Unsupervised Hyperspectral Image Change Detection via Deep Learning Self-Generated Credible Labels

Qiuxia Li, Hang Gong, Haishan Dai, Chunlai Li, Zhiping He, Wenjing Wang, Yusen Feng, Feng Han, Abudusalamu Tuniyazi, Haoyang Li, and Tingkui Mu

***Abstract*—Change detection (CD) aims to identify differences in scenes observed at different times. Hyperspectral image (HSI) is preferred for the understanding of land surface changes, since it can provide essential and unique features for CD. However, due to the high-dimensionality and limited data, the HSI-CD task is challenged. While model-driven CD methods are hard to achieve high accuracy due to the weak detection performance for fine changes, data-driven CD methods are hard to be generalized due to the limited datasets. The state-of-art method is to combine a single model-driven method with a data-driven convolutional neural network (CNN). Wherein the pseudolabels can be generated automatically by the model-driven method and then fed to CNN for training. However, the final detection accuracy is limited by the model-driven method which produces pseudolabels with one-sidedness and low accuracy. Therefore, the generation of credible pseudolabels is anticipated and crucial for such a combination. Herein, a novel strategy, the combination of two complementary model-driven methods, structural similarity (SSIM) and change vector analysis (CVA), is proposed to generate credible labels for training a subsequent CNN. The results show that the final accuracy is higher than that of SSIM and CVA. The main contributions of this article are threefold: First, a new paradigm for generating credible labels is proposed. Second, SSIM is used for the first time for HSI-CD tasks. Third, an unsupervised end-to-end framework is presented for the HSI-CD. Experimental results demonstrate the effectiveness of the proposed framework.**

***Index Terms*—Change detection (CD), deep learning, heterogeneous images, hyperspectral image (HSI), structural similarity (SSIM).**

Manuscript received May 30, 2021; revised July 4, 2021; accepted August 26, 2021. Date of publication September 1, 2021; date of current version September 20, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61775176 and Grant 62175196, in part by the National Major Special Projects of China under Grant GFZX04014308, in part by the Shaanxi Province Key Research and Development Program under Grant 2021GXLH-Z-058, Garnt 2020GY-131, and Grant 2021SF-135, in part by the Innovation Capability Support Program of Shaanxi under Grant 2021TD-57, in part by the Fundamental Research Funds for the Central Universities under Grant xjh012020021, and in part by the Natural Science Foundation of Shanghai under Grant 8ZR1437200. *(Corresponding author: Tingkui Mu.)*

Qiuxia Li, Hang Gong, Wenjing Wang, Yusen Feng, Feng Han, Abudusalamu Tuniyazi, Haoyang Li, and Tingkui Mu are with the Ministry of Education Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, Shaanxi Province Key Laboratory of Quantum Information and Quantum Optoelectronic Devices, School of Physics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: lqx0324@stu.xjtu.edu.cn; gh0000@stu.xjtu.edu.cn; wjwang@stu.xjtu.edu.cn; 3207650009@qq.com; 573378205@qq.com; abdusalam1112@stu.xjtu.edu.cn; lihaoyang@stu.xjtu.edu.cn; tkmu@mail.xjtu.edu.cn).

Haishan Dai is with the Shanghai Institute of Satellite Engineering, Shanghai Academy of Spaceflight Technology, Shanghai 201109, China (e-mail: dhs1314@126.com).

Chunlai Li and Zhiping He are with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China (e-mail: lichunlai@mail.sitp.ac.cn; hzping@mail.sitp.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3108777

I. Introduction

CHANGE detection (CD) aims to identifying the differences between bitemporal images obtained over the same geographical region at different times. Differential monitoring of bitemporal remote sensing images is essential for understanding land surface changes. It has been extensively applied to various applications, such as ecosystem monitoring [1], [2], land cover mapping [3], urban expansion research [4], [5], and resource management [6]. The complete CD process requires a series of comprehensive processing steps, which include selecting training data, accurately preprocessing image, extracting image features, designing the CD algorithm, and evaluating the CD performance.

The rapid development of sensor technologies facilitates the formulation of hyperspectral images (HSIs) [7]. The rich spectral information from HSI is helpful to achieve target detection, classification, as well as CD. HSI provides a dense sampling of target spectral signatures over a wide spectral wavelength range, which makes it possible to monitor land-cover precisely at a fine spectral scale [8]. The spectral information can distinguish the spectrally similar materials and describe the finer spectral changes [9], [10]. However, effectively exploiting the HSI data to accurately monitor land-cover changes is still a challenging task.

The existing hyperspectral image CD (HSI-CD) methods can be broadly classified into model-driven CD methods and data-driven deep learning CD methods [11]. Most model-driven CD methods do not rely on groundtruth change maps. Therefore, model-driven methods are still utilized at high frequencies, although evolving new techniques have been used for CD tasks. However, there are two challenges concerning model-driven methods. The first challenge is that the use of a single model is one-sided, as each model has unique limitations. A feasible strategy is to consider combining different model-driven methods in the design of CD algorithms. The key question is how to combine them. The second challenge is that the complex characteristics of HSI make it difficult to obtain the desired results by model-driven methods. To find a new way, deep learning with the convolutional neural network (CNN) is utilized extensively for HSI-CD recently.

Many research showed that deep learning has swept across the field of computer vision and remote sensing image interpretation due to the significant advantages in deep feature representation and nonlinear problem modeling [12], [13]. With end-to-end structure, deep learning methods can obtain depth features and global information from bitemporal images, which can be used to give change maps directly [14]. In general, supervised deep learning methods require a sufficient amount of ground reference datasets for training CNN. However, collecting ground reference data for diachronic HSIs is an expensive and time-consuming process. Therefore, the design of effective unsupervised automatic HSI-CD algorithms remains a critical problem.

For the problem of unsupervised HSI-CD, existing works usually use pseudolabel to train CNN [15]. The first step of these methods is to produce pseudolabels based on the *difference image* (*DI*) obtained by a single model-driven method. The second step is to train the formulated CNN. Such approaches intend to use the learning ability of the CNN and obtain a better result than the model-driven method. The drawbacks of such methods are accuracy limitation and criterion limitation. The accuracy limitation means that the wrong samples in the pseudolabels directly increase the error rate of the final detection results. It is difficult to remove this limitation even with the use of noise removal models for pseudo labels. The criterion limitation means that the discriminant criteria of CNN and model-driven methods are very similar because both of them rely on the same model-driven method. Therefore, it is meaningful to explore the combination of model-driven methods and deep learning methods for the achievement of HSI-CD concisely and efficiently.

Another challenge of the HSI-CD task is the great difficulty of detecting edge regions in the image, especially for low- and medium-resolution remote sensing images. In the general HSI-CD method, the change map is obtained from the discriminant of *DI*, and it is the binary map characterizing whether a pair of pixel points from T1 and T2 changes. Most of the remote sensing images with low resolution have mixed image elements in the edge region. When a pair of changed pixels contains a mixed image element, the *DI* is low. Obviously, the lower intensity change information in *DI* is difficult to detect by the model. The pseudolabel method also faces this challenge.

The accuracy and automation of HSI-CD would be effectively improved by solving the above problems. In this article, HSI-CD is considered as a segmentation problem. An unsupervised HSI-CD method using CNN is presented. The proposed framework is different from the supervised method, which learns features from the manually labeled data. The proposed method uses pseudolabels to train the network in an end-to-end manner. Different from the pseudolabels generated from a single model-driven CD method, the proposed framework balances the credits of pseudolabels generated from two selectively model-driven CD methods, changed vector analysis (CVA) and structural similarity (SSIM) [16]. Such pseudolabels are not only highly credible, but also comprehensive, as they contain both strongly changed pixels and weakly changed pixels. Experimental results demonstrate the effectiveness and flexibility of the proposed framework. The major contributions of this article can be summarized as follows.

1) A new paradigm using two complementary model-driven methods for the generation of credible labels is proposed.
2) SSIM is used for the first time for the HSI-CD task.
3) An unsupervised end-to-end framework is proposed for the HSI-CD.

The rest of this article is organized as follows. The related work is introduced in detail in Section II. Section III presents the proposed methodology. Experiments and discussions are given in Section IV. Finally, Section V concludes the article.

II. Related Work

The binary CD is one of the typical and popular CD technologies of recent decades [17]. Its objective is to classify all pixel pairs in bitemporal images into change class and no-change class. This section discusses model-driven CD methods and deep learning CD methods separately.

Model-driven CD methods can be broadly categorized into four classes: image arithmetical-based algorithms [18]; image transformation-based algorithms [19]; postclassification algorithms [20]; and methods focused on HSI-CD [21].

