This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images

Jia Liu, Maoguo Gong, *Senior Member, IEEE*, Kai Qin, *Senior Member, IEEE*, and Puzhao Zhang

***Abstract*— We propose an unsupervised deep convolutional coupling network for change detection based on two heterogeneous images acquired by optical sensors and radars on different dates. Most existing change detection methods are based on homogeneous images. Due to the complementary properties of optical and radar sensors, there is an increasing interest in change detection based on heterogeneous images. The proposed network is symmetric with each side consisting of one convolutional layer and several coupling layers. The two input images connected with the two sides of the network, respectively, are transformed into a feature space where their feature representations become more consistent. In this feature space, the different map is calculated, which then leads to the ultimate detection map by applying a thresholding algorithm. The network parameters are learned by optimizing a coupling function. The learning process is unsupervised, which is different from most existing change detection methods based on heterogeneous images. Experimental results on both homogenous and heterogeneous images demonstrate the promising performance of the proposed network compared with several existing approaches.**

***Index Terms*— Change detection, deep neural network, denoising autoencoder optical images, synthetic aperture radar images.**

I. Introduction

DETECTION of changes on the surface of the earth is becoming increasingly important for monitoring environments and resources [1]. With the advance of remote sensing technology, the earth can be observed via remote sensing imagery. Accordingly, changes on the earth's surface can be identified by using image change detection techniques [2]. Change detection is defined as the process of identifying variations of an object or a phenomenon by observing it at different times [3]. It plays a key role in many real-world applications, e.g., urban growth tracking [4], land use monitoring [5], and disaster evaluation [6]. Given two remote sensing images of the same scene but taken at different times, image preprocessing techniques, including denoising and coregistration, are first applied. Denoising is used to reduce the impact of corruptions for accurately detecting changes. The coregistration [7] is necessary, because the multitemporal images may have different sizes and contain objects of interest at different locations, scales, and/or viewpoints. Image coregistration aligns the raw multitemporal images into one coordinate system via image transformations (e.g., translation, rotation, and scaling), aiming to make any pair of pixels at the same position in two aligned images denote the same geolocation. Then, by comparison and analysis of the two preprocessed images, changed areas are detected. Change detection mainly focuses on the comparison and analysis process. Considering the importance of change detection in earth resources surveys, environmental monitoring, regional planning, and global macroresearch, it has attracted a lot of research interests [8]–[15]. There exist various kinds of remote sensing data acquired by different imaging sensors mounted on satellites or aircrafts, e.g., very high resolution (VHR) optical images, multispectral/hyperspectral images, synthetic aperture radar (SAR) images, and polarimetric SAR images. Accordingly, change detection techniques can be divided into two categories based on homogeneous and heterogeneous images, respectively. Here, "heterogeneous" means different remote sensing methods and we define the heterogeneous images as remote sensing images captured by different types of sensors with different characteristics, e.g., optical and radar sensors. Then, the homogeneous images are the multitemporal images acquired by homogeneous sensors between which the intensity of unchanged pixels are assumed to be linearly correlated [16]. Different sensors may capture distinct statistical properties [17] of the same ground object, and thus, inconsistent representations of heterogeneous images may make change detection more difficult than using homogeneous images. This paper is based on heterogeneous optical and SAR images.

Change detection based on homogeneous images has been widely investigated and many existing methods can detect changes with high accuracy. Optical and SAR images are two main sources for such studies. For homogeneous images, it is easy to derive a difference map by the direct computation of pixelwise difference [18]–[20]. Furthermore, context information is usually used to reduce the impact of noise, misregistration, and other distortions [21], [22]. Pixels corresponding to changes are then identified by analysing

Manuscript received January 13, 2016; revised August 9, 2016; accepted November 30, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61422209, in part by the National Program for Support of Top-Notch Young Professionals of China, and in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130203110011.

J. Liu, M. Gong, and P. Zhang are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China *(Corresponding author: M. Gong, e-mail: gong@ieee.org)*.

K. Qin is with the School of Computer Science and Information Technology, RMIT University, Melbourne, VIC 3001, Australia.

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/TNNLS.2016.2636227

2162-237X © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the difference map. Thresholding methods are often used in this process [23], [24]. However, due to the difficulty of estimating intensity distributions for different types of difference maps in thresholding methods, many advanced approaches were proposed, such as level set-based methods [25], neural network-based methods [26], [27], and fuzzy clustering-based methods [28], [29].

Detecting changes based on heterogeneous images is very promising but more challenging. Nowadays, because optical images are easier and cheaper to be obtained than SAR images, large volumes of optical images become available [16]. However, acquiring high-quality optical images is subjected to clear weather and good sunlight due to the passive property of optical sensors. In contrast, SAR sensors are active and thus insensitive to weather and sunlight conditions [28]. With the complementary properties of these two types of sensors, detecting changes from optical and SAR images accurately is of great significance for many real applications, especially for the immediate evaluation of emergency disasters. In such cases, the preevent SAR image is usually unavailable and the qualified postevent optical image cannot be obtained immediately. However, due to the difficulty of calculating pixelwise difference between heterogeneous images, there exist only a few works based on heterogeneous images [6], [16], [30]–[34]. One major challenge of dealing with heterogeneous images comes from the distinct feature representations of ground objects in different types of images, which increases the difficulty of obtaining the difference map. Optical sensors measure the intensity of the reflected light in the visual and near-infrared spectral bands. Accordingly, the appearances of ground objects in an optical image are determined by the surface reflection characteristics of these objects, the scene illumination conditions, and the sensor perspective. SAR sensors are active, measuring the backscatter of a transmitted signal. This signal is typically with a narrow microwave frequency band and sampled in the range direction. Therefore, the appearances of ground objects in an SAR image are determined by the geometry and dielectric properties of these objects and the transmitting/receiving configuration of the SAR sensor. Consequently, ground objects in optical and SAR images would demonstrate remarkably different appearances.

The problem of change detection based on heterogeneous images can be addressed by using classification-based methods [30], [31]. Postclassification comparison (PCC) [5], [35] is one of such methods applicable to both homogeneous and heterogeneous images. In PCC, the classification map of each image is first derived independently. Then, these obtained classification maps are compared and summarized to detect changes. The accuracy of such methods depends strongly on the performance of classification algorithms. Meanwhile, such methods may suffer from the accumulation of classification errors. Compared with PCC, multidate classification [36], [37] tends to obtain more accurate results and provides class transition information. However, this approach requires sufficient labeled pixels for training the classifier, which might not be available in practice. Recently, the research on change detection based on heterogeneous images is becoming popular. A method for building damage assessment was proposed in [6]. In this method, by rendering and matching the estimated parameters of buildings, the scene in optical images can be represented by its SAR signatures. Then, the changes of buildings are detected by comparison. The approach in [32] transforms one of the two images used for change detection to achieve the similar characteristics to the other image by using the copula theory. However, to learn the appropriate copula, labeled pixels must be used for training. Manifold learning [33], kernel canonical correlation analysis [34], and Bayesian nonparametric model associated with a Markov random field [16] were also attempted to detect changes based on heterogeneous images. However, these methods rely on hand-crafted analysis of image properties and learn the latent correlations between heterogeneous images via a set of manually labeled unchanged pixels. This limits the application of such methods.

In this paper, we use two heterogeneous images (one optical image and one SAR image) for change detection by transforming them into a consistent feature space via a specially designed deep neural network. Deep neural networks are one of the most popular machine learning techniques, which imitate the information processing mechanism of mammalian brains and have the properties of automatical feature learning and hierarchical information representation. Then, the properties and the features of raw images can be extracted and transformed automatically for comparison via the deep architecture. Therefore, it is difficult to define the consistent feature space, which is generated automatically during the training process. In particular, we had successfully applied a deep neural network to detect changes based on homogeneous images in [38]. Due to the powerful learning capability of this network, the difference map generation and the analysis are simultaneously achieved by the network during the learning process. In this paper, we propose a symmetric convolutional coupling network (SCCN) for change detection based on heterogeneous images. There are three major characteristics of SCCN. First, the network has a symmetrical structure with each of the two sides composed of convolution and coupling layers for feature extraction and transformation. Second, feature transformation is made simultaneously from both sides of the network via hierarchical coupling layers by minimizing a coupling function, which measures the pixelwise difference summed over unchanged pixels. Third, the network parameters are first initialized by layerwise feature extractors incorporated with certain noise models of images to reduce the impact of noise and extract useful features, which facilitate the learning process. Different from the above-mentioned existing change detection methods based on heterogeneous images, the proposed method is fully unsupervised where unchanged pixels are automatically identified and utilized during the learning process.

