. 1/2, pp. 69–80, 2006.

[13] M. Hasanlou and S. T. Seydi, “Hyperspectral change detection: An experimental comparative study,” Int. J. Remote Sens., vol. 39, no. 20, pp. 7029–7083, 2018.

[14] W. A. Malilla, “Change vector analysis: An approach for detecting forest changes with landsat,” in Proc. LARS Symp., 1980, Art. no. 385.

[15] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemometrics Intell. Lab. Syst., vol. 2, no. 1–3, pp. 37–52, 1987.

[16] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, “A novel sequential spectral change vector analysis for representing and detecting multiple changes in hyperspectral images,” in Proc. IEEE Geosci. Remote Sens. Symp., 2014, pp. 4656–4659.
Liangzhen Liu received the B.S. degree in electronic information engineering from the Hunan Institute of Science and Technology, Yueyang, China, in 2019, where he is currently working toward the M.S. degree in information and communication engineering. His research interests include image processing, hyperspectral image classification, change detection, and reconstruction.

Shulun Tan is currently working toward the B.S. degree in electronic information engineering with the Hunan Institute of Science and Technology, Yueyang, China. His research interests include hyperspectral image processing and change detection.

Guoyun Zhang (Member, IEEE) received the B.S. degree in industrial electrical and automation from Xiangtan University, Xiangtan, China, in 1993, and the M.S. and Ph.D. degrees in control theory and control engineering from Hunan University, Changsha, China, in 2002 and 2006, respectively. In 1995, he joined the Hunan Institute of Science and Technology, Yueyang, China. He spent half a year (January 2013 to June 2013) as a Visiting Scholar with George Fox University, Newberg, OR, USA. He is currently a Full Professor with the School of Information Science and Technology, Hunan Institute of Science and Technology. His research interests include digital signal processing, application technology, and hyperspectral image processing.

Wujing Li received the B.E. degree in software engineering and Ph.D. degree in computer science and technology from Sichuan University, Chengdu, China, in 2007 and 2012, respectively. He is currently a lecturer with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China. His main research interests include image denoising, image restoration, image enhancement, and hyperspectral image processing.

Bing Tu (Member, IEEE) received the M.S. degree in detection technology and automatic equipment from the Guilin University of Technology, Guilin, China, in 2002, and the Ph.D. degree in mechanical engineering from the Beijing University of Technology, Beijing, China, in 2013. From 2015 to 2016, he was a Visiting Researcher with the Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA, which is supported by the China Scholarship Council. He is currently a Professor with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China. His research interests include sparse representation, pattern recognition, and analysis in remote sensing.