1) Image arithmetical-based algorithms directly compare pixel values in a pixel-wise manner with arithmetical operations from bitemporal images to produce *DI*. Thresholds of *DI* are applied to classify pixels into the changed class or unchanged class. The most commonly used arithmetical operations are image subtraction [22], image regression [23], and image rationing [24]. The most widely used algorithm is CVA [25], which is a typically unsupervised method using spectral vector subtraction. Some modified CVA algorithms have also been proposed [26], [27]. However, CVA algorithms do not consider contextual information for pixels that are treated independently, and thus they are sensitive to noise and misregistration errors.
2) Image transformation-based algorithms transform image spectral into a specific feature space to emphasize changed pixels and suppress unchanged ones. Principal component analysis (PCA) [28] and iterative reweighted multivariate alteration detection [29], [30] are the most common algorithms in the image transformation-based algorithms for dimensionality reduction. In addition, slow feature analysis (SFA) can extract slowly changing features from time series [31], [32]. SFA can be used for CD by suppressing invariant pixels and highlighting the changing pixels [33]. These methods highly depend on empirically designed algorithms for the extraction of discriminative features, which are usually hard to achieve satisfying results on high-resolution images.
3) Postclassification algorithms rely on a priori knowledge for training the classifier [34]. Bitemporal images are independently classified first. Then, the changed areas are extracted through a direct comparison of the classification results. Consequently, postclassification algorithms are highly sensitive to the availability of prior knowledge as well as the selection of the classifier.

4) The methods that focus on HSI-CD can be at pixel level. For example, Chen and Wang [35] proposed a spectrally–spatially regularized low-rank and sparse decomposition model. Wu *et al.* [36] proposed a hyperspectral anomalous CD method based on joint sparse representation. In contrast, there are also methods at the sub-pixel level. In [37], HSI-CD based on spectral unmixing was investigated and validated on carefully prepared synthetic datasets and also with real datasets. In [38], a linear mixture model was used to analyze the estimated endmembers and abundances in each HSI.

Deep learning CD methods transform the image into a feature space using CNN and distinguish changes. It can be divided into two categories according to how they manage bitemporal images: early-fusion methods and postfusion methods. Daudt *et al.* [39] proposed supervised three fully convolutional network architectures for multitemporal CD including early-fusion and post-fusion.

The early-fusion method stack all bitemporal images as input [40]. Peng *et al.* [41] proposed a supervised end-to-end early-fusion framework for very-high-resolution satellite image CD. It extracted the features of the fused image using Unet++ without pre-training.

The postfusion methods take each of the bitemporal images as input [42]. The features of both images are extracted by each of the two independent pipelines in the network, and then the features are fused in the final part of the network to obtain the change map. Zhang *et al.* [43] proposed a deeply supervised post-fusion method for CD in high-resolution bitemporal remote sensing images. The method employed a deeply supervised image fusion network consisting of a fully convolutional two-branch structure and a deeply supervised difference discrimination network. The former was used for depth feature extraction and the latter for depth feature discrimination. Recently, several novel HSI-CD method based on deep learning was proposed. Li *et al.* [44] proposed a noise modeling-based unsupervised postfusion HSI-CD framework, which adopted a two-branch full convolutional network to extract the features of bitemporal images separately. The unsupervised noise modeling module can alleviate the accuracy limitation caused by pseudolabels. Song *et al.* [45] proposed an unsupervised post-fusion HSI-CD architecture based on a recurrent 3D fully convolutional network. The network, consisting of spectro-spatial and temporal modules, was able to extract the spectro-spatial features of HSIs while recording and analyzing the multi-temporal variation information of HSI. Wang *et al.* [15] proposed a general end-to-end 2-D CNN (GETNET) unsupervised post-fusion HSI-CD framework, in which mixed-affinity matrices were formed and features were extracted for classification. The pseudolabels of the GETNET were selected from the result of the single CVA method. Du *et al.* [46] proposed an unsupervised post-fusion DSFA framework, which extracted invariant paired pixels from the single CVA's result as training samples. The two trained depth networks were used to transform the bitemporal images separately. The invariant pairwise pixels were suppressed and the changed pairwise pixels were highlighted using SFA constraints.

III. Methodology

In this section, we introduce the proposed unsupervised HSI-CD framework in detail. The framework structure is shown in Fig. 1, which shows the process of acquiring a change map from the prechange image T1 and the post-change image T2. The proposed approach consists of three main components: obtaining change maps using CVA and SSIM, respectively; generating training samples to feed the specified CNN; and training the CNN to obtain the CD results.

Module 1: Obtaining change maps using CVA and SSIM, respectively. CVA algorithm is an algebraic operation based on pixel points. SSIM is a pixel-block-based detection method. For CVA, each pixel point is independent of the neighboring pixel points, whereas for SSIM, the neighboring pixel points affect the result.

Module 2: Generating the training set to feed the specified CNN. The pixels with the same value of both change maps are selected as high confidence pixels. Pixels detected as changing by both algorithms are used to make a positive sample set, and pixels detected as unchanging by both algorithms are used to make a negative sample set. The final training set is constructed based on the two sample sets. In this case, the training set labels are called self-generated credible labels because there is no manual inference and discrimination.

Module 3: Training the CNN to obtain the CD results. The identification rule of CD is mastered by the model via learning the sufficient training set. The optimal model can directly infer the complete change map from the biphasic images.

A. Module 1: Obtaining Change Maps Using CVA and SSIM

Both CVA and SSIM use clustering of *DI* to obtain change maps. The two algorithms obtain *DI* based on different ways and implement clustering based on the same Kmeans.

CVA is a common way to identify binary information based on comparison operators, which is a typically unsupervised method using spectral vector subtraction [27]. For the CD task with binary classification, CVA first generates the magnitude of change as *DI* by spectral vector subtraction, then obtains the change map by clustering. Considering T1 and T2 as two input HSIs, the *DI* of CVA is computed by

$$DI_{(i,j)} = \sqrt{\sum_{k=1}^{K} \{[P_{(i,j,k)}(\mathrm{T2}) - P_{(i,j,k)}(\mathrm{T1})]^2\}} \quad (1)$$

where i and j are the spatial coordinates, $P_{(i,\,j,\,k)}(\mathrm{T1})$ and $P_{(i,\,j,\,k)}(\mathrm{T2})$ are the pixel values of band k from T1 and T2, respectively. In addition, K is the number of bands. Then, the change map CM1 shown in Fig. 2 is obtained by K means clustering of the *DI*.

SSIM was first introduced for image similarity measurement methods based on structural information degradation [16]. It is an advanced method for comparing the structure between reference and distorted images. It can express the degree of similarity between two images. Therefore, SSIM can determine

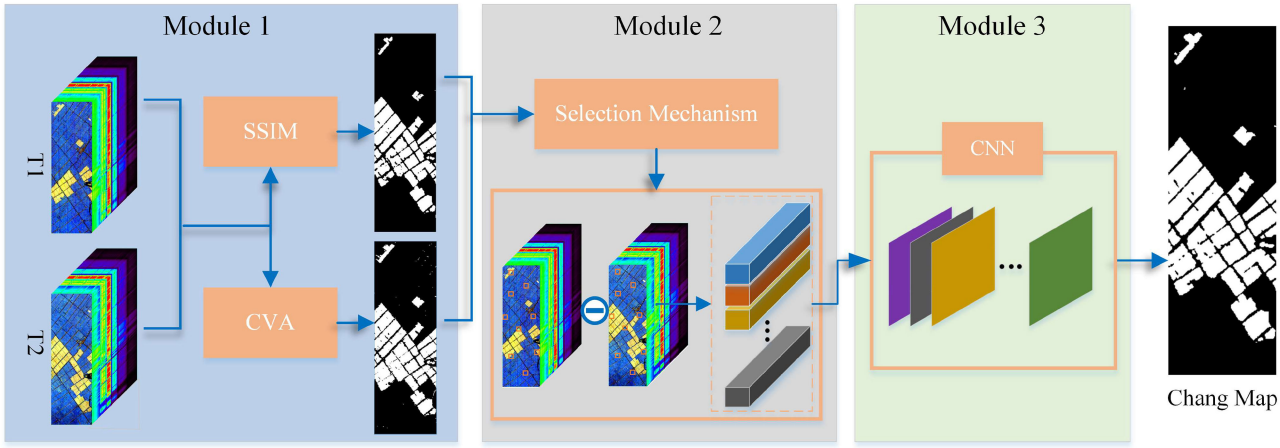


Fig. 1. Illustration of the proposed HSI-CD framework. Module 1: obtaining change maps using two model-driven methods, respectively. Module 2: generating the training set to feed the specified CNN. Module 3: training the CNN to obtain the CD results.