The rest of this paper is organized as follows. In Section II, the problem statement and the background of deep neural networks are introduced. The proposed SCCN and its learning process are described in detail in Section III. In Section IV, the experimental settings and the results are discussed. Finally, Section V concludes this paper and mentions the future work.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

II. Background

A. Problem Statement

Suppose two images I_1 and I_2 were acquired on two different dates t_1 and t_2, respectively, and coregistered. The corresponding ground objects in the two images are geometrically aligned. The goal of change detection is to generate a binary map BM of the same size to I_1 and I_2, which indicates the locations of the changes happening between t_1 and t_2. In other words, change detection is to label each pixel of the two coregistered images as either changed or unchanged. This process can be formulated as

$$\mathrm{BM} = F(I_1, I_2) = f_A(f_1(I_1) \ominus f_2(I_2)) \tag{1}$$

where $\ominus$ is the difference operator. f_1 and f_2 represent feature extraction. The intensity, spectral, structure, and statistical features can be extracted and compared to generate a difference map by $\ominus$. f_A analyzes the obtained difference map and accordingly classify each pixel to generate the BM. Many techniques follow this formulation including classification-based approaches and heterogeneous images-based change detection methods. In classification-based approaches, f_1 and f_2 represent two classifiers trained independently or jointly. For heterogeneous images, f_1 and f_2 can be trained on unchanged pixels by considering sensors' physical properties, the noise models, and local joint distributions [16].

Our proposed method also follows this formulation. In this paper, a symmetric network is established, which consists of two sides with a similar architecture. f_1 and f_2 correspond to the two sides of the network, which take as input I_1 and I_2, respectively. f_1 and f_2 map I_1 and I_2 into a high-dimensional feature space, aiming to make unchanged areas in these two images to demonstrate similar characteristics in such a space. To achieve this goal, a coupling function measuring the total difference over the unchanged pixels is defined. This function drives the learning of the parameters of the network. During learning, the network parameters are initialized through a feature learning model, i.e., denoising autoencoder (DAE), by considering the noise models of the two input images. Before detailing our proposed method, deep neural networks and feature learning models are first introduced.

B. Deep Neural Networks and Feature Learning Models

Intuitively, deep neural networks are neural networks involving many hidden layers. Compared with traditional shallow neural networks, deep neural networks are capable of automatically extracting relevant features from raw data in a hierarchical manner during the learning process. Convolutional neural network (CNN) is one of the earliest deep neural networks, which has become one of the most popular networks now mainly due to its remarkably outstanding performance achieved in a variety of image processing tasks [39]. However, the most notable breakthrough in deep neural networks is a fast greedy learning algorithm designed for deep belief networks (DBNs) [40]. Following that, the field of representation learning (also called feature learning) was established [41] and accordingly, various feature learning models were proposed [42]–[46]. Deep neural networks have been successfully applied to various fields, such as speech processing, nature language processing, computational biology, and image processing [47]–[50]. Especially, they have been used to deal with other heterogeneous data [17], [51].

1) CNN: CNN was specially designed for image (video) data in which local spatial (spatial-temporal) information is crucial. In CNN, a convolution layer is composed of a set of feature maps generated by applying multiple trainable convolution kernels (i.e., filters) to the previous layer. The pooling layer follows the convolution layer to subsample each of its involved feature maps so as to get rid of redundant features and producing robustness to spatial shifting. Through several alternating convolution and pooling operations, a feature vector representing the input image is eventually generated. The parameters of the network including convolution kernel parameters and activation function biases are learned by using stochastic gradient descent with backpropagation algorithm. The training process is supervised. The representation of the input image is automatically learned instead of hand-crafted.

2) DBN: DBN is a generative model consisting of multiple stacked restricted Boltzmann machines (RBMs) [40]. CNN provides a suitable network structure for learning features from images, while DBN inspires the unsupervised layerwise feature learning and pretraining for deep architectures. The greedy layerwise learning algorithm proposed in [40] resolves the issues encountered when the traditional backpropagation algorithm is applied to train deep neural networks. Specifically, the bottom-level RBM is first trained on the input data, and accordingly, the outputs of hidden units can be inferred. The inferred outputs of hidden units are then used as the input data to train the RBM at a higher level. Unsupervised learning had a catalytic effect in reviving interest in deep neural networks [52]. The features extracted by unsupervised stacked models can be used for initializing a deep neural network or directly fed into other learning models. Although it has been overshadowed by purely supervised learning in many learning problems recently, the success of supervised learning satisfies the situation in which there exist enough labeled training samples. In this problem, there are even no labeled samples. Therefore, the unsupervised learning is necessary here for automatically learn the label information.

3) Feature Extractors: RBM and autoencoder [53] are two basic feature learning models. RBM is a probabilistic model in which the parameters are trained to maximize the probability the network assigns to visible data. Unlike RBM, autoencoder is a neural network model where the outputs of hidden units are computed by

$$\boldsymbol{h}^j = s(W^j \boldsymbol{v} + \boldsymbol{b}^j). \tag{2}$$

Given the outputs of hidden units which form a hidden vector, a decoder is used to reconstruct the visible vector from the hidden vector

$$\boldsymbol{v}'^j = W'^j \boldsymbol{h} + \boldsymbol{b}'^j \tag{3}$$

where W' and $\boldsymbol{b}'$ are weight matrix and bias vector of the decoder, respectively. Generally, W' is constrained to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

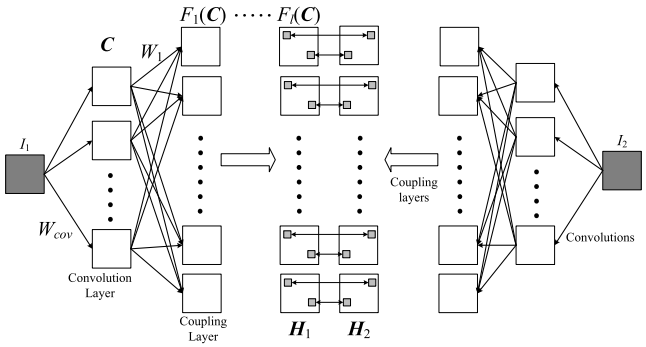


Fig. 1. Structure of SCCN. In the coupling layers, all the pixels share the same weight matrix, e.g., W_1 for the first coupling layer.

$W' = W^T$. A good representation should retain enough information of its corresponding visible vector, and accordingly, the network parameters are learned by minimizing the reconstruction error of the visible data $L(\boldsymbol{v}, \boldsymbol{v}')$.

DAE proposed in [42] increases robustness to corruptions. Since remote sensing images, especially SAR images, are easily contaminated by noise, we give more attentions on DAE. DAE is a simple autoencoder variant, aiming to recover a clean input from its corrupted version. This can be achieved by first deliberately corrupting visible data $\boldsymbol{v}$ to produce its corrupted version $\tilde{\boldsymbol{v}}$. $\tilde{\boldsymbol{v}}$ can be obtained via a stochastic corruption process $\tilde{\boldsymbol{v}} \sim q(\tilde{\boldsymbol{v}}|\boldsymbol{v})$. $\tilde{\boldsymbol{v}}$ is then mapped to generate $\boldsymbol{h}$ and $\boldsymbol{v}'$ as autoencoder does. Next, the network parameters are learned by minimizing the error between $\boldsymbol{v}$ and $\boldsymbol{v}'$. After training, the input $\boldsymbol{v}$ is mapped to generate its representation $\boldsymbol{h}$ by using the trained parameters W and $\boldsymbol{b}$. The corruption process is important for DAE. Different noise models can be incorporated into the learning process depending on different applications. In this paper, we implement the corruption process by considering the noise models of SAR and optical images, respectively.

III. Methodology

Given two heterogeneous images I_1 and I_2 captured on different dates t_1 and t_2, respectively, we aim to identify changes happening between t_1 and t_2 based on these two images. This will be achieved by first transforming I_1 and I_2 into a feature space in which these two images have more consistent feature representations, and then identifying changes based on such feature representations. Specifically, we develop an SCCN to implement feature transformation. In the transformed feature space, each pixel has a feature vector associated with it, and accordingly, the input image is represented by a set of feature maps. The pixelwise difference is then calculated based on these feature maps to detect changes.

A. Network Structure

The proposed SCCN is a symmetric network with each side consisting of one convolution layer and several coupling layers, as shown in Fig. 1. Since we aim to estimate the pixelwise difference instead the regional difference, the pooling layer is not used here. Following up the convolution layer, several coupling layers sequentially transform the feature maps obtained by the convolution layer into a feature space in which the two input images fed from the two sides of SCCN have more consistent feature representations. Because we aim to compute the pixelwise difference, the connections to any coupling layer only apply to the same pixel positions and each pixel in any feature map of a coupling layer use the same set of connecting weights to reduce the total amount of network parameters. This amounts to 1×1 convolution kernels for the input feature maps. As shown in Fig. 1, W_1 is applied to all the pixels.

SCCN takes two input images of the same size from its two sides, respectively. In the middle of SCCN, the two pixelwise representations of the two input images are used to define the objective function, i.e., the coupling function which intends to make any pair of unchanged pixels at the same position in two input images have similar representations, and consequently, the pairs of changed pixels would demonstrate disparate feature representations that can be easily detected. In this paper, the number of convolutional and coupling layers as well as the number of neurons per layer is set the same for both sides of the network. Therefore, SCCN has a symmetric structure with respect to its two sides.

B. Formulation

Suppose an image I fed to one side of the network is transformed via one convolution and several coupling layers, denoted by $\boldsymbol{H} = F_{\mathrm{CC}}(I) = F_{\mathrm{map}}(F_{\mathrm{cov}}(I))$. According to the structure of SCCN shown in Fig. 1, the final difference map is generated by carrying out the following three operations.

