## GENERAL INSTRUCTIONS

- **Authors: Carefully check the page proofs (and coordinate with all authors); additional changes or updates WILL NOT be accepted after the article is published online/print in its final form. Please check author names and affiliations, funding, as well as the overall article for any errors prior to sending in your author proof corrections.**
- **Authors: Please check ALL author names for correct spelling, abbreviations, and order of first and last name in the byline, affiliation footnote, and bios.**
- **Authors: We cannot accept new source files as corrections for your article. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.**
- **Authors: Please note that once you upload your changes, they will be entered and your article finalized. The proofing process is now complete and your article will be sent for final publication and printing. If you would like an additional proof to review, this should be noted as it is not IEEE policy to send multiple proofs. Once your article is posted on Xplore, it is considered final and the article of record. No further changes will be allowed at this point so please ensure scrutiny of your final proof.**

## QUERIES

Q1. Author: Please confirm or add details for any funding or financial support for the research of this article.
Q2. Author: Please provide the subject in which author Peijun Du received the Ph.D. degree.
Q3. Author: Please provide the year in which author Dongmei Chen received the B.A. and master's degree.

# Change Detection Based on Low-Level to High-Level Features Integration With Limited Samples

Xin Wang, Peijun Du, *Senior Member, IEEE*, Dongmei Chen, Sicong Liu, *Member, IEEE*, Wei Zhang, and Erzhu Li

***Abstract*—Detailed land cover change in multitemporal images is an important application for earth science. Many techniques have been proposed to solve this problem in different ways. However, accurately identifying changes still remains a challenge due to the difficulties in describing the characteristics of various change categories by using single-level features. In this article, a multilevel feature representation framework was designed to build robust feature set for complex change detection task. First, four different levels of information from low level to high level, including pixel-level, neighborhood-level, object-level, and scene-level features, were extracted. Through the operation of extracting different level features from multitemporal images, the differences between them can be described comprehensively. Second, multilevel features were fused to reduce the dimension and then used as the input for supervised change detector with initial limited labels. Finally, for reducing the labeling cost and improving the change detection results simultaneously, active learning was conducted to select the most informative samples for labeling, and this step together with the previous steps were iteratively conducted to improve the results in each round. Experimental results of three pairs of real remote sensing datasets demonstrated that the proposed framework outperformed the other state-of-the-art methods in terms of accuracy. Moreover, the influences of scene scale for high-level semantic features in the proposed approach on change detection performance were also analyzed and discussed.**

***Index Terms*—Active learning, attribute profiles (APs), change detection, convolutional neural network (CNN), multilevel feature, object feature, scene feature.**

## I. Introduction

LAND use and land cover change is one of the most important components of Earth science under the condition of frequent interaction between humans and the natural system. Earth observation from remote sensing satellites provides a great opportunity for monitoring the land surface dynamic changes in wide geographical areas compared with traditional *in-situ* investigation, which is difficult and time-consuming. Nowadays, both long-term (e.g., yearly) and short-term (e.g., daily) satellite observations produce large amounts of multitemporal images in data archives. Therefore, automatic techniques are required to effectively discover, describe, and detect changes that have occurred in multitemporal images. Change detection is a process of finding changes or phenomenon by observing images at different times [1]. Various change detection techniques through multitemporal remote sensing images have been proposed [2]–[5]. They are widely used in different applications, including natural (e.g., wildfires and glacial retreat) and anthropogenic disturbances (e.g., deforestation and urbanization) [6]–[9].

Change detection can be broadly categorized into unsupervised or supervised methods [2]. The unsupervised methods perform a direct comparison of multitemporal images acquired on different dates [10]. These techniques do not rely on prior knowledge of research areas and are suitable for some sudden change applications, such as the monitoring of landslide [11], deforestation [12], and burned areas [13]. However, the results of unsupervised methods are easily affected by some external factors, including illumination variations, changes of atmospheric conditions, and poor sensor calibration, which normally occur at different acquisition times. In contrast, the supervised techniques demonstrate the advantages of robustness in dealing with different image acquisition conditions [14]. For example, postclassification comparison techniques obtain results from the classified maps at different dates. As long as the quality of ground truth samples is high enough, it can achieve the satisfactory change detection results with transition categories. Due to its low requirements for the consistency between different images, it has been applied to change detection between multisource images [15], [16] and time series land cover change analysis [17]. However, the misclassification of any temporal image will affect the final change detection due to the error propagation mechanism in this method. Moreover, preparing training samples for each temporal image is a time consuming and expensive process. Therefore, another supervised change detection techniques based on stack or difference of multitemporal images and supervised change detector such as SVM were proposed [18]. These methods take the information of multitemporal images into account and only need the samples related

