GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection

Qi Wang, *Senior Member, IEEE*, Zhenghang Yuan, Qian Du, *Fellow, IEEE*, and Xuelong Li, *Fellow, IEEE*

***Abstract*— Change detection (CD) is an important application of remote sensing, which provides timely change information about large-scale Earth surface. With the emergence of hyperspectral imagery, CD technology has been greatly promoted, as hyperspectral data with high spectral resolution are capable of detecting finer changes than using the traditional multispectral imagery. Nevertheless, the high dimension of the hyperspectral data makes it difficult to implement traditional CD algorithms. Besides, endmember abundance information at subpixel level is often not fully utilized. In order to better handle high-dimension problem and explore abundance information, this paper presents a general end-to-end 2-D convolutional neural network (CNN) framework for hyperspectral image CD (HSI-CD). The main contributions of this paper are threefold: 1) mixed-affinity matrix that integrates subpixel representation is introduced to mine more cross-channel gradient features and fuse multisource information; 2) 2-D CNN is designed to learn the discriminative features effectively from the multisource data at a higher level and enhance the generalization ability of the proposed CD algorithm; and 3) the new HSI-CD data set is designed for objective comparison of different methods. Experimental results on real hyperspectral data sets demonstrate that the proposed method outperforms most of the state of the arts.**

***Index Terms*— 2-D convolutional neural network (CNN), change detection (CD), deep learning, hyperspectral image (HSI), mixed-affinity matrix, spectral unmixing.**

Manuscript received December 11, 2017; revised May 7, 2018; accepted June 18, 2018. Date of publication July 24, 2018; date of current version December 24, 2018. This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1002202, in part by the National Natural Science Foundation of China under Grant 61773316, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, in part by the Fundamental Research Funds for the Central Universities under Grant 3102017AX010, and in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences. *(Corresponding author: Qi Wang.)*

Q. Wang is with the School of Computer Science, the Center for OPTical IMagery Analysis and Learning, and the Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@gmail.com).

Z. Yuan is with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: zhenghangyuan@mail.nwpu.edu.cn).

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

X. Li is with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: xuelongli@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2849692

I. Introduction

THE comprehensive cognition of the global change is a critical task for providing timely and accurate information in the Earth's evolution. Change detection (CD), as an important application of identifying differences of multitemporal remote sensing images, finds great use in land-cover mapping [1], natural disaster monitoring [2], resource exploration [3], and so forth. The complete CD process is mainly divided into three steps: 1) image preprocessing, where the multitemporal images are spatially and radiometrically processed with geometric correction, radiometric correction, and noise reduction; 2) change difference image generation, where the change difference images are obtained to contrast changed and unchanged regions of multitemporal images; and 3) evaluation, where a certain measure is used to assess the performance.

Hyperspectral image (HSI) with high spectral resolution can provide richer spectral information than other remote sensing images, such as synthetic aperture radar images [4] and multispectral images (MSIs) [5]. Thus, HSIs have the potential to identify finer changes [6], with the detailed composition of different objects being reflected. Though the HSI-CD methods have been developed, it does not mean that the HSI-CD problem can be readily solved, since CD is a complicated process and can be affected by many factors. The main challenges for HSI-CD are summarized as follows.

A. Mixed Pixel Problem

In actual HSI-CD, mixed pixels commonly exist. A mixed pixel is a mixture of more than one distinct substance [7]. If one mixed pixel is roughly categorized into a certain kind of substance, it will not be expressed accurately. In general, mixed pixels may limit the improvement of a hyperspectral task if not being carefully handled.

B. High Dimensionality

The high dimension of the hyperspectral data makes it difficult to handle in some CD algorithms. In spite of the fact that feature extraction or band selection [8] is utilized to reduce the dimensionality, the detailed information may be lost to some extent.

C. Limited Data Sets

The data sets for HSI-CD are relatively limited since constructing a ground-truth map that reflects real change information of ground objects requires considerable time and effort. Both field work and manual labeling are time-consuming and expensive.

Aiming at the aforementioned three challenges, our motivations are described in detail from two aspects. First, an in-depth analysis of internal pixel is necessary, but many related studies are case studies on limited problems about HSI-CD. Spectral unmixing is a process to decompose a mixed pixel into a collection of endmembers and the corresponding abundance maps. An abundance map [7] indicates the proportion of a corresponding endmember present in the pixel, containing much detailed subpixel composition. By considering the useful information, unmixing in HSI-CD not only enhances the performance [9], but also provides easily interpretable information for the nature of changes [10]. Though spectral unmixing has been widely studied, research on HSI-CD by unmixing is still in an early stage. Existing studies in the literature are case studies on certain problems and mainly focus on linear or nonlinear unmixing without combining them in HSI-CD.

Second, deep learning can better handle the problem of high dimension and exploit effective features, but the features are not fully mined in many HSI-CD methods. By learning which part or what kind of combination of multisource data is critical, deep learning offers flexible approaches to deal with HSI. Generally, a change difference image can be obtained from the two preprocessed images with the trained deep neural network. However, many HSI-CD methods based on deep learning do not fully exploit the feature contained between the spectra of corresponding pixels in multitemporal images. These methods mainly analyze the changes of pixels by 1-D spectral vector. The features between the spectra of corresponding pixels contain richer information, which is often neglected.

Based on the aforementioned two motivations, this paper develops a new general framework, namely general end-to-end 2-D convolutional neural network (CNN) (GETNET), for HSI-CD task. The main contributions are summarized as follows.

1) A simple yet effective mixed-affinity matrix is proposed to better mine the changed patterns between two corresponding spectral vectors on a spatial pixel. Different from the previous methods, mixed-affinity matrix converts two 1-D pixel vectors to a 2-D matrix for CD task, which provides more abundant cross-channel gradient information. Moreover, mixed-affinity matrix can effectively deal with the multisource data simultaneously and make it possible to learn representative features between the spectra in the GETNET.
2) A novel end-to-end framework named GETNET based on a 2-D CNN is developed to improve the generalization ability of the HSI-CD algorithm. To the best of our knowledge, GETNET is the first one to fuse HSI and abundance maps together to handle the changes caused by different complicated factors. The proposed approach outperforms (or is competitive to) other methods across all the four data sets without parameter adjustment. Moreover, this approach offers more potential than other methods on some special data sets given that unmixing information is properly used.
3) A new HSI-CD data set named "river" is designed for objective comparison of different methods. The changed information on this data set is relatively complicated, and the experimental results show that this data set we create is reliable and robust for the quantitative assessment of CD task.

The remainder of this paper is organized as follows. Section II reviews the related work. Section III describes the detail of the proposed method. Section IV evaluates the performances of different CD approaches. Finally, we conclude this paper in Section V.

II. Related Work

The most common and simple methods of CD are image differencing and image ratioing, mainly designed for single-band images. These two methods are easy to implement but their applicable scope is limited. When handling multiband remote sensing images, classical CD methods fall primarily into four categories.