1) Convolution: This operation is performed via the convolution layer in SCCN. It transforms an input image I into several feature maps by applying several convolution kernels, i.e., $\boldsymbol{C} = F_{\mathrm{cov}}(I)$

$$\boldsymbol{C}^i = s\left(W^i_{\mathrm{cov}} * I + \boldsymbol{b}^i_{\mathrm{cov}}\right) \tag{4}$$

where $\boldsymbol{C}^i$ is the ith feature map obtained by the ith convolution kernel W^i_{cov}. $\boldsymbol{b}^i_{\mathrm{cov}}$ represents the bias value associated with the ith feature map. $*$ denotes the convolution operation. In SCCN, the sigmoid function is used as the activation function. Given n_{cov} convolution kernels $W_{\mathrm{cov}} = \{W^1_{\mathrm{cov}}, \ldots, W^{n_{\mathrm{cov}}}_{\mathrm{cov}}\}$ associated with n_{cov} bias values $\boldsymbol{b}_{\mathrm{cov}} = \{\boldsymbol{b}^1_{\mathrm{cov}}, \ldots, \boldsymbol{b}^{n_{\mathrm{cov}}}_{\mathrm{cov}}\}$, n_{cov} feature maps will be generated as the output of the convolution layer, i.e., $\boldsymbol{C} = \{\boldsymbol{C}^1, \ldots, \boldsymbol{C}^{n_{\mathrm{cov}}}\}$.

Convolution kernels are often designed manually, such as Laplacian of Gaussian kernels and Gabor kernels. In deep neural networks, the parameters of such convolution kernels are learned during the training process. The learning procedure can be supervised via backpropagation algorithms or unsupervised via feature learning techniques.

2) Mapping: This operation is performed via several coupling layers in SCCN. It transforms the feature maps $\boldsymbol{C}$ obtained by the convolution layer into a feature space where the two input images have more consistent feature representations and thus can be directly compared, i.e., $\boldsymbol{H} = F_{\mathrm{map}}(\boldsymbol{C})$

$$F^j_i(\boldsymbol{C})(x, y) = s\left(W^j_i F_{i-1}(\boldsymbol{C})(x, y) + \boldsymbol{b}^j_i\right) \tag{5}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where $i = 1, 2, \ldots, l$ denotes the index of a coupling layer. $F_i^j(\boldsymbol{C})$ denotes the jth feature map of the ith coupling layer and (x, y) represents the pixel position. Accordingly, $F_i(\boldsymbol{C})(x, y)$ stands for the feature vector at the position (x, y) of the ith coupling layer. W_i is an $n_i \times n_{i-1}$ weight matrix, which contains the weights of the connections between the $(i-1)$th layer and the ith layer, where n_i denotes the number of feature maps in the ith layer. $\boldsymbol{b}_i$ denotes the bias vector for the ith coupling layer. In this operation, there are l coupling layers and two types of parameters, i.e., the connection weight matrices $W_{\text{map}} = \{W_1, W_2, \ldots, W_l\}$ and the bias vectors $\boldsymbol{b}_{\text{map}} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_l\}$. The feature maps obtained at the convolution layer, denoted by $F_0(\boldsymbol{C}) = \boldsymbol{C}$ propagate through the l coupling layers to produce new feature representation $\boldsymbol{H} = F_{\text{map}}(\boldsymbol{C}) = F_l(\boldsymbol{C})$.

3) Difference Map Generation: Two heterogeneous images I_1 and I_2 fed to the two sides of SCCN are transformed into more consistent feature representations $\boldsymbol{H}_1 = F_{\text{CC1}}(I_1)$ and $\boldsymbol{H}_2 = F_{\text{CC2}}(I_2)$, respectively. Here, $F_{\text{CC1}}()$ and $F_{\text{CC2}}()$ stand for the operations with respect to the two sides of the network, respectively. Although SCCN is symmetric according to its original definition, the number of coupling layers with respect to the two sides of the network is allowed to be different to adapt to the different properties of the two input images. The difference map is generated by calculating the pixelwise distances between $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ as follows:

$$D(x, y) = \|\boldsymbol{h}_1(x, y) - \boldsymbol{h}_2(x, y)\|_2 \tag{6}$$

where D denotes the difference map in which the value of each pixel is the pixelwise difference between the finally obtained feature representations ($\boldsymbol{H}_1$ and $\boldsymbol{H}_2$) of the two input images. $\boldsymbol{h}_1(x, y)$ and $\boldsymbol{h}_2(x, y)$ represent the learned feature representations (i.e., feature vectors) of the pixels at position (x, y) in two input images, respectively. Since $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are more consistent feature representations of the two input images, the simple Euclidean distance is used here to measure the difference. The larger distance means that the corresponding pixel is more distinct.

C. Learning

Based on the principle that unchanged regions should demonstrate smaller pixelwise difference, the coupling function is defined as

$$\begin{aligned} \min \ F_{\text{cop}}(\theta, P_u) = &\sum_{(x,y)} P_u(x, y)\|\boldsymbol{h}_1(x, y) - \boldsymbol{h}_2(x, y)\|_2 \\ &- \lambda \sum_{(x,y)} P_u(x, y) \\ \text{s.t. } 0 \le P_u(x, y) \le 1 \end{aligned} \tag{7}$$

where θ represents the set of network parameters, i.e., $\theta = \{W_{\text{cov}}, \boldsymbol{b}_{\text{cov}}, W_{\text{map}}, \boldsymbol{b}_{\text{map}}\}$. P_u is a probability map with the value at each position (x, y) indicating the probability that no change occurs at this position. The regularization term $-\lambda \sum_{(x,y)} P_u(x, y)$ is used to prevent a full zero-probability map, i.e., $P_u = \{P_u(x, y) \equiv 0\}$. Here, λ is a user-defined weighting parameter.

Algorithm 1 Learning Procedure in SCCN

1.Initialization: randomly initializing P_u, $P_u(x, y) = \text{rand}(0, 1)$.
2.Updating θ: fixing P_u, minimizing the objective function in Eq. (7) with respect to θ by using the back-propagation algorithm.
3.Updating P_u: fixing θ, updating $P_u(x, y)$ via Eq. (8) for each position (x, y).
4.Repetition: repeating steps 2 and 3 until the value of the objective function in Eq. (7) is unchanged.

This definition is easy to understand. The aim is to align the unchanged regions and highlight the changed regions. The unchangedness probability map P_u encodes pixel label information, which is learned with the network parameter set θ. The first part of (7) stands for the weighted sum of pixelwise similarities between the learned feature representations of two input images. The larger $P_u(x, y)$ is, the more the pair of pixels at position (x, y) in two input images contributes to the learning process. The second part of (7) imposes a regularization on $P_u = \{P_u(x, y)\}$ to avoid the null probability map denoting all pixels in input images are labeled as changed, which can seldom occur in practical scenarios where there exist at least a few unchanged pixels. The pixels of nonzero $P_u(x, y)$ drive the learning of network parameters. A user-defined parameter λ is used to control the balance between the two terms. Note that both θ and P_u need to be optimized. We use alternative convex search [54] to minimize this objective function. Specially, the two sets of parameters θ and P_u are optimized alternately. Given a fixed θ, we can obtain the optimal P_u. Then, by fixing P_u, we can obtain the optimal θ. When P_u is fixed, θ is optimized by using the backpropagation algorithm [39], which is widely used to optimize neural network parameters. When θ is fixed, $\|\boldsymbol{h}_1(x, y) - \boldsymbol{h}_2(x, y)\|$ is computed as a constant. Then, $F_{\text{cop}}(\theta, P_u)$ is a monotonic function with respect to P_u, i.e., $F_{\text{cop}}(\theta, P_u) = \sum_{(x,y)}(\|\boldsymbol{h}_1(x, y) - \boldsymbol{h}_2(x, y)\|_2 - \lambda)P_u(x, y)$. Accordingly, the optimal $P_u(x, y)$ is the boundary of the interval, i.e., 0 or 1. The optimal $P_u(x, y)$ can be derived as

$$P_u(x, y) = \begin{cases} 1 & \|\boldsymbol{h}_1(x, y) - \boldsymbol{h}_2(x, y)\|_2 < \lambda \\ 0 & \|\boldsymbol{h}_1(x, y) - \boldsymbol{h}_2(x, y)\|_2 \ge \lambda. \end{cases} \tag{8}$$

Then, the defined probability $P_u(x, y)$ degenerates to binary value. Therefore, P_u can be taken as the index matrix of unchanged pixels for training the network parameters. Then, minimizing the objective function means to minimize the coupling error of unchanged pixels, which highlights the changed pixels. The parameter learning procedure which aims to minimize the objective function in (7) is summarized in Algorithm 1.