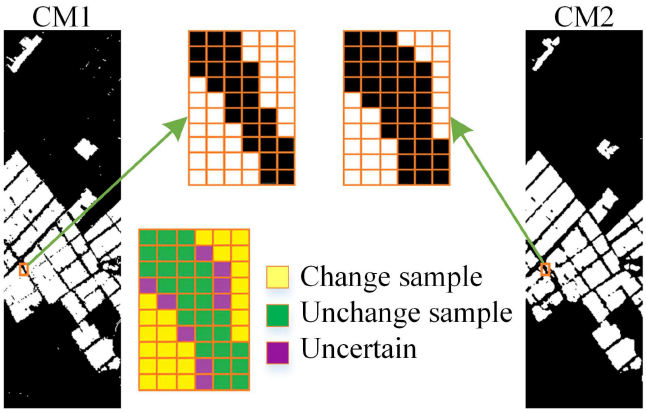


Fig. 2. Selection mechanism for high-confidence pixel points.

whether a pair of pixels has changed or not in CD tasks. The SSIM index is, based on the statistical similarity measurements, defined as follows:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}. \tag{2}$$

To take into account the influence of neighboring pixels, the sliding windows are used to perform the selection of the parts of the image. The average value of each window is the score of its central pixel. Thus, the inputs x and y are two windows of image. The statistical parameters μ_x, μ_y, σ_x^2, and σ_y^2 are the means and variances of x and y. The σ_{xy} represents the statistical covariance between x and y. C_1 and C_2 are constants. The three measurements considered in building the similarity between the x and y are given as follows.

Luminance:

$$L(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}. \tag{3}$$

Contrast:

$$C(\mathrm{x},\mathrm{y})=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}. \tag{4}$$

Structure:

$$S(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}. \tag{5}$$

The constants C_1 and C_2 are given as follows: $C_1=(K_1G)^2$ and $C_2=(K_2G)^2$, with K_1 and $K_2<<1$, $C_3=C_2/2$, and G is the dynamic range of the pixel values ($G=255$ for 8-b image). The small constants C_1, C_2, and C_3 are used to avoid the case of the denominator is zero, and have almost no effect on the accuracy of the final result. Let the center of the sliding window traverse the entire image to get the SSIM matrix as large as the original image. For HSI, the SSIM score of a pixel is the average of the scores of all bands, which is computed by

$$\mathrm{SSIM}_{(\mathrm{HSI})}=\frac{1}{K}\sum_{k=1}^{K}(\mathrm{SSIM}_{(k)}) \tag{6}$$

where $\mathrm{SSIM}_{(k)}$ is the SSIM matrix of the kth band in HSI and K is the number of bands. Thus, DI of a pair of pixels can be expressed as

$$DI=1-\mathrm{SSIM}_{(\mathrm{HSI})}. \tag{7}$$

Similarly, the change map CM2 shown in Fig. 2 is also obtained by K means clustering of the DI.

B. Module 2: Generating Credible Labels and Dataset

This module obtains the training set from the obtained change maps CM1 and CM2 shown in Fig. 2, using the two model-driven methods. It is a bridge that connects the model-driven method with deep learning. The module consists of three steps: selecting high confidence pixels, generating two kinds of sample sets, generating the training set. As seen, this automatically generated training set is a pseudotraining set without any human labor.

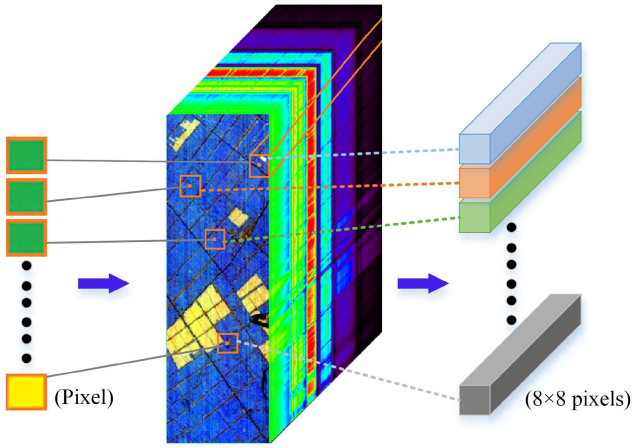


Fig. 3. Process of a sample set generation.

Since the two obtained change maps, CM1 and CM2, are not exactly correct, the high-confidence samples need to be selected by balancing them. The selection mechanism is shown in Fig. 2. Both CM1 and CM2 are binarized images with the same shape as T1 and T2. On the change map, the pixel with a value of 1 indicates that the ground object at the corresponding location has changed. The pixel with a value of 0 denotes that the ground object at the corresponding location has not changed. The pixel points with the same value on both change maps are the ones with higher confidence. The selection strategies can be formulated as

$$\begin{cases} \text{Changed sample if CM1} = \text{CM2} = 1 \\ \text{Unchanged sample if CM1} = \text{CM2} = 0 \\ \text{Uncertain if CM1} \neq \text{CM2} \end{cases}. \tag{8}$$

As seen, in our selection mechanism, only a portion of pixels that are simultaneously recognized by the two algorithms can be considered credible pseudolabels. That is, other one-sided pixels of change maps are uncertain and discarded temporally. However, we can use the selected credible pseudolabels to train the CNNs, and then let the uncertain pixels be predicted unambiguously. The sample set of T1 and the sample set of T2 are generated with these pixel points as the center. The generation process is shown in Fig. 3. The final training set consists of two sample sets combined in different ways, such as subtractive combination and concatenation combination.

C. Module 3: Training the CNN to Obtain the CD Results

The CNN plays the role of extracting features, discriminating features, fusing features, and mapping classification in the framework. The ability of the CNN to detect change information is continuously improved by learning the training set adequately. The final change map is predicted by the trained CNN model.

In order to fully demonstrate the proposed framework, three methods, including a postfusion method and two early fusion methods, are discussed. For the convenience of description, the post-fusion method is named model P, the concatenation early-fusion method is named model E1, and the subtractive early-fusion method is named model E2. Different methods employ different generating methods of training sets and different frames of CNNs. In all three models, the input to the CNN is 3-D data.

The inputs of model P are the two independent sample sets from the bitemporal images. The sample sets are the original information of the images. Therefore, the selected network consists of a feature extraction module and a feature fusion module. The integrated CNN structure is shown in Fig. 4. Among them, the two-branch Unet networks are responsible for feature extraction [47], and the feature fusion module is responsible for feature fusion, difference discrimination, and mapping classification. Thus, in the complete network, backpropagation is performed from the last disparity discrimination layer to the first feature extraction layer. The advantage of this model is that the original features of the image are utilized. Bitemporal images are transformed into the same feature space by the Unet networks for effective difference discrimination.

The inputs of model E1 are the concatenation of the two sample sets. The two sample sets are denoted as Sets 1 and 2. The specific structure of the model is shown in Fig. 1. The image fusion strategy is expressed as (9). Therefore, the model just uses a single Unet network as shown in Fig. 4. It is responsible for both the original feature extraction and the difference discrimination at the same time.

The input of model E2 is the absolute difference between the two sample sets, which represents the intensity change map. The image fusion strategy is shown in (10). The network only needs to perform the difference discrimination and mapping. Therefore the model E2 in Fig. 4 just contains a single convolutional layer as the core, takes *DI* as input and change map as output. The advantage of such network is that it is easy to train and suitable for simple binary classification tasks with a small training set due to a small number of trainable parameters. According to the input information and network structure of the network, the model E2 will get the test result speedily.

$$\text{Input} = \text{concat}(\text{Set1}, \text{Set2}) \tag{9}$$

$$\text{Input} = \text{abs}(\text{Set1} - \text{Set2}). \tag{10}$$

In addition, when only a small number of training samples are available, data enhancement is critical to generalize the network. To increase the diversity of the sample sets, a series of image preprocessing methods are used for data enhancement, such as image noise addition, image random brightness, and image random contrast.

IV. Experiments

A. Datasets

To the best of our knowledge, currently, there are few public datasets available for HSI-CD. In this article, two datasets are used to validate the proposed method. The two datasets have their own characteristics, such as data size, change type, time interval.