A. Image Arithmetical Operation

The most common approach is change vector analysis (CVA) [11], which generates magnitude and direction of change by spectral vector subtraction. Besides, there are many comparative analyses based on CVA, such as improved CVA, modified CVA, and CVA posterior-probability space [12]. Typically, based on the polar CVA, C^2VA [13] describes the CD problem in a magnitude-direction 2-D representation produced by a loss compression procedure. In addition, Chen *et al.* [14] apply tensor algebra to CD and propose 4-D higher order singular value decomposition to capture comprehensive changed features.

B. Image Transformation

Methods based on image transformation convert MSIs into a specific feature space to emphasize changed pixels and suppress unchanged ones. As a linear transformation technique, principal component analysis (PCA) [15] decorrelates images [16]. However, PCA heavily relies on the statistical property of images and is especially vulnerable to unbalanced data [17]. Based on the canonical correlation analysis, multivariate alteration detection (MAD) [18] is an unsupervised method to analyze multitemporal images CD. Nevertheless, MAD is not applicable to affine transformation. Then, iteratively reweighted MAD (IR-MAD) [19] extends MAD by applying different weights to observations. However, IR-MAD still cannot utilize the notable relationship between multitemporal bands. The subspace-based CD method [20] takes the observed pixel of one HSI as target and establishes the background subspace by utilizing the corresponding pixel in the other image.

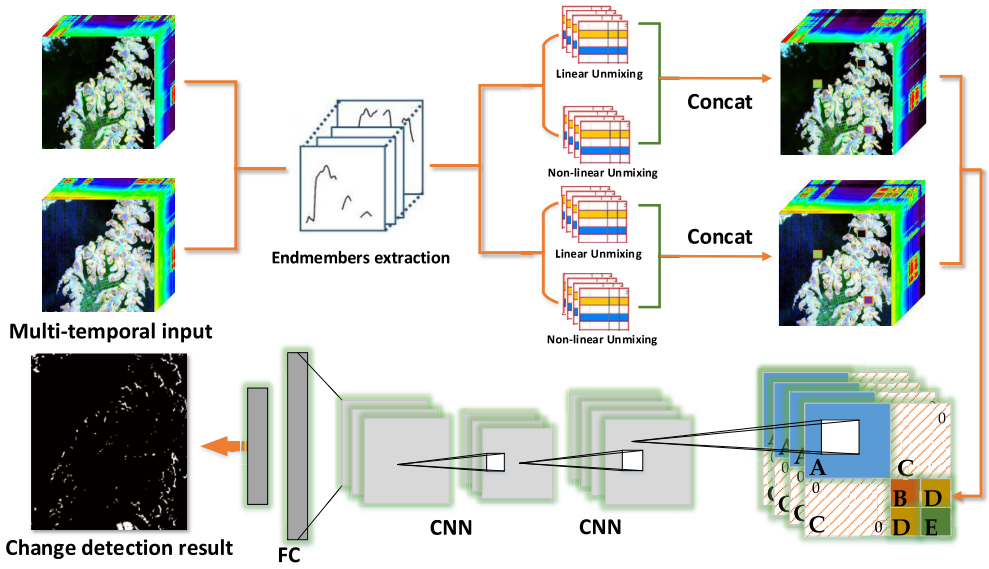

Fig. 1. Overview of the GETNET CD method for multitemporal HSIs.

C. Image Classification

These methods are supervised schemes and have prior knowledge for the training of a classifier. Among them, postclassification comparison [13] takes the CD as an image pair classification task, where the two classified images are compared pixel by pixel. The pixels that fall into different categories are considered as changed regions. This is also the case of classification of differential features [21] and compound classification [1].

D. Other Advanced Methods

Research on deep learning has found significant use in CD of multiband images. For example, Zhang *et al.* [15] develop a deep belief network-based feature extraction method, which concatenates a pair of multitemporal vectors to output the deep representation used for CD. By reshaping pixels to vectors, Zhao *et al.* [22] present the restricted Boltzmann machine-based deep learning methods to learn representation for image analysis. In addition, more new approaches have been put forward to solve the problems of HSI-CD in the recent studies. Liu *et al.* [23] describe a hierarchical CD approach, aiming at distinguishing all the possible change types by considering discriminable spectral behaviors. Based on the statistical analyses of spectral profiles, Han *et al.* [24] adopt an unsupervised CD algorithm for denoising without dimensionality reduction.

The aforementioned four categories can be applied to HSIs, but the accuracy may be improved if mixed pixels are considered to handle pixel-level change. In order to address mixed pixels in HSIs, unmixing is necessary. HSI-CD by unmixing has obvious advantages of providing subpixel-level information. However, unmixing in HSI-CD is very challenging and has not been extensively studied so far. The work in [25] obtains the change difference image by calculating changes between the corresponding abundance maps of multitemporal images. In addition to spectral information, superpixels are utilized to integrate spatial information in the unmixing process [26]. These two approaches detect changes in the abundance of multitemporal images for each endmember simply by difference operation, limited to certain problems. Liu *et al.* [27] analyzes abundance maps of changed and unchanged endmembers and their contribution to each pixel. Besides, sparse unmixing is first explored for HSI-CD by Ertürk *et al.* [10] with spectral libraries. Many studies about unmixing in change analysis are limited to case studies and need to be further explored.

III. Methodology

The overview of the proposed CD method for multitemporal HSIs is shown in Fig. 1. First, spectral unmixing algorithms are used to deal with two preprocessed HSIs. Then, the abundance maps obtained by linear and nonlinear unmixing interact with the HSIs and form mixed-affinity matrices. Next mixed-affinity matrices are processed by the GETNET. In the end, the final CD result is the output of the deep neural network.

In this section, we introduce our method from three aspects: hybrid unmixing for subpixel-level mining, mixed-affinity matrix generation for information fusion, and GETNET with "separation and combination."

A. Hybrid-Unmixing for Subpixel-Level Mining

In order to obtain subpixel-level information and improve generalization ability, spectral unmixing is utilized in the proposed method. A complete spectral unmixing process contains endmember extraction and abundance estimation. In previous studies, various endmember extraction algorithms (EEAs) are developed, and any EEA is applicable to the proposed method. In this paper, an automatic target generation process (ATGP) [28] is adopted to find potential target endmembers. After endmembers are obtained, the next step is to achieve abundance estimation, namely, abundance maps for each pixel.

Aiming at achieving abundance estimation, mixture models can be divided into two categories: linear and nonlinear. Either has advantages and is applicable to certain cases [29], [30]. Nevertheless, few works pay attention to combining them in HSI-CD. In order to utilize both the advantages, this paper focuses on hybrid-unmixing, namely, not only linear but also the nonlinear mixture model.