Since SCCN is bilateral, minimizing the coupling function in (7) with respect to P_u and θ may result in degenerated feature representations, i.e., $\boldsymbol{h}_1(x, y) \equiv \boldsymbol{h}_2(x, y) \equiv C$, where C is a constant. In such case, there are no constraints on the feature space. The consistent feature space should align the unchanged regions and meanwhile represent the images well.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

To address this issue, we first apply DAE to each side of SCCN. Then, the network parameters of one side, derived by DAE, are fixed. Next, the network parameters of another side are learned together with P_u by minimizing (7) via backpropagation using the parameters obtained by DAE as the starting point. Then, the consistent feature space generated by unsupervised learning can well represent the input images. Meanwhile, the network is pretrained via unsupervised feature learning method, aiming to obtain a good initialization to facilitate the subsequent backpropagation algorithm. Although unsupervised learning plays little role in many supervised learning problems, it is necessary to form a good initialization for both θ and P_u in this problem. Note that in this paper, the target label that indicates the changed or unchanged status of a pixel in the input image is completely unknown in advance. Therefore, the method is unsupervised. The method can also be semisupervised if any label information is available in advance and integrated into P_u.

D. Unsupervised Pretraining

To use DAE, we need to first model the corruption process, i.e., determining $\tilde{\boldsymbol{v}} = q(\tilde{\boldsymbol{v}}|\boldsymbol{v})$. In this paper, we first consider the noise models of SAR and optical images. SAR images are severely corrupted by speckle noise due to the active nature of SAR sensors. Because the surfaces of ground objects are not smooth, the phase of the echo signal is not coherent. Therefore, SAR images obtained by coherent imaging suffer from distorted intensity. Speckle noise in homogeneous areas is distributed following Gamma distribution $n_{\text{SAR}} \sim \Gamma(L, L^{-1})$, where L is the number of looks of SAR sensors. Since the speckle noise model is multiplicative, the corruption process for speckle noise is defined by

$$\tilde{\boldsymbol{v}}^i = \boldsymbol{v}^i n_{\text{SAR}} \tag{9}$$

where n_{SAR} is a random number which follows the Gamma distribution. For optical images, pixel intensity is usually influenced by additive Gaussian noise $n_{\text{opt}} \sim \mathcal{N}(0, \sigma^2)$. Accordingly, the corruption process for Gaussian noise is defined by

$$\tilde{\boldsymbol{v}}^i = \boldsymbol{v}^i + n_{\text{opt}} \tag{10}$$

where n_{opt} is a random number which follows the Gaussian distribution. After modeling these two types of corruption processes, DAE is utilized to initialize the network parameters layer by layer. For each layer, DAE is implemented by minimizing the total reconstruction error defined as follows:

$$f_e(W, \boldsymbol{b}) = \sum_{(x,y)} L(\boldsymbol{v}(x, y), \boldsymbol{v}'(x, y)) \tag{11}$$

where $\boldsymbol{V}$ is the set of feature maps in the visible layer. $\boldsymbol{v}(x, y)$ and $\boldsymbol{v}'(x, y)$ denote the original and reconstructed feature vector, respectively, at the pixel position (x, y). $f_e(W, \boldsymbol{b})$ denotes the sum of reconstruction errors over feature vectors at all pixel positions. Here, the reconstruction error is defined as the mean square error, i.e., $L(\boldsymbol{v}(x, y), \boldsymbol{v}'(x, y)) = \|\boldsymbol{v}(x, y) - \boldsymbol{v}'(x, y)\|_2^2$. The learned W and $\boldsymbol{b}$ will serve as the initialization of the network parameters of the current layer. DAE can be seen as a way to define and learn a manifold [55]. It is similar to the method in [16] where a vector belonging to a manifold that relates to the properties of different sensors is estimated for each sliding window.

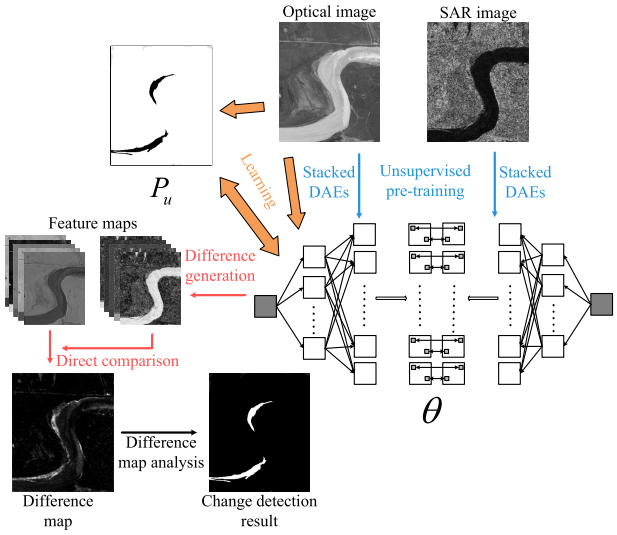


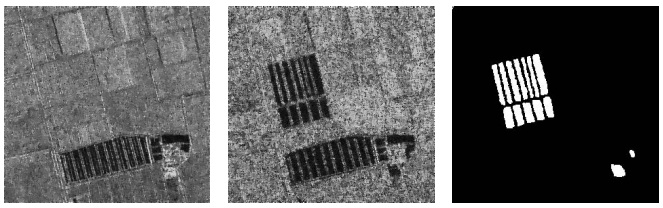
Fig. 2. Work flow of SCCN. Different steps are marked by different colors. The network parameter set θ is first pretrained by stacked DAEs. Then, the unchanged regions are aligned by optimizing the coupling function, i.e., the learning process. Finally, the input images are fed into the trained network and obtain the feature maps. The difference map is generated by comparing the feature maps.

E. Analysis

Given SAR and optical images I_1 and I_2, DAEs with respect to speckle and Gaussian noise models are first implemented to initialize the parameters of SCCN. For the side of SCCN connected with the SAR image, the speckle noise model is employed. For the other side connected with the optical image, the Gaussian noise model is utilized. Then, the learning procedure for minimizing the coupling function is executed to further adjust the parameters. Eventually, I_1 and I_2 are fed to the network to generate the difference map. We show how SCCN works in Fig. 2.

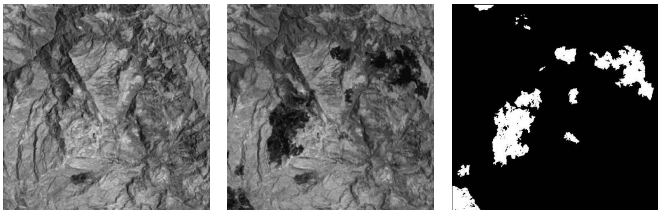
The basic principle of SCCN is to minimize the pixelwise difference that the unchanged regions present in two heterogeneous images captured on different dates. Then, by direct comparison of the transformed feature maps, the changed regions are highlighted. Unsupervised pretraining is also important, which facilitates the backpropagation algorithm and learns a useful representation of input images to form the consistent feature space. Usually, the local intensity distribution reflects the physical properties of ground objects while is influenced by noise in remote sensing images. Via DAE-based pretraining, the impact of noise is reduced and robust features are extracted. Then, unchanged areas present in heterogeneous images usually achieve similar local intensity distribution in the feature space after pretraining. Therefore, some unchanged areas show larger correlations and are easier to be aligned by the network, which can induce the method to find correct unchanged pixels. Then, P_u can be correctly initialized at first

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b) (c)

Fig. 3. Farmland data set. (a) SAR image acquired in 2008. (b) SAR image acquired in 2009. (c) Reference image.



(a) (b) (c)

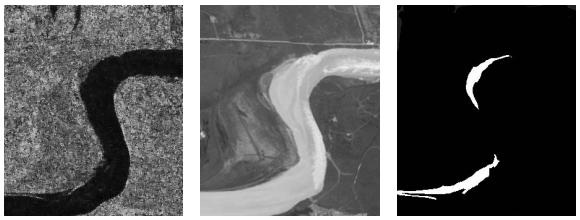
Fig. 4. Mexico data set. (a) Optical image acquired in 2000. (b) Optical image acquired in 2002. (c) Reference image.

iteration though it is randomly generated. As a consequence, by minimizing the coupling function, SCCN is expected to highlight the correct changed regions. Meanwhile, unsupervised learning forms the feature space that can well represent the two images.

In SCCN, there exist some user-defined parameters that may influence the performance of change detection. For example, the parameter λ influences the number of detected unchanged pixels for learning network parameters. Furthermore, the number of network layers and the proportion of the changed regions in the input images may also impact the performance. Moreover, the performance of SCCN may be sensitive to the random initialization of P_u.

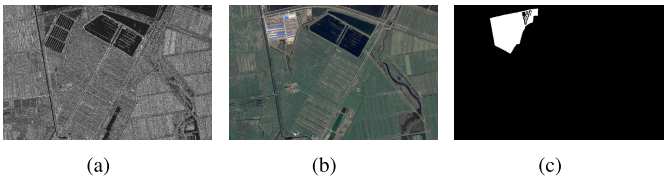
IV. Experimental Study

In this section, we evaluate the effectiveness of SCCN and investigate the factors that may influence the performance of SCCN based on one set of SAR images, one set of optical images, and two sets of heterogeneous images (i.e., SAR and optical images). We also compare the proposed SCCN with some existing change detection techniques to demonstrate its superiority.



(a) (b) (c)

Fig. 5. Yellow River data set. (a) SAR image acquired in 2008. (b) Optical image acquired in 2010. (c) Reference image.