The two datasets are from earth observing-1 (EO-1) hyperion hyperspectral sensor. The EO-1 hyperion sensor provides

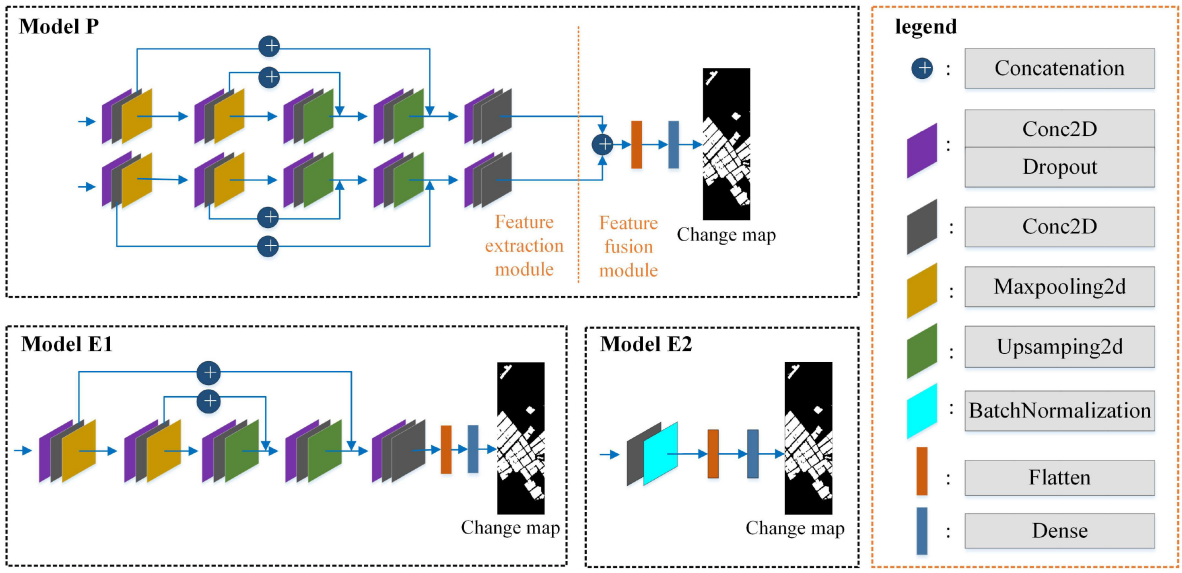


Fig. 4. Network structure of three models P, E1, and E2.

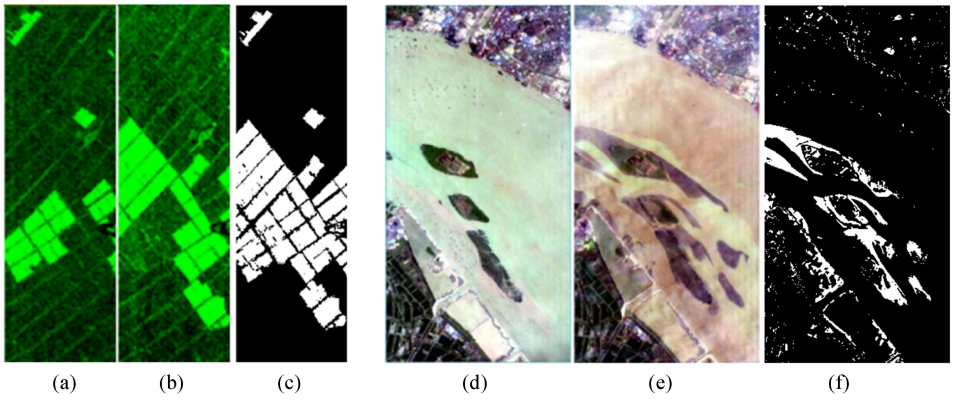


Fig. 5. Experimental datasets. (a) Farmland imagery on May 3, 2006. (b) Farmland imagery on April 23, 2007. (c) groundtruth change map of Farmland dataset. (d) River imagery on May 3, 2013. (e) River imagery on December 31, 2013. (f) Groundtruth change map of River dataset.

HSIs with a swath width of 7.7 km and a spectral range of 0.4–2.5 μm [48], [49]. In the meantime, it offers a spatial resolution of 30 meters and a spectral resolution of 10 nm [50]. Each dataset consists of three images, which are bitemporal images and a groundtruth change map. The bitemporal images with time difference contain change information. The white pixels in the binarized groundtruth change map indicate the changed classes and the black pixels indicate the unchanged classes. The two datasets are described in detail as follows.

The first is a Farmland dataset and its groundtruth change map [51]. It features bitemporal HSIs covering the area near the city of Yancheng, Jiangsu province, China, as shown in Fig. 5(a) and (b). The two HSIs were acquired on May 3, 2006, and April 23, 2007, respectively. The HSI has 155 bands with a spatial size of 450 $\times$ 140. Fig. 5(c) shows the corresponding groundtruth change map, which is obtained by manually marking based on the specialist knowledge. The main changes in this dataset are the conversion between bare soil and crops. The changed regions contain 18 277 pixels, whereas the unchanged regions contain 42 723 pixels.

The second is a River dataset and its groundtruth change map [15]. The bitemporal images were acquired on May 3, 2013, and December 31, 2013, respectively, in Jiangsu province, China. Fig. 5(d) and (e) shows the bitemporal HSIs, which consist of 198 bands with a spatial size of 463 $\times$ 241. Fig. 5(f) shows the corresponding groundtruth change map. The changed regions contain 12566 pixels, whereas the unchanged regions contain 99017 pixels.

B. Parameter Setup

This experiment is implemented on the Windows system, TensorFlow 2.0 framework, and 2080Ti graphics processing unit. To achieve optimal results for the three models P, E1, and E2, the optimizer SGD is used, and just 30 epochs are trained. The performance of the proposed methods are compared with the five existing model-driven CD methods, including

CVA [25], SSIM [16], image ratioing (imageratio) [52], image regression (ImageRegr) [23], and dynamic PCA (DPCA) [53], and two data-driven methods, including early-fusion FCN [54], and postfusion Siamese network [42]. Among them, the results of CVA and SSIM are obtained by Python, and the results of other model-driven CD methods are obtained by Matlab. Both data-driven methods use CVA to generate pseudolabels. The size of the sliding window for SSIM is taken as 3, and the appropriate odd value is recommended in the ranges of 3–9.

C. Evaluations Measures

Evaluation measures are critical to analyzing the performance of CD methods. To verify our proposed method, three evaluation metrics are applied, namely overall accuracy (OA), F1-score, and Kappa coefficient. In the calculation of evaluation metrics, four indexes are adopted: true positives (TPs), the number of changed pixels that are correctly detected; true negatives (TNs), the number of unchanged pixels that are correctly detected; the false positives (FPs), the number of unchanged pixels that are wrongly detected as changed pixels; and the false negatives (FNs), the number of changed pixels that are detected as unchanged pixels. In the CD task, Precision and Recall also have been commonly used as evaluation metrics. A large value of Precision denotes a small number of false alarms, and a large value of recall represents a small number of missed detections. F1-score combines both precision and recall metrics. F1-score and OA reveal the overall performance, where their larger values will lead to better performance. The larger value of the Kappa coefficient, the better the performance of the corresponding method. First, Precision and Recall are defined as

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \tag{12}$$

Then, OA and F1-score are defined as

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{13}$$

$$\text{F1} - \text{score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}. \tag{14}$$

The Kappa coefficient is employed as a consistency test, which is an index to evaluate the accuracy of classification. In the CD task, the Kappa coefficient indicates the consistency between the predicted change map and the groundtruth change map. The Kappa coefficient is denoted as

$$\text{Kappa} = \frac{\text{OA} - \text{OP}}{1 - \text{OP}} \tag{15}$$

$$\text{OP} = \frac{(\text{TP}+\text{FP})(\text{TP}+\text{FN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})^2} + \frac{(\text{FN}+\text{TN})(\text{FP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})^2}. \tag{16}$$

D. Experimental Results and Discussions

In this section, we first analyze the complementarity of the used two model-driven methods, which are the key to the proposed methods. Then, the three models, P, E1, and E2, are compared. The best one is selected. Finally, the selected model is analyzed in detail.

The complementarity of the two model-driven methods, SSIM and CVA, is the key to break the accuracy limit of every single model. Since both algorithms use Kmeans to obtain the change map by clustering of the *DI*, comparing the *DI* and the change map obtained by the two algorithms can illustrate their complementarity. First, there are differences between the two algorithms, mainly due to their detection units and discriminative mechanisms. Fig. 6 shows the *DIs* of the two algorithms for each of the four bands (7, 57, 117, 147). The darker color represents the more intense variation. Therefore, strong changes refer to change information with high value in *DI*, and weak changes refer to change information with low value in *DI*. Obviously, the differences of *DIs* obtained by CVA and SSIM are significant, especially for the first three bands. Besides, the *DI* differences of SSIM for each band are larger than that of CVA. These phenomena demonstrate the variability of the two algorithms at each band.