The linear mixture model assumes that each pixel is a linear combination of endmembers and the reflectance $\boldsymbol{r}$ of one pixel can be expressed as

$$\boldsymbol{r} = \sum_{i=1}^{m} w_{l_i} \boldsymbol{x_i} + \varepsilon = \boldsymbol{X} \boldsymbol{w_l} + \varepsilon \tag{1}$$

$$\textbf{s.t.} \sum_{i=1}^{m} w_{l_i} = 1 \tag{2}$$

$$0 \le w_{l_i} \le 1, \quad i = 1, 2, 3, \ldots, m \tag{3}$$

where $\boldsymbol{r}$ represents a $b \times 1$ vector, b is the number of HSI bands, m is the number of endmembers, $\boldsymbol{X}$ is a $b \times m$ matrix of endmembers, $\boldsymbol{x_i}$ is the ith column of $\boldsymbol{X}$, $\boldsymbol{w_l}$ is an $m \times 1$ vector of linear abundance fraction, w_{l_i} denotes the linear proportion of the ith endmember, ε is a $b \times 1$ vector of noise caused by the sensor and modeling errors. Equations (2) and (3) are two constraints of abundances nonnegative constraint and abundance sum-to-one constraint, respectively. To estimate the parameters, the fully constrained least squares (FCLS) [31] method is adopted. The linear abundance map $\boldsymbol{w_l}$ is the key information used in the proposed method.

The nonlinear mixture model can be formulated as

$$\boldsymbol{r} = \sum_{i=1}^{m} w_{ni} \boldsymbol{x_i} + \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij} \boldsymbol{x_i} \odot \boldsymbol{x_j} + \varepsilon \tag{4}$$

$$\textbf{s.t.} \ \forall i \ge j : a_{ij} = 0 \tag{5}$$

$$\forall i < j : a_{ij} = w_{ni} w_{nj} \tag{6}$$

$$\sum_{i=1}^{m} w_{ni} = 1, \quad w_{ni} \ge 0 \tag{7}$$

where $\odot$ is the pointwise multiplication operator, a_{ij} represents the bilinear parameter to model the nonlinearity of HSIs, w_{n_i} denotes the nonlinear proportion of the ith endmember, and other parameters have the same meaning in (1). In this paper, the bilinear-Fan model (BFM) [30] is employed to obtain the parameters, and the nonlinear abundance map $\boldsymbol{w_n}$ is achieved. It is worth noting that both inputs in one data set share the same endmembers. Specifically, all the endmembers used in FCLS and BFM algorithms are the same and generated by the ATGP algorithm. By applying simple and effective methods, subpixel-level information is explored fully to the next step. As linear and nonlinear mixture models are appropriate for different cases, combining both methods in a unified framework can improve the generalization ability.

The size of HSI is $h \times w \times b$. For convenience, we stack all $\boldsymbol{w_l}$ (the number of $\boldsymbol{w_l}$ is $h \times w$) into a 3-D abundance data cube with the size of $h \times w \times m$. $\boldsymbol{w_n}$ is processed on the same way of $\boldsymbol{w_l}$. Then, the next step is to integrate the two types of abundance maps together with the hyperspectral data. In other words, three kinds of data are stacked neatly in sequence in the direction of HSI spectral domain, as shown in Fig. 2. The final multisource data cube, with the size of $w \times h \times (b + 2m)$, contains almost all the HSI information for the subsequent analysis.

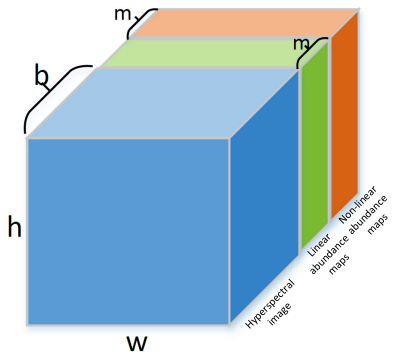

Fig. 2. Mixed data cube of HSI and unmixing abundance maps.

B. Mixed-Affinity Matrix Generation for Information Fusion

Different from the previous deep learning methods, this paper proposes a novel mixed-affinity matrix for the 2-D CNN algorithm. Assume that there are b bands to represent one pixel in HSI. Moreover, there are m different endmembers and each endmember has a complete linear and nonlinear abundance maps (the size of complete abundance map is $h \times w$). As described in Section III-A, both pixel- and subpixel-level information of the integrated multisource data cube is obtained. Hence, the similarity between different bands of the corresponding pixels is a kind of signature to describe this pixel, i.e., a certain band in one pixel interacts with other b bands in the corresponding pixel, respectively. Meanwhile, the similarity between different abundance maps is also another signature to indicate subpixel representation. Then, we define mixed-affinity matrix, which contains both pixel- and subpixel-level signatures to represent each pixel. The mixed-affinity matrix is constructed as

$$K_{ij} = 1 - (r_{1_{ij}} - r_{2_{ij}})/r_{2_{ij}}, \quad i, j = 1, 2, 3, \ldots, N \tag{8}$$

where K_{ij} is a 2-D mixed-affinity matrix of $n \times n$ and it represents the relationship between corresponding pixels of different wavelengths and abundance maps; $N = h \times w$; $n = b + 2m$; $r_{1_{ij}}$ is a $n \times 1$ vector of a pixel at time$_1$ in Fig. 2; and r_2 is another $n \times 1$ vector of the corresponding pixel at time$_2$. Mixed-affinity matrix is a 2-D matrix of $n \times n$ and it represents the relationship between corresponding pixels of different wavelengths and abundance maps, i.e., mixed affinity. As shown in Fig. 3, mixed-affinity matrix is divided into five different parts.

In part A, a certain band at time$_1$ in one pixel subtracts from other b bands in the corresponding pixel at time$_2$. Part A represents the pixel-level difference information. In part B, linear abundance map of a certain endmember for one pixel at time$_1$ subtracts from the corresponding linear abundance map at time$_2$. Similarly, in part E, nonlinear abundance map interacts with each other by subtraction. In part D, linear

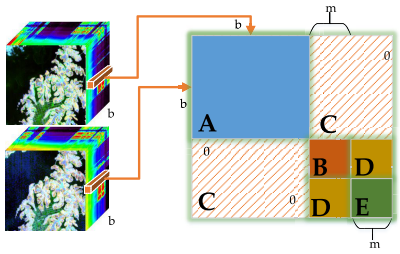

Fig. 3. Mixed-affinity matrix of one pixel.

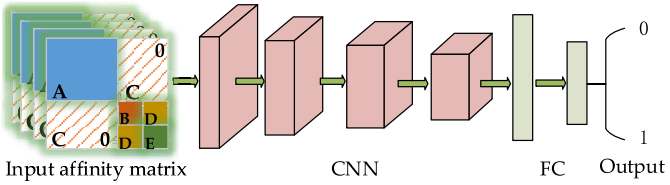

Fig. 4. Architecture of the proposed CNN with mixed-affinity matrices for CD.

abundance map subtracts from the corresponding nonlinear abundance map. Parts B, D, and E reveal subpixel-level difference information. As the affinity between hyperspectral data and abundance maps is meaningless, each K_{ij} in part C is set to 0. In general, the image spectral affinity and the abundance affinity are distributed in the top-left and bottom-right corner, respectively, in Fig. 3.

The greater value of K_{ij} is, the more similar it is between the corresponding pixels. After the mixed-affinity matrices are calculated, they are input to the proposed 2-D CNN. The architecture takes the CD as a binary classification task and outputs the classification result, as shown in Fig. 4. Remarkably, the similarity and difference of the corresponding pixels can be seen from the value of mixed-affinity matrix. Mixed-affinity matrix has important significance in three aspects as follows.