(a) (b) (c)

Fig. 6. Shuguang Village data set. (a) SAR image acquired in 2008. (b) Optical image acquired in 2012. (c) Reference image.

A. Data Sets

The first data set consists of two SAR images with the same size of 306×291 pixels, which were acquired for the same region on different dates, as shown in Fig. 3(a) and (b). These two SAR images captured by Radarsat-2 (Canadian satellite) in June 2008 and June 2009, respectively, cover the same piece of the farmland along the Yellow River in eastern China. The changed regions correspond to the corrupted farmland as shown in Fig. 3(c), which is the reference image indicating the actual changed regions. This reference image is obtained manually by integrating some prior information with image interpretation based on the input images. Note that the two SAR images are single-look and four-look, respectively, which increases the difficulty of change detection.

The second data set consists of two optical images acquired by Landsat-7 (US satellite) at urban Mexico in April 2000 and May 2002, respectively. These two images are extracted from Band 4 of the ETM+ images. The sizes of both images are 512×512 pixels. This data set shows the vegetation damage after the forest fire at urban Mexico. Fig. 4(a)–(c) shows the two optical images and the reference image, respectively.

The third data set consists of one SAR image and one optical image with the same size of 291×343 pixels, as shown in Fig. 5(a) and (b), respectively. The SAR image was acquired by Radarsat-2 at the Yellow River Estuary in June 2008. The optical image captured in September 2010 was obtained from Google Earth, which covers the same region as the SAR image does. Google Earth provides a large amount of VHR image data of the surface of the earth. The data provided by Google Earth integrate the imagery from both satellite and aerial photography. The satellite images are mainly from QuickBird (US VHR satellite) and Landsat-7. The aerial images are from the companies of BlueSky, Sanborn, IKONOS, and SPOT5. One can acquire VHR optical images of most of the surface for the earth for research. This data set is used to study the change of the Yellow River bank caused by the flood. Fig. 5(c) shows the reference image that illustrates the actual changed regions.

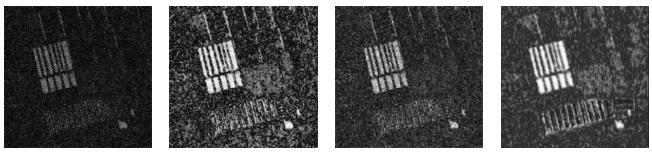
The fourth data set is composed of one SAR image and one RGB optical image, as shown in Fig. 6(a) and (b), respectively. This data set covers a piece of the farmland in the Shuguang Village of the Dongying City in China. The changed region corresponds to the new buildings built on the farmland, as shown in Fig. 6(c). The SAR and optical images have the same size of 921×593 pixels and were acquired in June 2008 and September 2012, respectively.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

B. Experimental Settings

1) Methods in Comparison: Change detection based on homogeneous images has been widely studied. There are many excellent methods detecting changes with high accuracy. Although the proposed method is mainly for heterogeneous images, the experiments on homogeneous images are to evaluate it with some well-developed change detection methods. For the first and second data sets composed of homogeneous images, we compare our method with the wavelet fusion-based method [28] and the deep neural network-based method [38]. The wavelet fusion-based method generates the difference map by fusing two complementary difference maps. The two difference maps are combined at the wavelet domain, which is a multiscale analysis method. The fused difference map complements each of the two difference maps and thus outperforms both of them. The deep neural network-based method takes as input the two images and outputs a binary map at the end of the network. This method reduces the influence of the difference map on the detection performance and outperforms many methods based on a difference analysis. For change detection based on heterogeneous images, not many existing techniques can be found, especially unsupervised techniques. Since PCC is a widely used framework for change detection, we compare our method with PCC-based method for the third and fourth data sets consisting of heterogeneous images. Because our proposed method is unsupervised, for fair comparison, we use the unsupervised image segmentation algorithm [56] to classify the two images. In our experiments, we set the sizes of all convolution kernels to 3×3 and specify 3 coupling layers for each side of the network with the same number of 20 feature maps at each hidden layer, i.e., $n_1 = n_2 = n_3 = 20$. Here, we employ the same number of hidden units per layer based on [42] to reduce the vast number of possible choices. Then, we use the same number of feature maps. We set $\lambda = 0.1$ for data sets composed of homogeneous images and $\lambda = 0.15$ for those consisting of heterogeneous images, because the coupling error for homogeneous images is intrinsically lower than that for heterogeneous images.

2) Evaluation Criteria: We represent the change detection result in the form of a binary map in which white pixels denote changed regions and black pixels for unchanged regions. The quantitative analysis is made to demonstrate the performance of our proposed method. First, false negative (FN) and false positive (FP) are calculated. FN is the number of pixels that are classified as unchanged ones but actually changed ones as indicated in the reference image. FP is the number of pixels that are unchanged ones in the reference image but wrongly classified as changed ones. Second, we compute the overall error (OE) and the classification accuracy (CA) by $\mathrm{OE} = \mathrm{FN} + \mathrm{FP}$ and $\mathrm{CA} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})$, where TN denotes true negative which is the number of pixels correctly classified as unchanged ones and TP for true positive which is the number of pixels correctly classified as changed ones. Finally, Kappa coefficient (KC) [57] is computed to measure the detection performance. KC is typically used to evaluate unsupervised image segmentation results. The higher the value of KC is, the higher the segmentation accuracy is.



(a) (b) (c) (d)

Fig. 7. Difference maps for the farmland data set generated by different methods. (a) Log ratio. (b) Mean ratio. (c) Wavelet fusion. (d) SCCN.

Specially, KC is calculated by

$$\mathrm{KC} = \frac{\mathrm{CA} - \mathrm{PRE}}{1 - \mathrm{PRE}} \tag{12}$$

where

$$\mathrm{PRE} = \frac{(\mathrm{TP} + \mathrm{FP}) \cdot (\mathrm{TP} + \mathrm{FN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2} + \frac{(\mathrm{FN} + \mathrm{TN}) \cdot (\mathrm{FP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2}. \tag{13}$$

We evaluate the quality of the difference map by using the receiver operating characteristics (ROCs) plot. The ROC plot can be used to evaluate the intrinsic quality of change detection independent of the threshold choice. It is drawn via two variables, i.e., true positive rate (TPR) and FP rate (FPR). Given a threshold, we compute TPR and FPR based on the difference map. TPR and FPR are calculated by

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$
$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}. \tag{14}$$

Then, a set of TPR and FPR values are obtained by using all the possible threshold values (i.e., $1, 2, \ldots, 255$). The ROC plot for a high-quality difference map should be close to the top-left corner of the coordinate system. The area under the ROC curve (AUC) provides a numerical measure. For a perfect difference map, AUC is equal to 1. A larger AUC value typically indicates higher quality of the difference map.

The following discusses the experimental results for each data set. Due to the use of random initialization, for SCCN, the reported values of evaluation criteria are the average values over 30 independent runs.

C. Experiments on the Farmland Data Set

A major difficulty of change detection based on SAR images is the influence of speckle noise. Usually, the difference map is calculated by using the log-ratio or mean-ratio operator to suppress speckle noise. A wavelet fusion-based method was proposed to fuse two difference maps [28]. First, we illustrate the difference maps obtained by the log-ratio operator, mean-ratio operator, wavelet fusion, and SCCN in Fig. 7. Since the deep neural network-based method directly outputs a binary classification map, no different map is generated by this method. The log-ratio operator transforms the multiplicative noise into additive one while suppressing the changing information. According to the reference image, the difference map obtained by the log ratio mainly shows unchanged

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
EVALUATION OF CHANGE DETECTION RESULTS ON THE FARMLAND DATA SET OBTAINED BY DIFFERENT METHODS

methods	AUC	FP	FN	OE	CA	KC
log-ratio	0.9164	2743	1989	4732	0.9469	0.5528
mean-ratio	0.9660	1051	1312	2363	0.9735	0.7560
wavelet fusion	0.9739	931	1377	2308	0.9741	0.7576
deep neural network	–	644	710	1354	0.9848	0.8627
SCCN	0.9916	768	779	1547	0.9826	0.8438

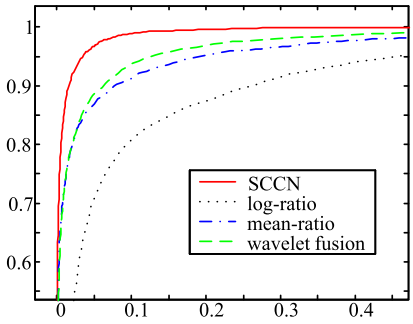


Fig. 8. ROC plots of the four difference maps for the farmland data set.

regions without clearly highlighting changed. The mean-ratio operator uses local information and thus is robust to speckle noise. However, the background of thus obtained difference map is quite obscure, as shown in Fig. 7. SCCN suppresses speckle noise via the convolutional layer and the DAE-based pretraining. Although the pixelwise differences with respect to unchanged regions are not well reduced, changed regions are clearly highlighted. We draw the ROC plot for each difference map and show them in Fig. 8. The quality of difference maps is clearly exhibited in the ROC plot. The ROC plot for SCCN is closest to the top-left corner of the coordinate system, which demonstrates the superiority of SCCN. For SAR images, the DAE-based pretraining plays a crucial role, while the coupling function also plays the role of implicitly aligning two input images.