Manuscript received June 25, 2020; revised August 17, 2020; accepted September 28, 2020. This work was supported in part by the Natural Science Foundation of China under Grant 41631176 and in part by the Natural Sciences and Engineering Research Council of Canada. *(Corresponding authors: Peijun Du; Dongmei Chen.)*

Xin Wang, Peijun Du, and Wei Zhang are with the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China (e-mail: wangxrs@126.com; dupjrs@gmail.com; zhangwrs@163.com).

Dongmei Chen is with the Department of Geography and Planning, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: chendm@queensu.ca).

Sicong Liu is with the College of Surveying and Geoinformatics, Tongji University, Shanghai 200092, China (e-mail: sicong.liu@tongji.edu.cn).

Erzhu Li is with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China (e-mail: liezrs2018@jsnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3029460

to change information once. Furthermore, some extended work has been proposed by introducing semi-supervised methods for reducing the tedious workload of labeling [19], [20]. In a word, supervised change detection methods can relax the strict prerequisite of radiometric consistency in remote sensing images and provide the superiority of discrimination of change category, which reflects the exact "from-to" information of changes [18].

Most traditional change detection methods are only focused on spectral signature changes in each individual pixel [21], [22], while the geometrical characteristics of change targets, especially for high-resolution images, are not fully modeled and preserved. This may increase the ambiguity due to abnormal spectral variations in isolated pixels and errors (e.g., coregistration errors), leading to more omission and commission errors in change detection. Therefore, some methods based on a local neighborhood were proposed by using spatial features with contextual information, including grey-level co-occurrence matrix (GLCM), morphological profile (MP), and morphological attribute profile (AP), and have achieved better results [18], [23], [24]. However, these methods can underperform due to poor feature representation of spectral reflectance and spatial features. They are also unable to obtain a smooth border close to the reality of ground truth. Hence, object-based change detection (OBCD) approaches, which utilize a new base unit to process multitemporal images and construct feature representation of targets, were proposed to address these problems [25], [26]. To this end, object-based information including spectral reflectance, texture, and shape of individual objects can be employed to change detection tasks [27]. These methods provide effective schemes to use different types of information to construct discriminative feature representations for change targets [28]. Nonetheless, it is often difficult to define the parameters that are suitable for all the objects in the entire image. Although OBCD and contextual-based methods take the spatial characteristics into account and can better deal with complex surface structures, the low- and medium-resolution images are mostly mixed pixels. Therefore, the change information of some pixels will not be fully expressed.

Deep learning algorithms, especially convolutional neural network (CNN), have drastically improved performance in understanding and identifying changes and their types from remotely sensed images [29], [30]. CNNs are capable of extracting high-level semantic features within scenes and have already been applied to various image processing tasks, including semantic labeling [31], [32], image classification [33]–[36], and target detection [37]. Since CNNs are able to capture rich information from objects or image parts, studies have exploited deep learning for change detection [38], [39]. However, most of them deal with binary change detection only, and are usually trained to accept RGB input images, whereas some satellite images are provided with near-infrared and other channels (e.g., Landsat and Sentinel-2) that are also important for change detection, especially for vegetation analysis [40]. In the application of using CNN, a large number of training samples are often required to train the network. However, it is extremely difficult to obtain such amount of samples containing change information of multitemporal images.

In summary, change detection tasks based on single-level features have faced challenges as follows.

1) For different resolution images and change detection requirements, a certain level of features is not sufficient to fully characterize the change information for multitemporal images.
2) Some features are sensitive to parameters such as size of the morphological filter or segmentation scale.
3) The identification of change class relies too much on a large number of training samples.

Therefore, a novel framework of employing different-level information to address the complex change detection problems using multitemporal remotely sensed images is proposed in this article. In this framework, low-level to high-level features, including spectral reflectance, APs, object features, and deep semantic features, are extracted and combined to characterize difference information of multitemporal images from different aspects. These features are then fused at pixel level for change detection. The active learning algorithm is also integrated for selecting the most informative training samples to iteratively optimize the extracted features and change detector in an efficient way.

Different from the existing approaches, the main innovative contributions of this article can be summarized as follows.