1) Mixed-affinity matrix is an efficient way for simultaneous processing of multisource information fusion, combining preprocessed hyperspectral data with linear and nonlinear abundance maps. In addition, pixel- and subpixel-level representation is achieved naturally.
2) The differences between the two 1-D pixel vectors are mapped to a 2-D matrix, which provides more abundant cross-channel gradient information. By mining the differences among b channels and $2m$ complete abundance maps, the utilization ratio of multisource information is maximized.
3) The powerful learning ability of the 2-D CNN can work seamlessly with the mixed-affinity matrix. On the one hand, deep learning is able to learn the significant features for the CD task. On the other hand, it can integrate information with local context according to different parts in the mixed-affinity matrix.

C. GETNET With "Separation and Combination"

CD can be seen as the binary classification of changed and unchanged pixels. The proposed classification model on the mixed-affinity matrix is different from a traditional CNN classification task. In a common image classification task, better performance is usually obtained by a deeper CNN model. This model commonly exploits residual layers or inception modules to learn powerful representations of images. Considering the specific CD task, we design a GETNET architecture for the vector pair classification problem. The end-to-end architecture can help ensure high accuracy of the whole system by optimizing it at all layers and generate the change difference image without stagewise optimization or error accumulation problem. Then, the designed network is introduced as follows.

1) "Separation and Combination": The proposed algorithm involves convolution operators on the mixed-affinity matrices. The structure of the mixed-affinity matrix is shown in Fig. 3, where the image spectral affinity and the abundance affinity are distributed in the top-left and bottom-right corner, respectively. Traditional convolution operation shares kernel weights on the whole input map. However, it is not suitable for the mixed-affinity matrix at the early convolution layer, since spectral feature has a different nature compared with the abundance map. In this paper, two different convolution kernels are operated on the mixed-affinity matrix. One is for the image spectral affinity in the top left, i.e., part A, and the other is for the abundance affinity in the bottom-right corner, i.e., parts B, D, and E. Because the value of part C is 0, any convolution kernel is applicable here. Then, we adopt the locally sharing weights convolution [32] for these parts.

After locally sharing convolution layers and max pooling layers on the mixed-affinity matrices, two kinds of higher level features of different parts, i.e., spectral feature and abundance feature, are learned automatically by the proposed network. The next step is to classify these features by the fully connected layers, which work as a classifier. Additionally, considering the truth that different data sets are sensitive to different types of features, the fully connected layers are employed to fuse these features to improve the generalization ability.

2) Training With Predetection: The overall architecture of the proposed network is shown in Fig. 4, and the detailed parameters are presented in Table I, where the "LS" stands for locally sharing weights convolution [32]. In this paper, HSI-CD is viewed as a binary classification task of changed and unchanged pixels. Mixed-affinity matrices are input into the 2-D CNN, and the outputs of the network are the class labels of pixels. A change difference image can be obtained from the two HSIs with the trained network. However, the ground-truth map is used to evaluate the different CD methods but not for training in the HSI analysis. Thus, we need to generate pseudodata sets to train it. Any HSI-CD method can be used to generate labeled samples. This paper adopts the conventional method CVA with high confidence threshold to generate pseudotraining sets with labels. Pixels which have high possibility of being classified correctly are selected as final labeled samples to train the network. The positive samples (changed samples) account for about 10% of the changed pixels. The number of negative samples (unchanged samples) is twice of the positive samples, namely that the ratio of positive to negative samples is 1:2.

TABLE I
ARCHITECTURE DETAILS FOR GETNET

Layers	Type	Channels	Kernel Size
LSConv1	LSConvolution + BN Activation(tanh)	32	5 x 5
MaxPool1	MaxPooling	-	2 x 2
LSConv2	LSConvolution + BN Activation(tanh)	64	3 x 3
MaxPool2	MaxPooling	-	2 x 2
LSConv3	LSConvolution + BN Activation(tanh)	128	3 x 3
MaxPool3	MaxPooling	-	2 x 2
LSConv4	LSConvolution + BN Activation(tanh)	96	1 x 1
MaxPool4	MaxPooling	-	2 x 2
FC1	Fully Connected + BN Activation(tanh)	512	-
FC2	Fully Connected	2	-

The number of training samples has a significant impact on the performance of classification. If the positive and negative training samples are unbalanced, the network may be biased to the class with more samples. To avoid this, the ratio between the positive and negative samples is set to 1:1 at first. However, more training samples are desired to train a deep network without pretrained models. According to the predetection method (CVA) output, the ratio between the changed and unchanged samples is usually far less than 1:1. Thus, more unchanged samples are needed to train the network. Then, the ratio of positive to negative samples is set to 1:2 in this model.

IV. EXPERIMENTS

In this section, extensive experiments are conducted using four real HSI data sets. First, we introduce the existing and newly constructed data sets used in the experiments. Then, the evaluation measures are described for the HSI-CD methods. In the end, the experimental performances of the proposed method and other state of the arts are analyzed in detail.

A. Data Sets Description

The most important reason for the lack of HSI-CD data sets is that it is difficult to construct a ground-truth map. Ground-truth map is a change difference image reflecting the real change information of ground objects, composed of changed and unchanged classes. Each data set contains three images, two real HSIs which are taken at the same position from different times, and a ground-truth map. Generally, the structure of the ground-truth map mainly depends on the on-the-spot investigation and some algorithms with high accuracy. For uncertain pixels, it is necessary to query their spectral values for further judgment. Constructing precise ground-truth map is vital for its quality that directly affects the accuracy assessment.

In this paper, all the HSI data sets used in the experiments are selected from Earth Observing-1 (EO-1) Hyperion images. The EO-1 Hyperion covers the 0.4–2.5-μm spectral range with 242 spectral bands. Moreover, it provides a spectral resolution of 10 nm approximately, as well as a spatial resolution of 30 m.

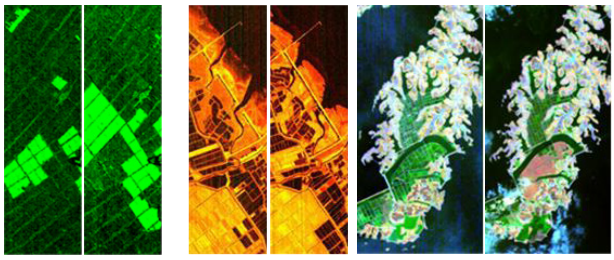

Fig. 5. Existing experimental data sets in this paper. (a) Farmland imagery on May 3, 2006. (b) Farmland imagery on April 23, 2007. (c) Countryside imagery on November 3, 2007. (d) Countryside imagery on November 28, 2008. (e) Poyang lake imagery on July 27, 2002. (f) Poyang lake imagery on July 16, 2004.