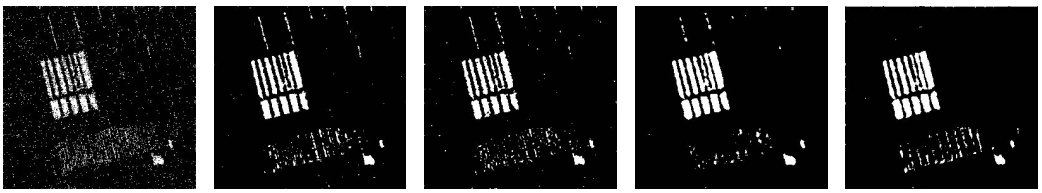
With a difference map, the change detection result can be obtained by using a thresholding algorithm [24]. The detection results based on the difference maps obtained by five methods are shown in Fig. 9. The log-ratio method cannot get rid of noise, because it just transforms noise. Accordingly, there exist many spots in its change detection result. The mean ratio, wavelet fusion, SCCN, and deep neural network-based methods can all reduce the impact of noise. Table I reports the quality analysis results based on six evaluation criteria as described in Section IV-B2. It can be observed that SCCN performed best among the four difference map-based methods. Notably, without using the difference map and thresholding-based method, the deep neural network-based method achieved the overall outstanding detection performance.

D. Experiments on the Mexico Data Set

Change detection based on optical images is easier than that based on SAR images, because noise in optical images is relatively easier to handle. For optical images, the substraction operator and the ratio operator are usually used to generate the difference map. Fig. 10 shows the difference maps generated by different methods. The substraction operator-based difference map can reveal unchanged regions well. However, changed regions are less obviously highlighted. In contrast, the difference map generated via the ratio operator highlights changed regions but not accurately reveals unchanged regions. The wavelet fusion-based method somewhat makes up the defects of these two operators to generate a better difference map. SCCN relies on a fundamentally different principle to generate the difference map. The ROC plots for the above-mentioned four difference maps are shown in Fig. 11. It can be observed that SCCN outperforms the other three methods in comparison. Finally, a thresholding algorithm is applied on the difference maps to generate the change detection results, as shown in Fig. 12. In the change detection map obtained via the substraction operator, some details in both changed and unchanged regions are missed. Some of such missing details in changed regions are recovered in the change detection map obtained by the ratio operator at the expense of more details missed in unchanged regions. In comparison, the wavelet fusion-based method yields the detection result that balances the missing details in changed and unchanged regions. Since the deep neural network-based method was mainly designed for dealing with speckle noise in SAR images, its performance degrades when dealing with optical images. Table II reports the values of evaluation criteria. It can be observed that SCCN outperforms the other compared methods.

The above two experiments based on homogeneous images demonstrate the effectiveness of SCCN. We find that the proposed SCCN cannot significantly outperform the compared methods. In fact, the experiments based on homogeneous images only aim to demonstrate that the proposed SCCN is applicable to homogenous images besides its prominent utility on heterogeneous images. Therefore, we do not expect that the proposed SCCN can significantly outperform those dedicated methods in comparison. Given the fact that homogenous images-based change detection has been widely studied but research on change detection based on heterogeneous images is relatively rare (and most existing works are supervised), the obvious advantage of our proposed method is its efficacy of detecting changes based on heterogeneous images in an unsupervised manner. In the following, we will evaluate SCCN based on heterogeneous images.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b) (c) (d) (e)

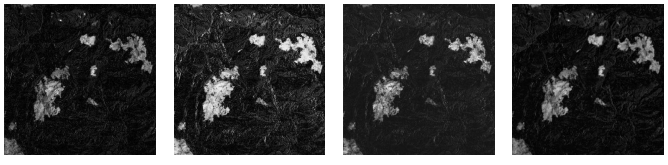
Fig. 9. Change detection maps for the farmland data set generated by different methods. (a) Log ratio. (b) Mean ratio. (c) Wavelet fusion. (d) Deep neural network. (e) SCCN.

TABLE II
EVALUATION OF CHANGE DETECTION RESULTS ON THE MEXICO DATA SET OBTAINED BY DIFFERENT METHODS

methods	AUC	FP	FN	OE	CA	KC
substraction	0.9869	2585	3056	5641	0.9785	0.8768
ratio	0.9874	3049	3199	6248	0.9761	0.8644
wavelet fusion	0.9899	2246	3311	5557	0.9788	0.8774
deep neural network	–	696	4266	4962	0.9811	0.8854
SCCN	0.9914	2290	2278	4568	0.9826	0.9011

TABLE III
EVALUATION OF CHANGE DETECTION RESULTS ON THE YELLOW RIVER DATA SET OBTAINED BY DIFFERENT METHODS

methods	AUC	FP	FN	OE	CA	KC
PCC	–	2863	1017	3880	0.9611	0.5064
SCCN without pre-training	0.9505	3494	642	4136	0.9586	0.5301
SCCN with 1 coupling layer	0.9515	2903	620	3523	0.9647	0.5513
SCCN with 2 coupling layers	0.9645	1882	1169	3051	0.9694	0.5740
SCCN with 3 coupling layers	0.9688	1060	1235	2295	0.9770	0.6154



(a) (b) (c) (d)

Fig. 10. Difference maps for the Mexico data set generated by different methods. (a) Substraction. (b) Ratio. (c) Wavelet fusion. (d) SCCN.

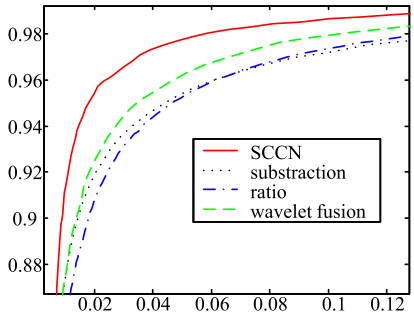


Fig. 11. ROC plots of the four difference maps for the Mexico data set.

E. Experiments on the Yellow River Data Set

This data set contains two heterogeneous images (one SAR image and one optical image). We first apply PCC to yield the classification and change detection results, as shown in Fig. 13. There are two classes in these two images, i.e., land and river. The change detection result is obtained by comparing the classification results obtained for these two images in which pixels having different class labels in two classification maps are identified as changed pixels. PCC is intuitive and simple which can be applied to various types of change detection tasks. The results obtained by SCCN are shown in Fig. 14. To comprehensively evaluate SCCN, we implement several SCCN variants, i.e., SCCN without pretraining, SCCN with one coupling layer, SCCN with two coupling layers, and SCCN with three coupling layers in each side, respectively. The difference maps obtained by these SCCN variants are shown in Fig. 14(a)–(d). We specify one coupling layer for SCCN without pretraining, because it is difficult to train an SCCN with multiple coupling layers only via backpropagation. The change detection results are obtained by using a thresholding algorithm and shown in Fig. 14(e)–(h). Table III reports the values of evaluation criteria. The comparisons in Fig. 14 and Table III reveal that pretraining is an indispensable step when training SCCN.

PCC is able to detect changes but easy to introduce accumulated errors from the inaccurate classification results. In comparison, SCCN achieved the superior performance on the Yellow River data Set. In the following, we further evaluate SCCN on a more complex data set.

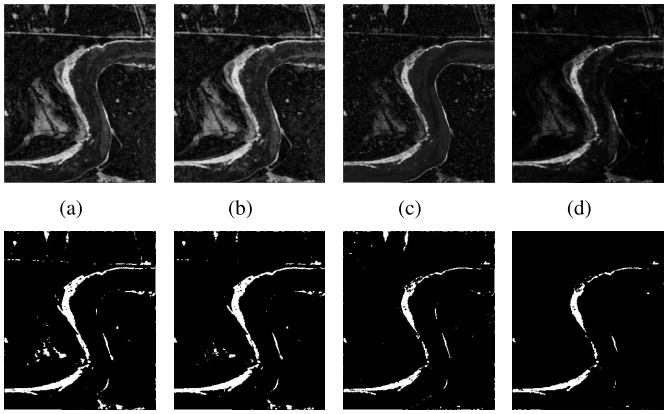
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

(a) (b) (c) (d) (e)

Fig. 12. Change detection maps for the Mexico data set generated by different methods. (a) Substraction. (b) Ratio. (c) Wavelet fusion. (d) Deep neural network. (e) SCCN.

(a) (b) (c)

Fig. 13. Classification and change detection results for the Yellow River data set. (a) Classification result for the SAR image. (b) Classification result for the optical image. (c) Change detection result.



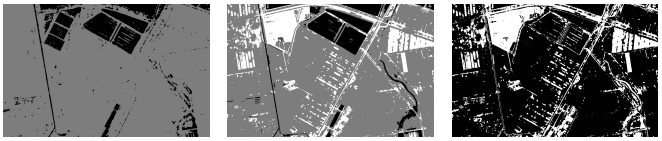
(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 14. Difference maps and change detection results obtained by SCCN on the Yellow River data set. (a) Difference map generated by SCCN without pretraining (and with one coupling layer). (b) Difference map generated by SCCN with one coupling layer. (c) Difference map generated by SCCN with two coupling layers. (d) Difference map generated by SCCN with three coupling layers. (e)–(h) Change detection results obtained from the difference maps [(a)–(d)], respectively, by using a thresholding algorithm.