The Kmeans is used to obtain the change map by clustering of the *DI*. Fig. 7 shows the visualization of the change maps obtained by the two algorithms. The four colors, respectively, represent the pixels corresponding to the four metrics of the evaluation measures. While green and gray represent TP and TN, respectively, purple and yellow indicate FP and FN, respectively. As seen, both algorithms can obtain full change maps but differ significantly in the detection of edge regions and discrete points. The completeness and differences of the change maps obtained from the two algorithms ensure that they are complementary to each other. Completeness ensures that the training samples are sufficient to train neural networks for getting strong detection ability. The differences in details provide the possibility of breaking the accuracy limitation and the criterion limitation. Therefore, if the training samples are pixel points that are doubly affirmed by these two algorithms, they will ensure a high confidence level. The models trained by the high confidence samples will perform well on other unselected samples.

Figs. 8 and 9 show the performances of the three models P, E1, and E2 on the Farmland dataset and the River dataset, respectively. As seen, model P with two-branch Unet networks achieves modest performance, model E1 with the deeper network leads to the worst performance, and model E2 with the single convolution layer provides the most accurate and fast performance. It is important to note that the model P implements the combination of low resolution and high resolution information and difference discrimination. The model E1 has both original feature extraction and difference discrimination capabilities. The model E2 has only difference discrimination capabilities. As seen, the difference discrimination capability of the used network is important, since the model E2 has the best performance. From another perspective, the model E1 is hard to combine the features extraction and difference discrimination, and thus, obtains the worst results. In contrast, the model E2 is characterized by its small structure, high accuracy, and low time consumption.

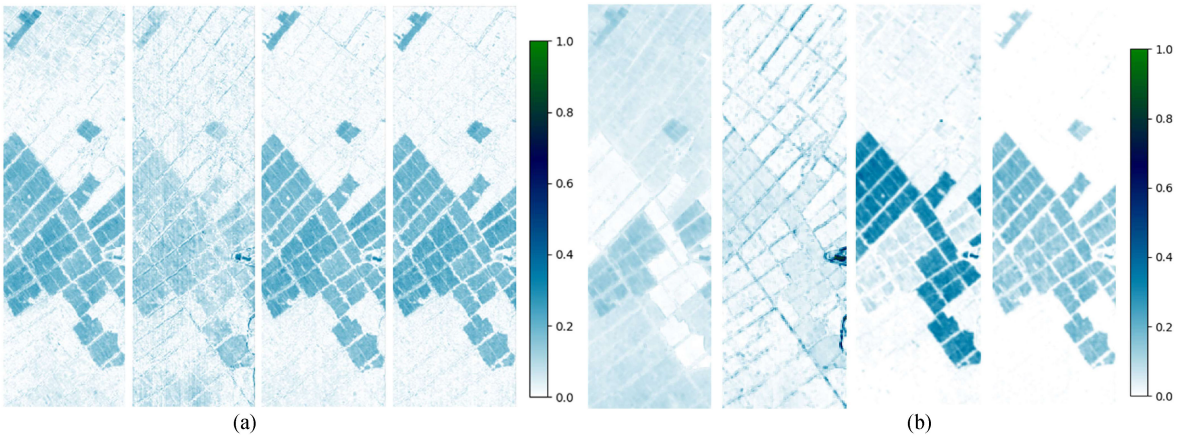


Fig. 6. *DIs* for each of four bands (7, 57, 117, 147) obtained by (a) change vector analysis and (b) structural similarity, respectively.

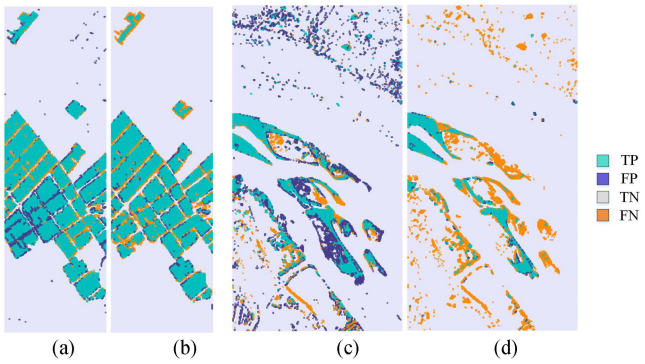


Fig. 7. Visualization of the change maps obtained by the two algorithms. The farmland change maps of (a) change vector analysis and (b) structural similarity. The river change maps of (c) change vector analysis and (d) structural similarity.

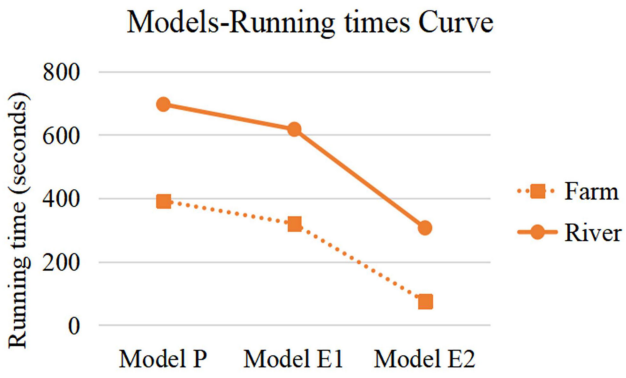


Fig. 8. Running times of three models P, E1 and E2 on (a) the farmland dataset (b) the river dataset, respectively.

To further demonstrate the performance of the model E2, extensive experiments and comparisons are implemented. The following contents including the results evaluation of the eight methods (five model-driven methods, the chosen model E2, and two data-driven methods), the change maps of eight methods, the visualization change maps of the five model-driven methods and the results evaluation of the model E2 based on different two model-driven methods, and results of different parameters in model E2.

It should be noted that the pseudotraining set still contains some error messages on the region of weak changes, which cannot be doubly confirmed as the correct changes by the two model-driven methods, CVA and SSIM. These weak changes with error messages cannot be fully learned by the used CNN. As a result, the higher accuracy and unstable results appeared when the training CNN does converge. Therefore, the results of the model with the best convergence are chosen for comparison.

Table I gives the results of eight methods on the farmland dataset and the River dataset. As seen, all the evaluation metrics on the river dataset are lower than those on the farmland dataset, which is related to the content characteristics of the dataset. The contents of the farmland dataset are crops and bare soil. Their change information is concentrated, and the change type is onefold. The river dataset mainly includes rivers, sand and gravel, buildings, and roads. The change information is fragmented, and change types are complex. Therefore, the characteristics of the River dataset are a challenge for all HSI-CD methods. The results of both data-driven methods are lower than the results of the CVA which is used to generate the pseudolabels. Such results validate that generating pseudolabels using a single model-driven method will face insurmountable accuracy limitations. In contrast, our method achieves the best results, which proves that the high-quality and comprehensive pseudolabels generated by the combination of CVA and SSIM can make a breakthrough in accuracy.

Fig. 10(a)–(e) shows the change maps of the Farmland dataset obtained just using the five model-driven methods. The change maps are diverse due to their different discriminative mechanisms. However, the change maps of CVA and SSIM algorithms also demonstrate the best completeness and differences. Both of them obtain the complete change maps, whereas the differences appear in the edge regions. While CVA detects each pixel pair independently, SSIM detects the block of pixels as a unit. Consequently, the results of CVA have some discrete points and many false negative samples in the edge region.

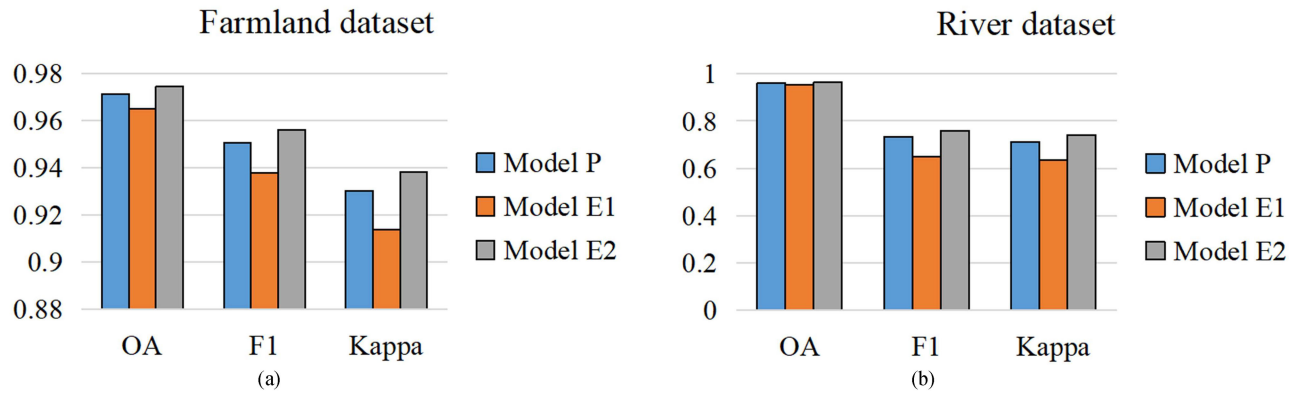


Fig. 9. Performance of three models P, E1 and E2 on (a) the farmland dataset and (b) the river dataset, respectively.