Though there are 242 spectral bands in Hyperion, the HSI quality can be harshly affected by the atmosphere. Typically, some of spectral bands are noisy with low signal-to-noise ratio (SNR). Therefore, according to the existing research and the environment for visualizing images analysis, bands with high SNR are selected in the following analysis. The experiments are carried out on the data sets after noise elimination and the ground-truth maps of the existing data sets come from [33].

1) Existing Data Sets: There are some available hyperspectral data sets for HSI-CD methods. The first data set "farmland" is shown in Fig. 5(a) and (b). It covers farmland near the city of Yancheng, Jiangsu province, China, with a size of 450 $\times$ 140 pixels. The two HSIs were acquired on May 3, 2006 and April 23, 2007, respectively. There are 155 bands selected for CD after noise elimination. Visually, the main change on this data is the size of farmland. In this experiment, about 20.95% pixels of the whole image are selected as labeled samples with 4400 positive samples and 8800 negative samples. There are more samples selected than other data sets because the proportion of changed pixels is higher.

The second data set "countryside" is shown in Fig. 5(c) and (d), which covers countryside near the city of Nantong, Jiangsu province, China, having a size of 633 $\times$ 201 pixels. The multitemporal HSIs were obtained on November 3, 2007 and November 28, 2008, respectively; 166 bands are selected for CD after noise elimination. Visually, the main change on this data is the size of rural areas. In this experiment, about 7.55% pixels are selected as labeled samples with 3200 positive samples and 6400 negative samples.

The third data set "Poyang lake" is shown in Fig. 5(e) and (f). It covers the province of Jiangxi, China, with a size of 394 $\times$ 200 pixels. The two HSIs were obtained on July 27, 2002 and July 16, 2004, respectively; 158 bands are available for CD after noise elimination. Visually, the main change type on this data is the land change. In this experiment, about 2.17% pixels are selected, where 570 positive samples and 1140 negative samples are taken as a labeled data.

2) Newly Constructed Data Set: The newly constructed data set "river"[1] is shown in Fig. 6. The two HSIs were acquired at

[1]This dataset can be downloaded from the following website. http://crabwq.github.io

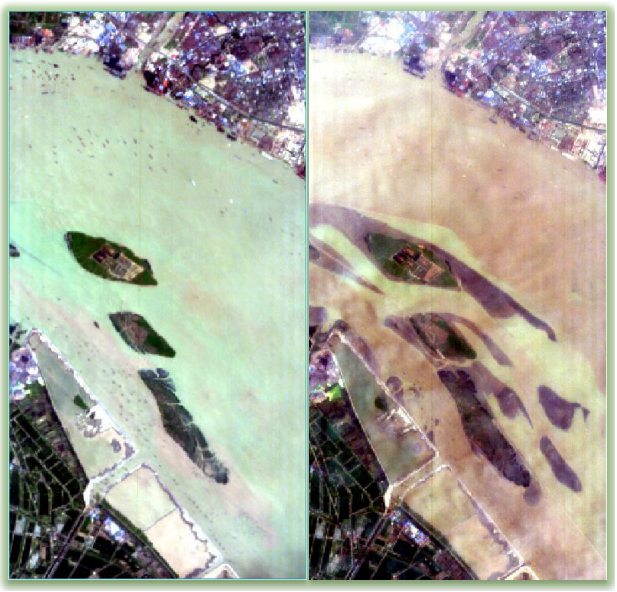

Fig. 6. Newly constructed data set of HSIs at two different times. (a) River imagery on May 3, 2013. (b) River imagery on December 31, 2013.

May 3, 2013 and December 31, 2013, respectively, in Jiangsu province, China. This data set has a size of 463×241 pixels with 198 bands available after noisy band removal. The main change type on this data is disappearance of substance in river. In this experiment, about 3.37% pixels are selected as labeled samples with 1250 positive samples and 2500 negative samples.

B. Evaluation Measures

In order to evaluate the performance of the CD methods, we compare the change difference images with ground-truth maps, in which white pixels represent changed portions and black pixels mean unchanged parts. The accuracy of result can directly reflect the reliability and practicability of the CD methods. Generally, through pixel-level evaluation, this paper adopts three evaluation criteria: overall accuracy (OA), kappa coefficient, and confusion matrix. In their calculation, there are four indexes: 1) true positives, i.e., the number of correctly detected changed pixels; 2) true negatives, i.e., the number of correctly detected unchanged pixels; 3) false positives, i.e., the number of false-alarm pixels; and 4) the false negatives, i.e., the number of missed changed pixels. The OA is defined as

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{9}$$

Kappa coefficient reflects the agreement between a final change difference image and the ground-truth map. Compared with OA, kappa coefficients can more objectively indicate the accuracy of the CD results. The higher value of kappa coefficient, the better the result of CD will be. Kappa coefficient is calculated as

$$\mathrm{Kappa} = \frac{\mathrm{OA} - P}{1 - P} \tag{10}$$

TABLE II
CONFUSION MATRIX

Confusion Matrix	Predicted		
		1	0
Actual	1	TP	FN
	0	FP	TN

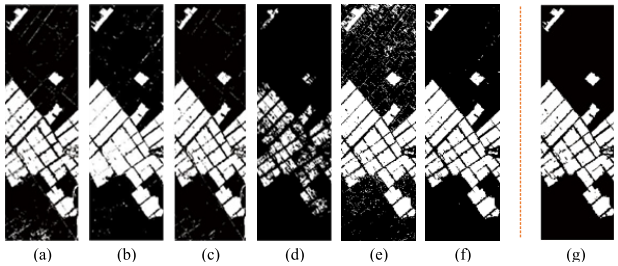

Fig. 7. Change difference images of different methods on the data set of farmland. The change difference image of (a) CVA, (b) PCA-CVA, (c) IR-MAD, (d) SVM, (e) CNN, (f) GETNET, and (g) ground-truth map.

where

$$P = \frac{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2} + \frac{(\mathrm{FN} + \mathrm{TN})(\mathrm{FP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2}. \tag{11}$$

As illustrated in Table II, a confusion matrix represents information about predicted and actual binary classifications (changed and unchanged classifications) conducted by CD algorithms, and it indicates how the instances are distributed in estimated and true classes.

C. Experimental Results

In this paper, the changes between HSIs are at pixel-level variation with no subpixel-level change involved. Spectral unmixing is utilized to obtain subpixel-level information, which can improve the performance of pixel-level CD. In order to demonstrate the effectiveness and generality of the proposed method, we compare it with other state-of-the-art methods, including CVA [34], PCA-CVA [35], IR-MAD [19], support vector machines (SVMs) [21], and patch-based CNN, which takes preprocessed HSIs as input. The principle for choosing these algorithms is due to their variety and popularity. Specifically, they employ different techniques to detect changes, such as image arithmetical operation, image transformation, and deep learning.