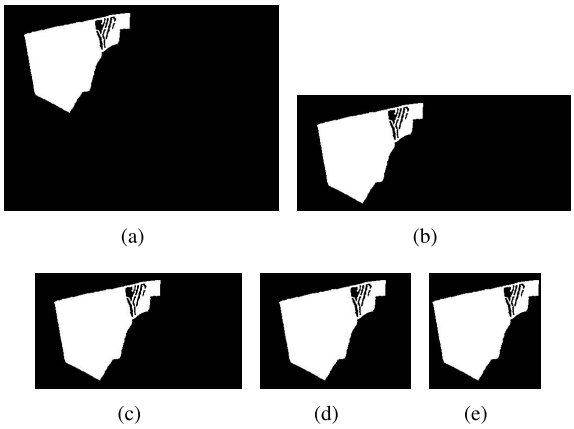
F. Experiments on the Shuguang Village Data Set

Two heterogeneous images contained in this data set are more complex than those contained in the Yellow River data set. There exist more ground objects and the optical image has three bands. First, we show the classification and change detection results in Fig. 15. There are two recognizable classes in the SAR image, i.e., farmland and water area. In fact, there also exist a few buildings which, however, are difficult to be recognized by using unsupervised classifiers. There are three recognizable classes in the optical image, i.e., farmland, water area, and building. However, some farmland regions are wrongly classified as building regions.



(a) (b) (c)

Fig. 15. Classification and change detection results for the Shuguang Village data set. (a) Classification result of the SAR image. (b) Classification result of the optical image. (c) Change detection result.



(a) (b)

(c) (d) (e)

Fig. 16. Changed regions of the subsets extracted from the Shuguang Village data set. (a) Subset A. (b) Subset B. (c) Subset C. (d) Subset D. (e) Subset E.

Accordingly, such errors are introduced into the change detection result. Because the coupling function aims to minimize the coupling errors for unchanged regions, the proportion of unchanged regions in the entire image may influence the performance. Therefore, we carry out image cropping to produce five subsets where unchanged regions take up different proportions. We show the reference images of these subsets in Fig. 16. The change detection results by applying SCCN on these subsets are shown in Fig. 17. Because the properties of some changed regions in the optical image are very similar to those of unchanged regions, some details in changed regions are missed, while unchanged regions are well preserved. With the decrease of unchanged regions, the details of changed regions are not lost. Table IV reports the values of evaluation criteria. It can be observed that the performance of SCCN decreases as the portion of unchanged regions shrinks, because less unchanged pixels are used to train the network. However, even with a small portion of unchanged regions, e.g., subset E, a satisfactory detection result was achieved.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
EVALUATION OF CHANGE DETECTION RESULTS ON THE YELLOW RIVER DATA SET OBTAINED BY DIFFERENT METHODS

methods		AUC	FP	FN	OE	CA	KC
PCC	Shuguang Village	–	97258	489	97747	0.8210	0.2569
SCCN	Shuguang Village	0.9593	5101	7800	12901	0.9764	0.6789
	subset A	0.9631	5207	6446	11653	0.9484	0.7019
	subset B	0.9571	5747	4435	10182	0.9197	0.7286
	subset C	0.9528	3675	5165	8840	0.9076	0.7348
	subset D	0.9525	2705	4776	7481	0.8928	0.7468
	subset E	0.9453	2851	3581	6432	0.8751	0.7445

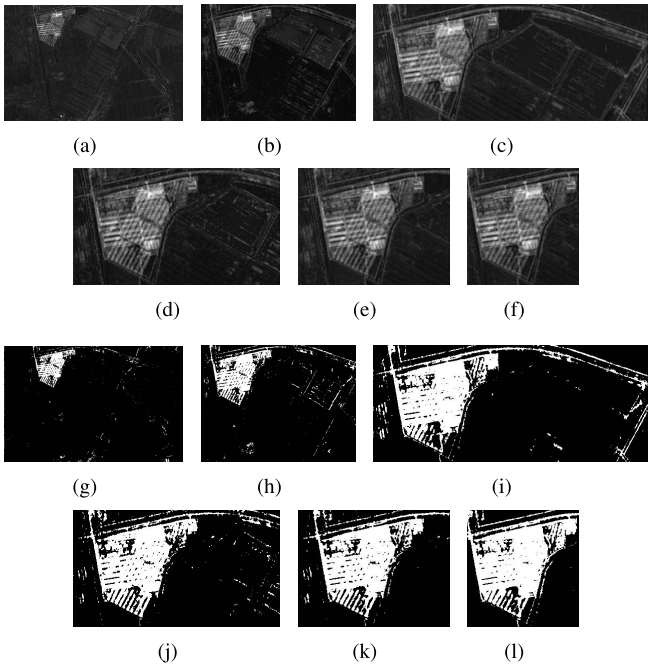


Fig. 17. Difference maps and change detection results of SCCN for several subsets of the Shuguang Village data set. (a) Difference map for the Shuguang Village data set. (b)–(f) Difference maps for subsets A–E, respectively. (h)–(l) Change detection results obtained by segmenting (a)–(f), respectively, using a thresholding algorithm.

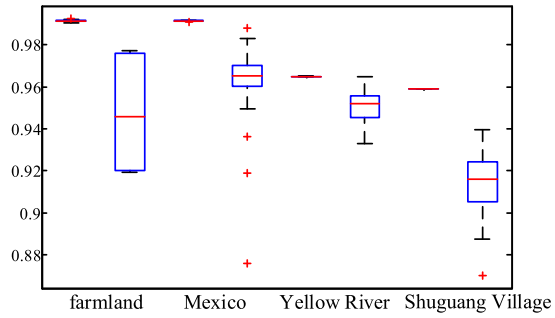


Fig. 18. Box plot of the AUC values on the four data sets. Left box is the result obtained by the method with pretraining and right box without pretraining.

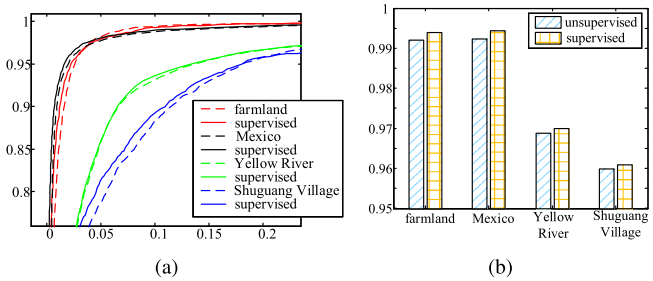


Fig. 19. Comparison between unsupervised and supervised ways of SCCN on the four data sets. (a) ROC plots. (b) Histograms of AUC values.

G. Experiments on P_u and λ

As discussed in Section III, the initialization of P_u may influence the performance and so as the choice of λ. Therefore, we make a sensitivity analysis on the initialization of P_u. Specially, we execute SCCN for 30 independent runs with each run using a randomly initialized P_u to obtain the value of AUC for each run. We draw the box plots of the obtained AUC values in Fig. 18. We also draw the box plots for SCCN without pretraining. It can be observed via the generally small variations that the detection performance of SCCN is less sensitive to the initialization of P_u. However, without pretraining, the performance of SCCN depends much more on the initialization. This demonstrates that pretraining helps to improve the robustness of SCCN to initialization. For SAR images, speckle noise is difficult to handle. Therefore, the variation of the performance on the farmland data set is larger than those of the other data sets.

SCCN is unsupervised in which the unchanged map, i.e., P_u is learned during the learning process as well. However, it is not clear whether P_u is correct. Therefore, we employ a supervised version of SCCN as the supervised change detection model where P_u in (7) is substituted by the ground-truth label map with 0 denoting the changed pixel and 1 unchanged. The supervised version of SCCN is trained in the same way as the unsupervised SCCN, although only θ needs to be learned based on the ground truth. The performance comparison is shown in Fig. 19, where Fig. 19(a) shows the ROC plots and Fig. 19(b) shows the AUC values. The ROC plots of supervised and unsupervised are almost coincident which demonstrates that the unsupervised P_u achieves approximate performance to that of real P_u.

We also make a sensitivity analysis on λ. Specially, we set λ as $0.01, 0.02, \ldots, 0.2$, respectively, and evaluate SCCN using different λ values on the four data sets. We plot the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

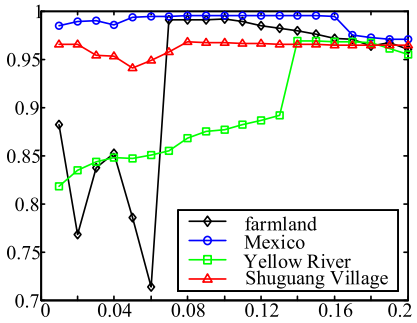


Fig. 20. Plot of the AUC values with respect to different λ values on the four data sets.