TABLE I
OA, F1, AND KAPPA OF EIGHT METHODS

Methods	Farmland Dataset	River Dataset
	OA / F1 / Kappa	OA / F1 / Kappa
ImageRatio	0.8874 / 0.8348 / 0.7522	0.9341 / 0.6527 / 0.6166
ImageRegr	0.8664 / 0.7229 / 0.6396	0.9401 / 0.6702 / 0.6373
DPCA	0.9495 / 0.9052 / 0.8712	0.9352 / 0.5706 / 0.5365
CVA	0.9689 / 0.9471 / 0.9250	0.9418 / 0.7165 / 0.6849
SSIM	0.9673 / 0.9423 / 0.9196	0.9555 / 0.6675 / 0.6459
FCN	0.9688 / 0.9470 / 0.9249	0.9416 / 0.7161 / 0.6844
Siamese network	0.9623 / 0.9371 / 0.9102	0.9508 / 0.6881 / 0.6617
Our model E2	**0.9746 / 0.9562 / 0.9383**	**0.9639 / 0.7592 / 0.7403**

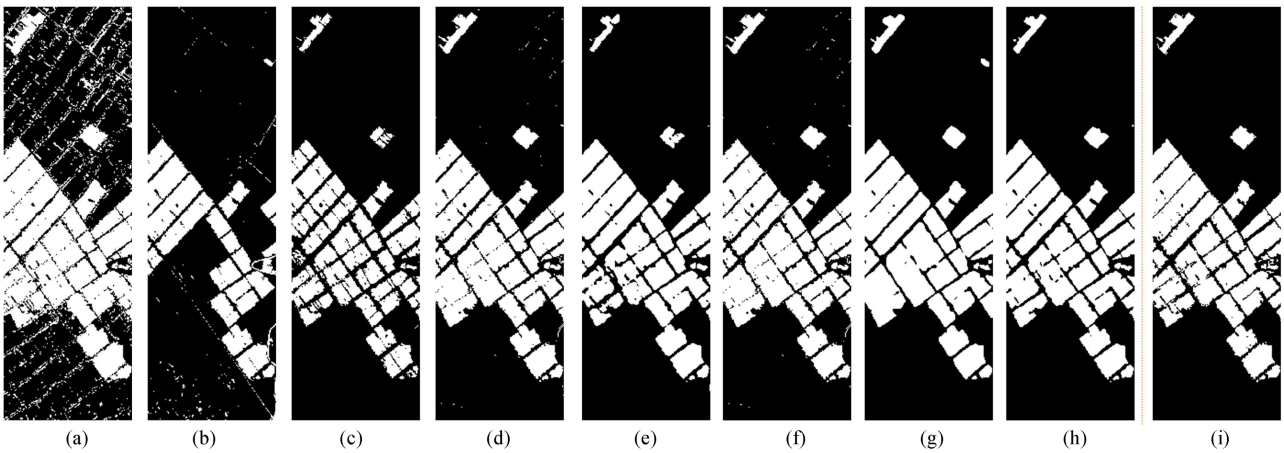


Fig. 10. CD results on the Farmland dataset using different methods. (a) ImageRatio. (b) ImageRegr. (c) Dynamic PCA. (d) Change vector analysis. (e) Structural similarity. (f) FCN. (g) Siamese network. (h) Our model E2. (i) Groundtruth.

The results of SSIM have no prominent discrete points. shows the change maps of the River dataset obtained using the five algorithms. Since the river dataset has complex change types and discrete change information, the change map of CVA has a large number of redundant white dots, and the SSIM lacks many white dots.

Figs. 10(f) and 11(f) show the change maps of the FCN method. It can be seen that the pseudo changes in the CVA are transmitted to the final change map. Figs. 10(g) and 11(g) show the change maps of the Siamese network method. A lot of details are lost in the change map, especially the edge areas. Suppression of false alarm rate leads to the increase of missed detection rate due to pseudo change information in the pseudolabel.

Figs. 10(h) and 11(h) show the change maps of the model E2. Both maps are close to the groundtruth change maps, respectively. The results can be considered as the neutralization of SSIM and CVA. Especially in the edge regions, which are neither as sharp as the CVA results nor as smooth as the SSIM

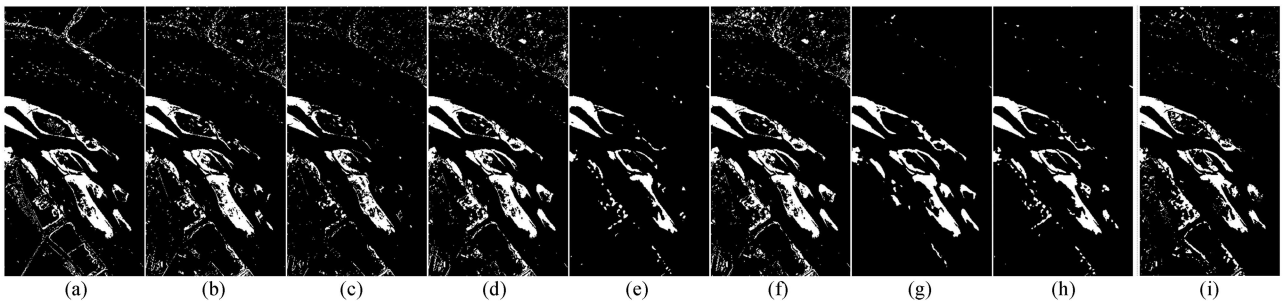


Fig. 11. CD results on the River dataset using different methods. (a) ImageRatio. (b) ImageRegr. (c) Dynamic PCA. (d) Change vector analysis. (e) Structural similarity. (f) FCN. (g) Siamese network. (h) Our model E2. (i) Groundtruth.

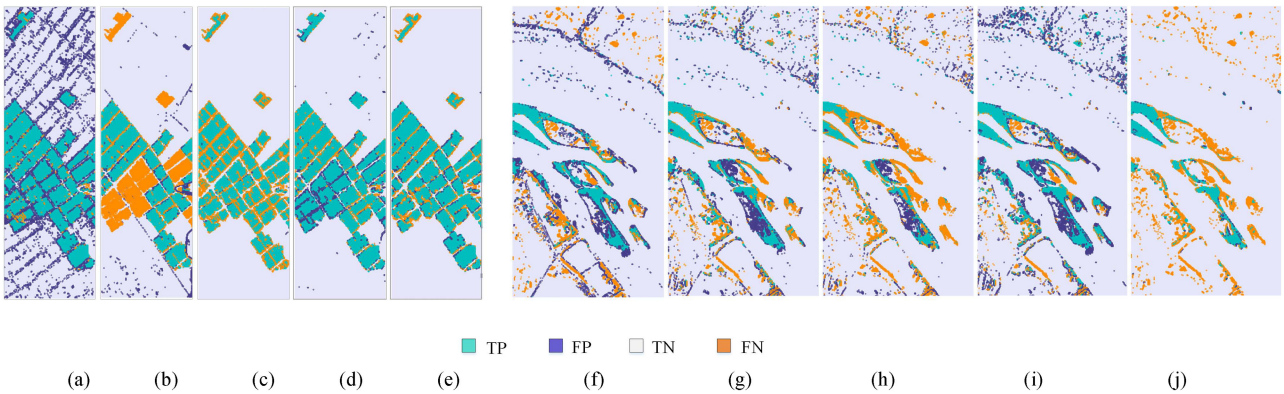


Fig. 12. Visualization of the Farmland dataset change maps obtained by (a) ImageRatio. (b) ImageRegr. (c) Dynamic PCA. (d) Change vector analysis. (e) Structural similarity. The visualization of the River dataset change maps obtained by (f) ImageRatio. (g) ImageRegr. (h) Dynamic PCA. (i) Change vector analysis. (j) Structural similarity.