1) First of all, this article first integrates different hierarchical features to comprehensively describe the change characteristics of the multitemporal images. The difference information can therefore be highlighted to distinguish the changes and their categories.
2) Second, the approach introduces active learning, which cannot only minimize the labeling cost, but also optimize the scene-level feature and change detector model simultaneously to iteratively improve the change detection results.

Three case studies using datasets with spatial resolutions from medium to high (Landsat 5, Sentinel-2, and UAV) were conducted and validated the effectiveness and universality of the proposed approach in multispectral image change detection. The remainder of this article is organized as follows. Section II reviews the state-of-the-art features used in change detection tasks and introduces their concepts. Section III illustrates the detailed theory and strategy of the proposed approach. Section IV presents the experimental results of the case studies. Finally, Sections V and VI contain discussion and conclusion, respectively.

## II. Related Techniques

### A. Morphological Attribute Profile

The APs, which execute the filtering operations on the max-tree representation of the analyzed image, have proven to be effective in extracting informative spatial features and geometrical structures in various remote sensing applications [41], such as land cover classification [42]–[44], target detection [45], [46], and change detection [23], [47], [48]. The filtering techniques are provided with the ability of attenuate the slight details and preserving the important characteristics of the regions simultaneously through opening and thinning operators by adjusting

the criterion of the different attributes [49]. In detail, the filtering operation implemented is based on the evaluation of how a given attribute $A$ is computed for every connected component of a grayscale image $f$ for a given reference value $\lambda$. For a connected component of the image $C_i$, the region is kept unaltered if the attribute meets a predefined condition [e.g., $A(C_i) > \lambda$]; otherwise, it is set to the grayscale value of the adjacent region with closer value, thereby merging $C_i$ to a surrounding connected component. When the region is merged into the adjacent region of lower (or greater) gray level, the corresponding operation performed is called thinning (or thickening). Given an ordered sequence of thresholds $\Lambda = \{\lambda_i | i = 0, 1, \ldots, L\}$, APs for the input component $F$ are obtained by applying a sequence of thinning and thickening operations as follows:

$$\mathrm{AP}(F) = \{\phi^{\lambda_L}(F), \phi^{\lambda_{L-1}}(F), \ldots, \phi^{\lambda_1}(F), \times F, \gamma^{\lambda_1}(F), \ldots, \gamma^{\lambda_{L-1}}(F), \gamma^{\lambda_L}(F)\} \quad (1)$$

where $\phi$ and $\gamma$ represent the extensive and antiextensive attribute filters, respectively, which rely on connected morphological operators (the underlying operators form an adjunction). In summary, the APs of a pixel are a function of the values of its adjacent pixels and can be used to characterize multiscale neighborhood-level information for images based on the attribute and criterion defined.

### B. Object-Based Image Segmentation

Object-based image analysis has been gaining much attention in the fields of remote sensing and geographical information science over the past decades [25]. It always starts with image segmentation, a process of partitioning a digital image into small separate regions (segments) according to certain criteria [50], in order to achieve a segmented image in support of objected-based feature representation. Creating representative image objects through image segmentation algorithms is crucial for object feature extraction and further remote sensing analysis. In this article, fractal net evolution approach (FNEA) was employed to execute the image segmentation task, which is a bottom-up region merging technique with a fractal iterative heuristic optimization procedure [51]. It starts with a single pixel and a pairwise comparison of its neighbors, with the aim of minimizing the resulting merged heterogeneity. The heterogeneity is determined using geometric shapes and the standard deviation of spectral properties as its basis. Compared with the existing image segmentation methods such as hierarchical stepwise optimization [52], iterative region growing using semantics [53], mean-shift [54], FNEA is adopted in this article based on the following advantages. First, it is a hierarchical segmentation method, which allows the differently sized geographic objects to be fully extracted by simply tuning the scale parameter. Second, merging criterion of FNEA not only considers spectral properties but also geometric information, thus irregularly shaped ground objects in multitemporal remotely sensed images can be extracted with relatively high accuracy. In addition, FNEA is mature and implemented in the commercial software eCognition, which is efficient and convenient for a user to handle, and has been successfully used in various OBIA applications [51], [55], [56].

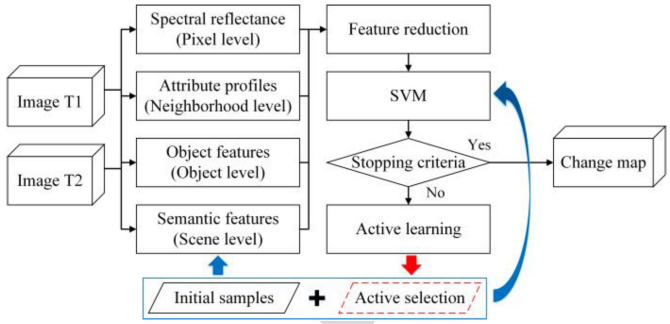



Fig. 1. Flowchart of the proposed change detection approach.