AUC values with respect to the λ values in Fig. 20 where the horizontal axis represents λ and the vertical axis denotes AUC. It can be obvious that the performance is not consistent under different λ values, especially when the value of λ is too small or too large. When λ is too small, very few unchange pixels can be derived to train the network. When λ is too large, many changed pixels are forced as unchanged ones, which deteriorates the detection performance. However, the values of λ corresponding to good performances constitute a wide range. In other words, it is easy to choose the value of λ. Furthermore, for homogeneous images, their image properties are intrinsically similar and accordingly the coupling error is small. Therefore, a relatively small λ may lead to the good performance. In comparison, heterogeneous images usually have very different properties, which results in large coupling error. In such a case, an appropriate value of λ is relatively large.

V. Conclusion and Future Work

We proposed a novel SCCN for change detection based on heterogeneous SAR and optical images. Change detection techniques have been investigated for decades. Recently, there shows an increasing research interest in change detection based on heterogeneous images. Different from most existing methods that rely on labeled pixels to learn the latent relation between two heterogeneous images, our proposed SCCN is fully unsupervised without using any labeled pixels. Each of the two sides of SCCN is composed of one convolutional layer and several coupling layers, which transform the two input images (fed to each side) into a feature space in which these two input images have more consistent feature representations. Eventually, the difference map is directly calculated via pixelwise Euclidean distances in this feature space. A coupling function is defined to drive the learning of network parameters. To give a good initialization to both network parameters and unchanged labels, we pretrain the network layer by layer through DAE by taking into account the noise models of the two input images.

Experimental results on both homogeneous images and heterogeneous images verified the effectiveness of SCCN and also demonstrated the superiority of SCCN over several existing approaches. In this paper, we only consider the unchanged pixels and only two images are used to detect changes. Our future work includes but is not limited to the detection of subtle changes, investigation of joint distribution of the changed and unchanged pixels, and the extension of SCCN to handle more than two heterogeneous images of different types.

References

[1] J. Gong, S. Haigang, M. Guorui, and Z. Qiming, "A review of multi-temporal remote sensing data change detection algorithms," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 37, no. B7, pp. 757–762, 2008.

[2] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[3] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[4] C. Tison, J. M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.

[5] K. Mubea and G. Menz, "Monitoring land-use change in Nakuru (Kenya) using multi-sensor satellite data," *Adv. Remote Sens.*, vol. 1, no. 3, 2012, Art. no. 26112.

[6] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.

[7] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, Jul. 2014.

[8] 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.

[9] A. Robin, L. Moisan, and S. L. Hegarat-Mascle, "An a-contrario approach for subpixel change detection in satellite imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1977–1993, Nov. 2010.

[10] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.

[11] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2963–2972, Dec. 2005.

[12] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[13] G. Quin, B. Pinel-Puyssegur, J.-M. Nicolas, and P. Loreaux, "MIMOSA: An automatic change detection method for SAR time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5349–5363, Sep. 2014.

[14] F. Chatelain, J.-Y. Tourneret, and J. Inglada, "Change detection in multisensor SAR images using bivariate Gamma distributions," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 249–258, Mar. 2008.

[15] H. Li, M. Gong, Q. Wang, J. Liu, and L. Su, "A multiobjective fuzzy clustering method for change detection in SAR images," *Appl. Soft Comput.*, vol. 46, pp. 767–777, Nov. 2015.

[16] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.

[17] I. Chaturvedi, Y.-S. Ong, and R. V. Arumugam, "Deep transfer learning for classification of time-delayed Gaussian networks," *Signal Process.*, vol. 110, pp. 250–262, May 2015.

[18] 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.

[19] L. Bruzzone and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.

[20] F. Bujor, E. Trouvé, L. Valet, J.-M. Nicolas, and J.-P. Rudant, "Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2073–2084, Oct. 2004.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[21] S. Marchesi, F. Bovolo, and L. Bruzzone, "A context-sensitive technique robust to registration noise for change detection in VHR multispectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1877–1889, Jul. 2010.

[22] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 307–311, Mar. 2012.

[23] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.

[24] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, Oct. 2006.

[25] Y. Bazi, F. Melgani, and H. D. Al-Sharari, "Unsupervised change detection in multispectral remotely sensed imagery with level set methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3178–3187, Aug. 2010.

[26] G. Pajares, "A hopfield neural network for image change detection," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1250–1264, Sep. 2006.

[27] A. Ghosh, B. N. Subudhi, and L. Bruzzone, "Integration of gibbs Markov random field and hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3087–3096, Aug. 2013.

[28] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.

[29] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98–109, Feb. 2014.

[30] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.

[31] Z.-G. Liu, G. Mercier, J. Dezert, and Q. Pan, "Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 168–172, Jan. 2014.

[32] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.

[33] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 1513–1517.

[34] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multitemporal cross-sensor images with automated kernel canonical correlation analysis," *J. Photogram. Remote Sens.*, vol. 107, pp. 50–63, Sep. 2015.

[35] J. R. Jensen, E. W. Ramsey, H. E. Mackey, Jr., E. J. Christensen, and R. R. Sharitz, "Inland wetland change detection using aircraft MSS data," *Photogram. Eng. Remote Sens.*, vol. 53, no. 5, pp. 521–529, 1987.

[36] Y. Qin, Z. Niu, F. Chen, B. Li, and Y. Ban, "Object-based land cover change detection for cross-sensor images," *Int. J. Remote Sens.*, vol. 34, no. 19, pp. 6723–6737, 2013.

[37] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observat. Geoinf.*, vol. 20, pp. 77–85, Feb. 2013.

[38] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2015.

[39] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[40] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[41] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[42] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[43] S. Rifai, P. Vincent, X. Müller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. 28th Int. Conf. Mach. Learn.*, Bellevue, WA, USA, Jun. 2011, pp. 833–840.

[44] H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area V2," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2007, pp. 873–880.

[45] N.-N. Ji, J.-S. Zhang, and C.-X. Zhang, "A sparse-response deep belief network based on rate distortion theory," *Pattern Recognit.*, vol. 47, no. 9, pp. 3179–3191, 2014.

[46] M. Gong, J. Liu, H. Li, Q. Cai, and L. Su, "A multiobjective sparse feature learning model for deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3263–3277, Dec. 2015.

[47] Y. Yuan, L. Mou, and X. Lu, "Scene recognition by manifold regularized deep learning architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2222–2233, Oct. 2015.

[48] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.

[49] H. Goh, N. Thome, M. Cord, and J.-H. Lim, "Learning deep hierarchical visual feature coding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2212–2225, Dec. 2014.

[50] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.

[51] A. S. Chandar, M. M. Khapra, B. Ravindran, V. Raykar, and A. Saha, "Multilingual deep learning," in *Proc. NIPS*, 2013, pp. 1–10.

[52] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

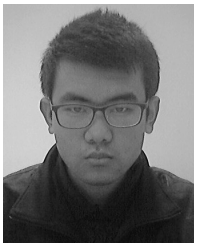
[53] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2007, pp. 153–160.

[54] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory Algorithms*. Hoboken, NJ, USA: Wiley, 2013.

[55] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, Helsinki, Finland, Jun. 2008, pp. 1096–1103.

[56] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy C-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573–584, Feb. 2013.

[57] R. L. Brennan and D. J. Prediger, "Coefficient kappa: Some uses, misuses, and alternatives," *Edu. Psychol. Meas.*, vol. 41, no. 3, pp. 687–699, 1981.



Jia Liu received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2013, where he is currently pursuing the Ph.D. degree.

His current research interests include computational intelligence and image understanding.

Maoguo Gong (M'07–SM'14) received the B.S. degree (Hons.) in electronic engineering and the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2003 and 2009, respectively.

Since 2006, he has been a Teacher with Xidian University. In 2008 and 2010, he was promoted to Associate Professor and Full Professor, with exceptive admission. His current research interests include computational intelligence with applications to optimization, learning, data mining, and image understanding.

Dr. Gong was a recipient of prestigious national program for the support of Top-Notch Young Professionals from the Central Organization Department of China, the Excellent Young Scientist Foundation from the National Natural Science Foundation of China, and the New Century Excellent Talent in University from the Ministry of Education of China.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

K. Qin (S'06–M'07–SM'12) received the Ph.D. degree from Nanyang Technological University, Singapore, in 2007.

From 2007 to 2009, he was a Post-Doctoral Fellow with the University of Waterloo, ON, Canada. He was with French National Institute for Research in Computer Science and Control, France, as a Post-Doctoral Researcher and an Expert Engineer. He is currently a Lecturer with RMIT University, Melbourne, VIC, Australia. His current research interests include evolutionary computation, machine learning, image processing, and massively parallel computing.

Dr. Qin co-authored journal papers are the most cited ones (Web of Science) among papers published in the 2006 and 2009 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION (TEVC), respectively. The 2009 paper describing a self-adaptive differential evolution algorithm won the 2012 IEEE TEVC Outstanding Paper Award. His conference papers describing an improved GPU-based differential evolution had been nominated for the best paper award at the 2012 Genetic and Evolutionary Computation Conference. He is currently chairing the IEEE Computational Intelligence Society task force on collaborative learning and optimization.

Puzhao Zhang received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2013, where he is currently pursuing the Ph.D. degree.

His current research interests include computational intelligence and image understanding.