For the setup of experiments, the proposed GETNET is implemented by using Keras with Tensorflow as backend. The modified detail of the CNN network architecture is shown in Table I. The GETNET is trained from scratch, and no pretrained models are used in this paper. The batch size used for training is set to 96. The optimizer is Adagrad with epsilon 10^{-8}, and the initial learning rate is set to 10^{-4} without decay. The training parameter setups are the same on all the data sets, which are trained with 30 000 steps. The OA and kappa coefficients of six CD methods on different data sets are shown in Table III. To evaluate the proposed GETNET,

TABLE III
OA AND KAPPA COEFFICIENTS OF GETNET AND OTHER DIFFERENT STATE-OF-THE-ART CD METHODS ON FOUR DATA SETS

Different methods	Index	The experiment datasets			
		farmland	countryside	Poyang lake	river
CVA	OA	0.9523	0.9825	0.9693	**0.9529**
	Kappa	0.8855	0.9548	0.8092	**0.7967**
PCA-CVA	OA	0.9668	0.9276	0.9548	0.9437
	Kappa	0.9202	0.8216	0.7259	0.7326
IR-MAD	OA	0.9604	0.8568	0.8248	0.8963
	Kappa	0.9231	0.8423	0.7041	0.6632
SVM	OA	0.8420	0.9536	0.9583	0.9046
	Kappa	0.6417	0.8767	0.7266	0.6360
CNN	OA	0.9347	0.9033	0.9522	0.9440
	Kappa	0.8504	0.7547	0.8412	0.6867
GETNET(without unmixing)	OA	0.9765 ± 0.00332	0.9758 ± 0.00649	0.9760 ± 0.00258	0.9497 ± 0.00119
	Kappa	0.9367 ± 0.00417	0.9438 ± 0.00219	0.8903 ± 0.00395	0.7662 ± 0.00377
GETNET(Ours, with unmixing)	OA	**0.9783 ± 0.00599**	**0.9851 ± 0.00126**	**0.9856 ± 0.00365**	0.9514 ± 0.00482
	Kappa	**0.9572 ± 0.00495**	**0.9651 ± 0.00173**	**0.9113 ± 0.00408**	0.7539 ± 0.00349

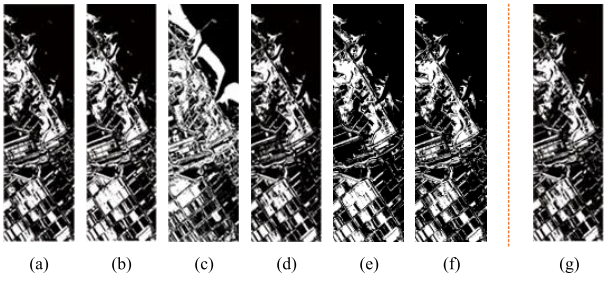

Fig. 8. Change difference images of different methods on the data set of countryside. The change difference image of (a) CVA, (b) PCA-CVA, (c) IR-MAD, (d) SVM, (e) CNN, (f) GETNET, and (g) ground-truth map.

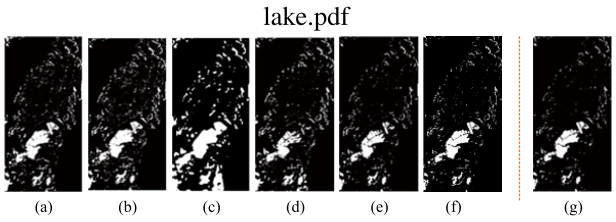

Fig. 9. Change difference images of different methods on the data set of Poyang lake. The change difference image of (a) CVA, (b) PCA-CVA, (c) IR-MAD, (d) SVM, (e) CNN, (f) GETNET, and (g) ground-truth map.

we run five times per data set and get the mean value together with the standard deviation as the final CD results. The results in Table III show that GETNET is robust across different data sets with the same parameter setting. The histogram of the OA comparisons on four data sets in different methods is shown in Fig. 11. The confusion matrices compare the performance of the CNN and GETNET methods in Fig. 12. The comparisons of OA in GETNET and GETNET without unmixing are shown in Fig. 13.

1) Experiments on the Farmland Data Set: As shown in Fig. 7, the main change is the size of farmland on this data set. Table III gives the details of the different methods on four data sets. After overall consideration of OA and kappa coefficient, it can be obviously found that GETNET is an effective method for CD task. With respect to OA,

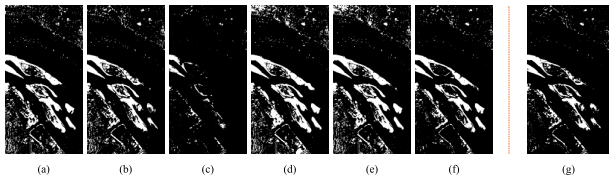

Fig. 10. Change difference images of different methods on the data set of river. The change difference image of (a) CVA, (b) PCA-CVA, (c) IR-MAD, (d) SVM, (e) CNN, (f) GETNET, and (g) ground-truth map.

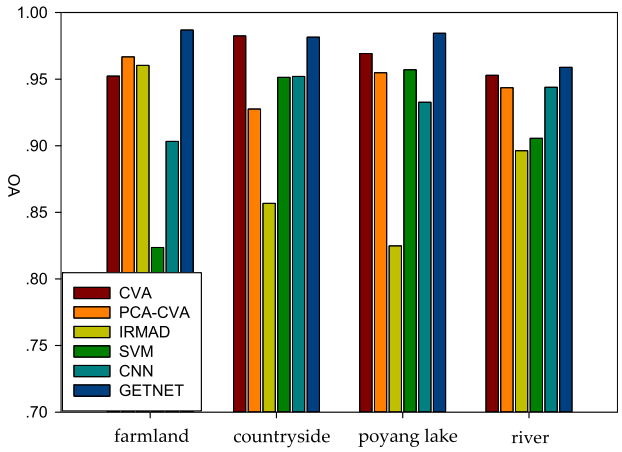

Fig. 11. OA comparisons on four data sets in different methods.

conventional CVA, PCA-CVA, and IR-MAD perform better than CNN, but the performance of SVM is poor. Generally, SVM, CNN, and GETNET use semisupervised learning to train the classifier and network with the same training sets. Therefore, noise in pseudotraining set can have negative effect on the accuracy of semisupervised learning. SVM, as a binary classifier, divides the image into changed and unchanged areas but cannot deal adequately with noise. Moreover, CNN-based deep learning is also not effective as expected since it is more easily affected by the noise in pseudotraining set. As can be clearly seen, the best satisfactory performance is provided by the proposed GETNET, since it makes full use of hyperspectral