TABLE II
OA, F1, AND KAPPA OF MODEL E2 BASED ON DIFFERENT COMBINATIONS OF TWO MODEL-DRIVEN METHODS

Methods	Farmland Dataset	River Dataset
	OA / F1 / Kappa	OA / F1 / Kappa
CVA | ImageRatio	0.9706 / 0.9504 / 0.9295	0.9603 / **0.7860** / **0.7643**
CVA | ImageRegr	0.9734 / 0.9534 / 0.9348	0.9563 / 0.7609 / 0.7369
CVA | DPCA	0.9706 / 0.9479 / 0.9276	0.9552 / 0.7227 / 0.6984
SSIM | ImageRatio	0.9742 / 0.9558 / 0.9376	0.9599 / 0.7152 / 0.6950
SSIM | ImageRegr	0.8951 / 0.7849 / 0.7189	0.9595 / 0.7174 / 0.6967
SSIM | DPCA	0.9673 / 0.9410 / 0.9184	0.9526 / 0.6396 / 0.6170
ImageRatio | ImageRegr	0.8930 / 0.7824 / 0.7145	0.9611 / 0.7698 / 0.7486
ImageRatio | DPCA	0.9722 / 0.9506 / 0.9113	0.9560 / 0.7061 / 0.6830
ImageRegr | DPCA	0.8771 / 0.7326 / 0.6697	0.9523 / 0.6915 / 0.6660
Our model E2	**0.9746** / **0.9562** / **0.9383**	**0.9639** / 0.7592 / 0.7403

results. The first reason is that the obtained pseudotraining set contains not only strong changes in the block regions, but also weak changes in the partial edge regions. The second reason is that CNN can not only accurately detect trained pixels, but also accurately detect edge region pixels that are not training samples.

The visualization change maps of the five model-driven methods are shown in Fig. 12. For the Farmland dataset, the change map of ImageRegr contains large areas of missed detection, does not satisfy the requirement of completeness. Therefore, ImageRegr is difficult to complement with other algorithms. For the River dataset, the change map of SSIM also contains many missing regions. Therefore, the complementarity of SSIM with other algorithms is weak.

Table II gives the results evaluation of the model E2 based on combining any two model-driven methods. Combining Fig. 12 and Table II reveals the impact of the complementarity of the two model-driven methods on the final change map. Different combinations of model-driven methods demonstrate different performances. Most of the combinations obtain more accurate change maps than the two model-driven change maps. The weakly complementary combinations obtain change maps with

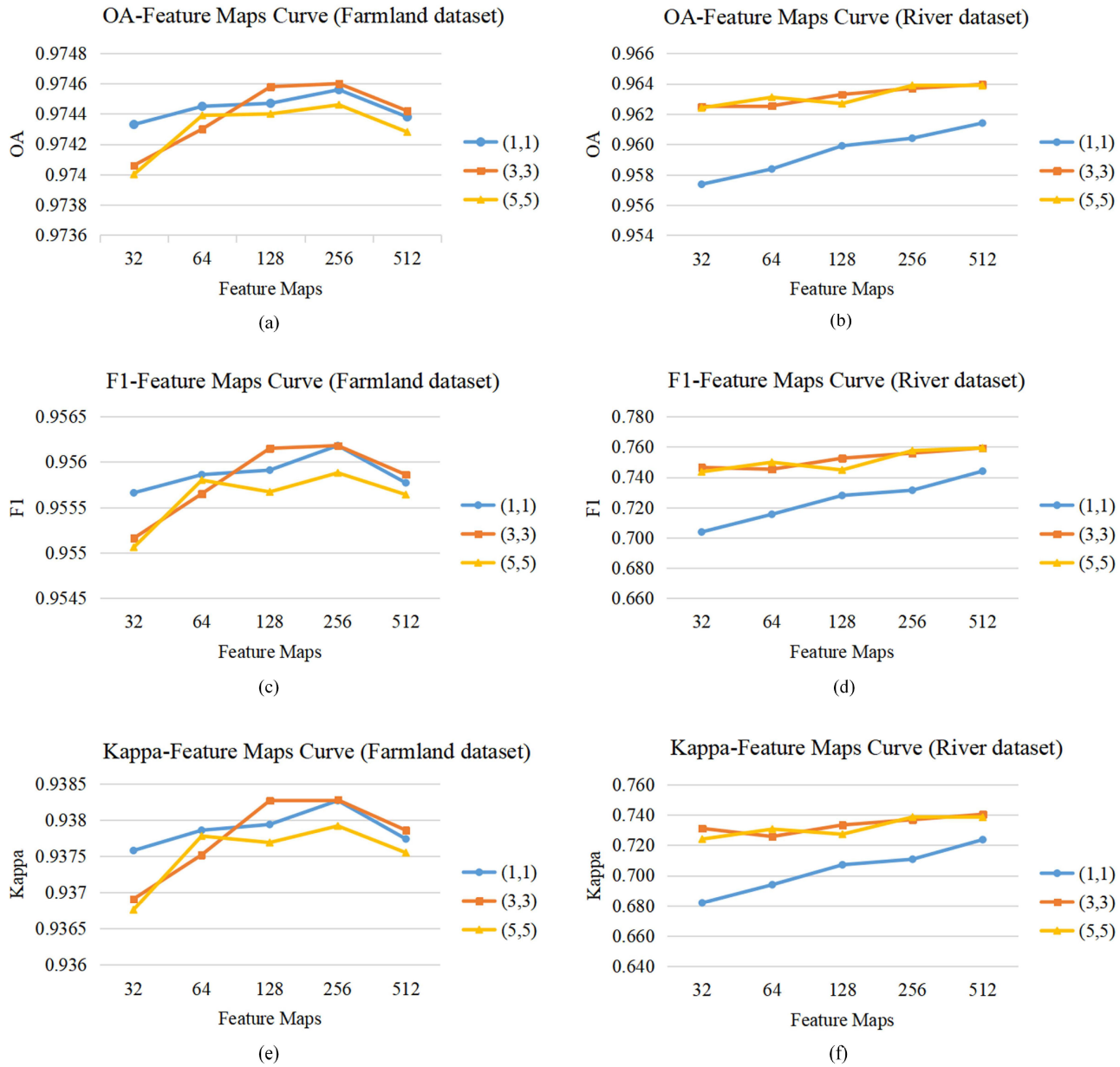


Fig. 13. Results based on the number and size of convolutional kernels. Farmland dataset of (a) overall accuracy, (c) F1, and (e) Kappa. River dataset of (b) overall accuracy, (d) F1, and (f) Kappa.

an accuracy between the two model-driven change maps. As seen, the model E2 combining CVA and SSIM provides the best performance. Importantly, for both datasets, the model E2 combing CVA and SSIM obtains the best results than just using any one of five model-driven methods. It can be concluded that only the combination of two model-driven methods with good complementarity can generate the complementary and credible pseudotraining set. Then the trained model can get the better detection capability over all regions of the image.

In addition, in model E2, the parameters of the single convolutional layer are crucial, including the number and size of convolutional kernels. Fig. 13 shows the relevant experimental results of different parameters. On each dataset, the convolution kernels of 256 and the convolution kernel size of (3, 3) are the best choices. The results of the model E2 in Table II, Figs. 10(h) and 11(h) are all based on these parameters.

V. Conclusion

Rather than simply propose a modified network based on the existed architectures, we focus on exploring a new paradigm to solve the task of HSI-CD. We propose to integrate the model-driven methods with deep learning methods in an end-to-end way. Especially, the *DIs* of the used two model-driven methods, CVA and SSIM, are the basis for generating the credible pseudotraining set. The proposed model E2 trained with the self-generated credible labels provides a new paradigm for the implementation of unsupervised HSI-CD. The lightweight of

the model E2 just using the single convolution layer ensures the timeliness of HSI-CD. In the future, we will explore the availability of the proposed method for CD tasks in heterogeneous bitemporal images, such as bitemporal synthetic aperture radar and optical images.

REFERENCES

[1] F. Wang and Y. J. Xu, "Comparison of remote sensing change detection techniques for assessing hurricane damage to forests," *Environ. Monit. Assessment,* vol. 162, no. 1-4, pp. 311–326, Mar. 2010.

[2] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to Tsunami-damage assessment," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 6, pp. 1658–1670, Jun. 2007.

[3] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the National Land Cover Database to Circa 2011," *Remote Sens. Environ.,* vol. 132, pp. 159–175, May 2013.

[4] C. N. Mundia and M. Aniya, "Analysis of land use/cover changes and urban expansion of Nairobi city using remote sensing and GIS," *Int. J. Remote Sens.,* vol. 26, no. 13, pp. 2831–2849, Jul. 2005.

[5] D. Wen, X. Huang, L. Zhang, and J. A. Benediktsson, "A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation," *IEEE Trans. Geosci. Remote Sens.,* vol. 54, no. 1, pp. 609–625, Jan. 2016.

[6] T. Hame, I. Heiler, and J. San Miguel-Ayanz, "An unsupervised change detection and recognition system for forestry," *Int. J. Remote Sens.,* vol. 19, no. 6, pp. 1079–1099, Apr. 1998.