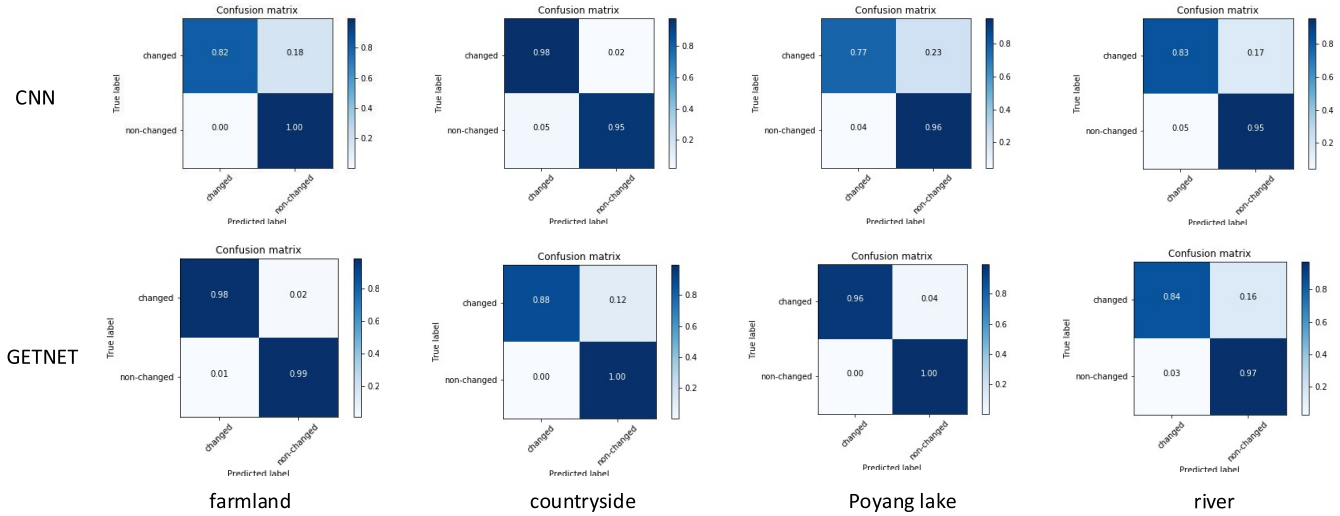

Fig. 12. Confusion matrices of the patch-based CNN and GETNET.

data with abundance information and high antinoise property. Even GETNET without unmixing works well and achieves the second-best performance. From another respect, the kappa coefficient of GETNET is also high, which means that the agreement between GETNET and the ground-truth map is almost perfect.

2) Experiments on the Countryside Data Set: The size of rural areas is the primary change, and the changed information is quite complicated as shown in Fig. 8. In this data set, the best performance is GETNET. Specifically, IR-MAD yields lower OA since it is not very sensitive to complicated changes. Because of dimension reduction, PCA-CVA loses a part of information producing worse performance than CVA. Although the OA of CNN reaches more than 90%, GETNET is appropriately 8% higher. This is primarily because the inputs of them are different. The patch-based CNN only exploits the spatial context information of the preprocessed HSIs, while GETNET utilizes the mixed-affinity matrix, which is an efficient way for simultaneous processing of hyperspectral data with abundance maps. It is worth noting that GETNET and CVA achieve a high OA of over 98%. GETNET without unmixng yields lower OA than CVA, since pseudotraining sets may be not sufficient to effectively train a deep network. Remarkably, abundance information provides subpixel-level presentation, which improves the performance of CD. Though this data set is more complicated, both OA and kappa coefficient of GETNET are even improved than other data sets.

3) Experiments on the Poyang Lake Data Set: On this data set, there are some dispersed changes covering the lake. The performance of IR-MAD is not very satisfactory in terms of both OA and kappa coefficient since it cannot separate various changes well. CVA, PCA-CVA, and SVM methods are relatively insensitive to changes in the width of the spacers. The kappa coefficients of PCA-CVA, IR-MAD, and SVM are also low. Though both CNN and GETNET are based on deep learning, their performances are different. As can be seen in Fig. 9, some of the obvious changes are not recognized well by CNN. However, GETNET is able to identify dispersed changes and changes in the width of the spacers. Overall, the best performance is generated by GETNET on this data set and GETNET without unmixing comes second at OA.

4) Experiments on the River Data Set: The disappearance of substance in river is an obvious change, and beyond that, there are some dispersed changes. Fig. 10 details the visual comparison of six different methods. In this data set, the best result is CVA but not GETNET. It is thought that GETNET should always perform better than CVA because of the pseudotraining sets generated by CVA. However, in a few data sets, these generated data sets may be insufficient to effectively train a deep network without parameter adjustment. The performances of CVA and GETNET are very close, both generating OA up to 95%. PCA-CVA and CNN have the similar accuracy about 94%. Besides, IR-MAD offers lower OA and kappa coefficient since it cannot identify irregular changes well. In addition, a remarkable problem is that the kappa coefficient of each method is relatively lower than the other data set. This is because the data set is more complicated and includes diverse changes. It also demonstrates that the data set we create is reliable.

5) Ablation Study of GETNET: In this paper, the mixed-affinity matrix is constructed as the input of 2-D CNN for discriminative features extraction. As shown in Fig. 3, abundance maps are used to provide extra subpixel information to improve the performance of CD. To validate the effectiveness of the unmixing information, we carry out the experiments of GETNET with and without abundance maps. The experimental results are shown in Fig. 13, which demonstrate that GETNET is almost always better than GETNET without unmixing information across all the four data sets, since subpixel-level representation is of good use to CD. Moreover, it can also be seen from the line chart that the performance of GETNET is stable and robust on different data sets and weight initials.

Although both GETNET and CNN are based on deep learning, the performance of GETNET is better according to Table III and Fig. 12. On the one hand, GETNET with end-to-end manner to train the deep neural network can avoid error