[7] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and EO-1 hyperion for mineral mapping," *IEEE Trans. Geosci. Remote Sens.,* vol. 41, no. 6, pp. 1388–1400, Jun. 2003.

[8] 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.

[9] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.,* vol. 160, pp. 1–14, Apr. 2015.

[10] M. T. Eismann, J. Meola, and R. C. Hardie, "Hyperspectral change detection in the presence of diurnal and seasonal variations," *IEEE Trans. Geosci. Remote Sens.,* vol. 46, no. 1, pp. 237–249, Jan. 2008.

[11] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 4, pp. 2848–2864, Apr. 2020.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM,* vol. 60, no. 6, pp. 84–90, Jun. 2017.

[13] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.,* vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, 2016, pp. 770–778.

[15] 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. 2019.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.,* vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens,* vol. 12, no. 10, May 2020.

[18] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 38, no. 3, pp. 1171–1182, May 2000.

[19] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geosci. Remote Sens. Lett.,* vol. 6, no. 4, pp. 772–776, Oct. 2009.

[20] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion," *Remote Sens. Environ.,* vol. 199, pp. 241–255, Sep. 2017.

[21] A. Erturk, M. D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.,* vol. 9, no. 2, pp. 708–719, Feb. 2016.

[22] B. A. Singh, "Change detection in the tropical forest environment of northeastern India using Landsat remote sensing and tropical," 1996.

[23] R. D. Jackson, "Spectral indexes in N-space," *Remote Sens. Environ.,* vol. 13, no. 5, pp. 409–421, 1983.

[24] W. J. Todd, "Urban and regional land-use change detected by using landsat data," *J. Res. US Geol. Surv.,* vol. 5, no. 5, pp. 529–534, 1977.

[25] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with Landsat," in *Proc. 6th Annu. Symp. Mach. Process. Remotely Sensed Data,* pp. 326–336, 1980.

[26] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.,* vol. 50, no. 6, pp. 2196–2212, Jun. 2012.

[27] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 1, pp. 218–236, Jan. 2007.

[28] H. Zhang, M. Gong, P. Zhang, L. Su, and J. Shi, "Feature-level change detection using deep representation and feature change analysis for multispectral imagery," *IEEE Geosci. Remote Sens. Lett.,* vol. 13, no. 11, pp. 1666–1670, Nov. 2016.

[29] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.,* vol. 64, no. 1, pp. 1–19, Apr. 1998.

[30] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.,* vol. 16, no. 2, pp. 463–478, Feb. 2007.

[31] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comput.,* vol. 14, no. 4, pp. 715–770, 2002.

[32] L. Wiskott, P. Berkes, M. Franzius, H. Sprekeler, and N. Wilbert, "Slow feature analysis," *Scholarpedia*, vol. 6, no. 4, 2011.

[33] C. Wu, L. Zhang, and B. Du, "Kernel slow feature analysis for scene change detection," *IEEE Trans. Geosci. Remote Sens.,* vol. 55, no. 4, pp. 2367–2384, Apr. 2017.

[34] A. A. Abuelgasim, W. D. Ross, S. Gopal, and C. E. Woodcock, "Change detection using adaptive fuzzy neural networks: Environmental damage assessment after the Gulf war," *Remote Sens. Environ.,* vol. 70, no. 2, pp. 208–223, Nov. 1999.

[35] Z. Chen and B. Wang, "Spectrally-Spatially regularized low-rank and sparse decomposition: A novel method for change detection in multitemporal hyperspectral images," *Remote Sens.,* vol. 9, no. 10, pp. 1–21, 2017.

[36] C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogramm. Remote Sens.,* vol. 146, pp. 137–150, 2018.

[37] A. Erturk and A. Plaza, "Informative change detection by unmixing for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.,* vol. 12, no. 6, pp. 1252–1256, Jun. 2015.

[38] D. Qian, L. Wasson, and R. King, "Unsupervised linear unmixing for change detection in multitemporal airborne hyperspectral imagery," in *Proc. Int. Workshop Anal. Multi-Temporal Remote Sens. Images*, 2005.

[39] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[40] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo, and R. Gherardi, "Street-view change detection with deconvolutional networks," *Auton. Robots,* vol. 42, no. 7, pp. 1301–1322, Oct. 2018.

[41] D. Peng, Y. Zhang, and H. Guan, "End-to-End change detection for high resolution satellite images using improved UNet++," *Remote Sens.,* vol. 11, no. 11, Jun. 2019.

[42] J. Bromley, I. Guyon, Y. Lecun, E. Sckinger, and R. Shah, "Signature verification using a siamese time delay neural network," in *Proc. 7th Adv. Neural Inf. Process. Syst.*, 1993, pp. 669–688.

[43] C. Zhang *et al.*, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.,* vol. 166, pp. 183–200, Aug. 2020.

[44] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.,* vol. 11, no. 3, Feb. 2019.

[45] A. Song, J. Choi, Y. Han, and Y. Kim, "Change detection in hyperspectral images using recurrent 3D fully convolutional networks," *Remote Sens.,* vol. 10, no. 11, Nov. 2018.

[46] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.,* vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," May 2015.
[48] M. Folkman, J. Pearlman, L. Liao, and P. Jarecke, "EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration," *Proc. SPIE* vol. 4151, pp. 40–51, 2001.
[49] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.,* vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
[50] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, "Hyperion, a space-based imaging spectrometer," *IEEE Trans. Geosci. Remote Sens.,* vol. 41, no. 6, pp. 1160–1173, Jun. 2003.
[51] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing,* vol. 148, pp. 363–375, 2015.
[52] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.,* vol. 25, no. 12, pp. 2365–2407, Jun. 2004.
[53] 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, 2008.
[54] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 39, no. 4, pp. 640–651, Apr. 2017.

Qiuxia Li received the B.E. degree in physics from Yunnan University, Yunnan, China, in 2019. She is currently working toward the Ph.D. degree with the School of Physics, Xi'an Jiaotong University, Xi'an, China.

Her research interests include hyperspectral remote sensing image change detection based on deep learning methods.

Haishan Dai received the B.S. degree of optical information science and technology, in 2009, and the master's degree in physics, in 2012 from Xi'an JiaoTong University, Xi'an, China.

He is currently an Engineer with the Shanghai Institute of Satellite Engineering, China Aerospace Science and Technology Group Co., Ltd., Beijing, China. His research interests include satellite system engineering, satellite on-board data processing and intelligent algorithm for remote sensing data.

Chunlai Li received the B.S. degree in optoelectronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree in electronic science and technology from the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2009.

He is currently a Researcher with the Key Laboratory of Spatial Active Opto-Electronic Techniques, Shanghai Institute of Technical Physics, Chinese Academy of Sciences. His current research interests include weak infrared spectrum signal processing, refrigeration infrared focal plane information acquisition, and infrared hyperspectral data processing.

Zhiping He received the M.S. degree in optoelectronic engineering from the Zhejiang University, Zhejiang, China, in 2003, and the Ph.D. degree in Shanghai Institute of Technology Physics of the Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, in 2009.

He has long been engaged in the research of space optics and optoelectronic technology, and he devotes himself to the research of key technologies involved in space spectral technology and active-passive composite optical systems. The instrument developed by his team has been successfully applied to the Chang'E 3, Chang'E 4, Chang'E 5, and Mars exploration projects in China.

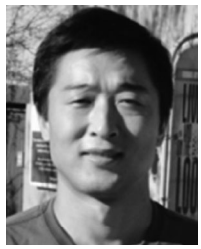
Wenjing Wang, photograph and biography not available at the time of publication.

Yusen Feng, photograph and biography not available at the time of publication.

Feng Han, photograph and biography not available at the time of publication.

Abudusalamu Tuniyazi, photograph and biography not available at the time of publication.

Haoyang Li, photograph and biography not available at the time of publication.



Tingkui Mu received the Ph.D. degree in physics from Xi'an Jiaotong University, Xi'an, China, in 2012.

He was a Postdoctoral Fellow with the College of Optical Sciences, University of Arizona, Tucson, AZ, USA, between 2014 and 2016. He is currently an Associate Professor of the School of Physics, Xi'an Jiaotong University. He is the Director of Shaanxi Provincial Optical Society, the committee member of the Optoelectronics Technology Committee of the Chinese Aerospace Society, the member of the Chinese Optical Society, Chinese Optical Engineering Society, the Shaanxi Provincial Physical Society, the Optical Society of America (OSA) and the Society of Photo-Optical Instrumentation Engineers (SPIE). He has authored or coauthored more than 50 research papers and 16 authorized invention patents. His research interests include integrated acquisition and processing of multi-dimensional optical information, computer vision, spectral and polarimetric imaging technique and sensors.