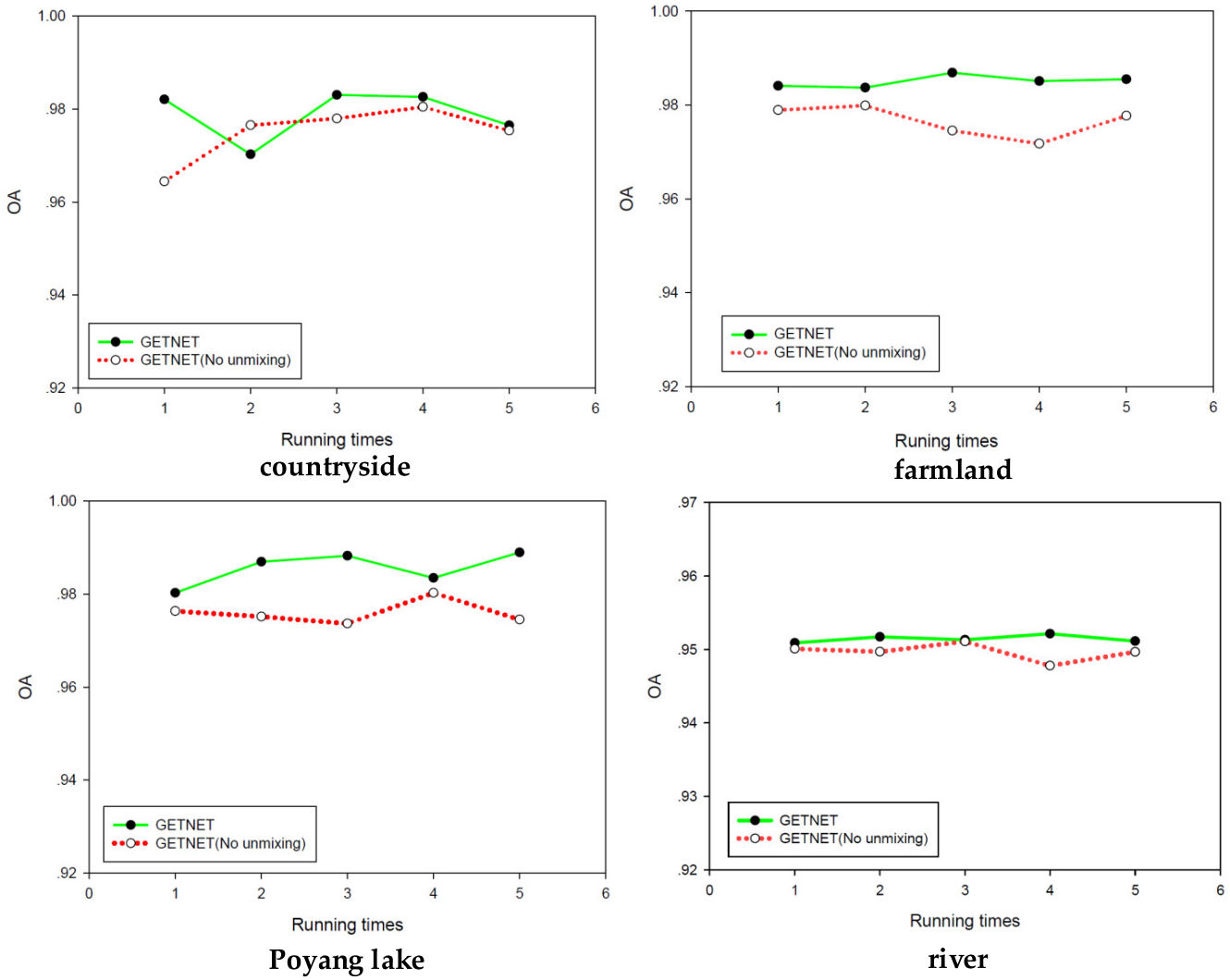

Fig. 13. Comparisons of OA in GETNET and GETNET without unmixing.

accumulation of some deep learning methods with pipeline procedure. On the other hand, the fusion of multisourced data, including hyperspectral data and unmixing abundance information, makes GETNET to mine the characteristics of spectral gradient producing higher robustness, accuracy, and generalization.

The proposed method outperforms the other CD methods in most of the cases and is robust for different HSIs. The data sets used in the experiments involve multiple change types, which demonstrate the merits of GETNET. The merits are mainly due to the novel mixed-affinity matrix with abundant information and a unique design of GETNET.

V. Conclusion

In this paper, a general framework named GETNET is proposed for the HSI-CD task by the 2-D CNN. First, a novel mixed-affinity matrix is designed. It is not only an efficient way for simultaneous processing of multisource information fusion, but also provides more abundant cross-channel gradient information. Next, based on deep learning, GETNET is developed to learn a series of more significant features fully mined with excellent capabilities of generalization and robustness. Moreover, the new HSI-CD data set is designed for objective comparison of different methods. In the end, the proposed method is evaluated on the existing and newly constructed data sets and shows its effectiveness through extensive comparisons and analyses.

References

[1] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.

[2] A. Koltunov and S. L. Ustin, "Early fire detection using non-linear multitemporal prediction of thermal imagery," *Remote Sens. Environ.*, vol. 110, no. 1, pp. 18–28, 2007.

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

[4] T. T. Lê, V. A. Tran, H. T. Pham, and X. T. Tran, "Change detection in multitemporal SAR images using a strategy of multistage analysis," in *Proc. Int. Conf. Geo-Spatial Technol. Earth Resour.*, 2017, pp. 152–165.

[5] S. G. A. Usha, S. Vasuki, S. G. A. Usha, and S. Vasuki, "Unsupervised change detection of multispectral imagery using multi level fuzzy based deep representation," *J. Asian Sci. Res.*, vol. 7, no. 6, p. 206, 2017.

[6] 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.

[7] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[8] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2018.2828161.
[9] C.-C. Hsieh, P.-F. Hsieh, and C.-W. Lin, "Subpixel change detection based on abundance and slope features," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, Jul./Aug. 2006, pp. 775–778.
[10] A. Ertürk, M.-D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 9, no. 2, pp. 708–719, Feb. 2015.
[11] O. A. C. Júnior, R. F. Guimarães, A. R. Gillespie, N. C. Silva, and R. A. T. Gomes, "A new approach to change vector analysis using distance and similarity measures," *Remote Sens.*, vol. 3, no. 11, pp. 2473–2493, 2011.
[12] 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.
[13] 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, May 2012.
[14] Z. Chen, B. Wang, Y. Niu, W. Xia, J. Q. Zhang, and B. Hu, "Change detection for hyperspectral images based on tensor analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 1662–1665.
[15] 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.
[16] 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.
[17] J. S. Deng, K. Wang, Y. H. Deng, and G. 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.
[18] 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, 1998.
[19] 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.
[20] 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.
[21] H. Nemmour and Y. Chibani, "Multiple support vector machines for land cover change detection: An application for mapping urban extensions," *I SPRS J. Photogram. Remote Sens.*, vol. 61, no. 2, pp. 125–133, 2006.
[22] J. Zhao, M. Gong, J. Liu, and L. Jiao, "Deep learning to classify difference image for image change detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 411–417.
[23] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, Jan. 2015.
[24] Y. Han, A. Chang, S. Choi, H. Park, and J. Choi, "An unsupervised algorithm for change detection in hyperspectral remote sensing data using synthetically fused images and derivative spectral profiles," *J. Sensors*, vol. 2017, Aug. 2017, Art. no. 9702612, doi: 10.1155/9161.
[25] A. Ertürk 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.
[26] A. Ertürk, S. Ertürk, and A. Plaza, "Unmixing with slic superpixels for hyperspectral change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2016, pp. 3370–3373.
[27] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Multitemporal spectral unmixing for change detection in hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 4165–4168.
[28] H. Ren and C.-I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
[29] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.
[30] J. Yu, D. Chen, Y. Lin, and S. Ye, "Comparison of linear and nonlinear spectral unmixing approaches: A case study with multispectral TM imagery," *Int. J. Remote Sens.*, vol. 38, no. 3, pp. 773–795, 2017.
[31] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[32] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2518–2525.
[33] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363–375, Jan. 2015.
[34] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with Landsat," in *Proc. LARS Symp.*, 1980, p. 385.
[35] M. Baisantry, D. S. Negi, and O. P. Manocha, "Change vector analysis using enhanced PCA and inverse triangular function-based thresholding," *Defence Sci. J.*, vol. 62, no. 4, pp. 236–242, 2012.

Qi Wang (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer Science, the Unmanned System Research Institute, and the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

Zhenghang Yuan received the B.E. degree in information security from Northwestern Polytechnical University, Xi'an, China, in 2017, where she is currently pursuing the M.S. degree in computer science with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning.

Her research interests include hyperspectral image processing and computer vision.

Qian Du (S'98–M'00–SM'05–F'17) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2000.

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Xuelong Li (M'02-SM'07-F'12) is currently a Full Professor with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China, and also with the University of Chinese Academy of Sciences, Beijing, China.