### C. Convolutional Neural Network

CNN is one of the most well-known deep learning methods, and is superior to other deep network algorithms owing to its ability of preserving the geometry of the image. Particularly, it maintains the interconnection between pixels, and thus preserves the spatial information of the images [57]. In general, a typical CNN consists of three types of layers, namely the convolution layer, the pooling layer, and the fully connected layer [58]. The convolution layer extracts information from previous layers and acts as a filter in the image domain. The values of the filter determine the type of information to be extracted. The pooling layer reduces the size of data and preserves the most important information of the input. In each pooling layer, compressed features are determined by subsampling of a small selected rectangle, in which the average or the maximum value in a region is used to replace the values of this region in the input features [59]. The fully connected layer is the reasoning part of the network, in which each neuron receives the information from all neurons in the previous layers to make the final decision of the input data.

## III. Proposed Change Detection Model

The proposed approach aims to investigate a proper way to integrate multilevel spectral–spatial information to improve representation and discrimination for change detection. The general flowchart of the proposed framework is shown in Fig. 1 with the following steps.

1) Given two preprocessed remote sensing images, multilevel difference information is extracted by spectral reflectance (pixel-level), AP filters (neighborhood-level), object feature extraction (object-level), and CNN model with initial training samples (scene-level).
2) The extracted features are fused through the Fractional-Order Darwinian Particle Swarm Optimization (FODPSO) method [60] to reduce their dimension.
3) The reduced features are put into the SVM to generate the change detection results and the corresponding posterior probabilities.

4) If the result does not meet the criteria given in advance, additional training samples will be selected and added by introducing the active learning algorithm. Steps 1) to 4) will be repeated until the result meets the requirement.

### A. Multilevel Feature Extraction

Extracting different level features from multitemporal images is the first and important step of the proposed approach. First, pixel-level features of bitemporal images can be obtained by directly using the preprocessed image bands. Second, as far as the neighborhood-level features are concerned, APs perform a contextual analysis of images considering measures computed on adjacent pixels. The attributes used in AP filtering for each temporal image in this article are as follows:

1) $s$, standard deviation of the gray-level values of the pixels, which measures the homogeneity of the connected regions;
2) $a$, the area of connectivity area, which is related to the size of the connected regions;
3) $d$, length of the diagonal of the box bounding the region, which is related to the folding degree of the connected regions; and
4) $i$, moment of inertia, which measures the elongation of the connected regions.

These four attributes are adopted in this research for two reasons. On the one hand, artificial changes present very different structures with different characteristics. For example, block buildings can be differentiated through the profiles constructed by using the area attribute. In turn, linear entities such as roads and streets can be distinguished by using moment of inertia attributes. On the other hand, using these four different APs cannot only comprehensively characterize complex surface change, but also reduce the collinearity among the feature spaces.

Third, to extract object-level features, FNEA is used for remote sensing image segmentation, thanks to its efficiency and stability. In FNEA, spectral and spatial information are considered to define a series of relatively homogeneous polygons, and several pixels or existing objects are merged into a larger one based on the following parameters: scale, color against shape weight, and smoothness against compactness weight [49]. Changing these criteria will change the shape and size of the objects produced by segmentation, allowing an image to be segmented at different scales. In order to optimize the scale parameters of FNEA, estimation of scale parameter (ESP) proposed in [61], [62], which relies on the potential of the local variance to detect scale transitions in geospatial data, is used to achieve the appropriate image segments for each temporal image. After image segmentation, all the pixels within a segment receive the same value of the feature computed for the entire segment. Multiple features are subsequently calculated based on the objects segmented in each temporal image as the object-level features, including object spectral features, object shape features, and object texture features.

Finally, to obtain the semantic features based on scene level, a CNN model is designed and conducted in the approach. An initialized training patch set $D$ corresponding to a limited number of randomly selected labeled pixels is first chosen and enlarged by data augmentation (DA), in which transformation strategies are used to augment the training set $D$ including flipping horizontally, flipping vertically, rotating 90 Â°, rotating 180 Â°, and rotating 270 Â° clockwise. After conducting DA, the training sample set $D$ has a fivefold increase and is denoted as $D_A$. It is worth noting that the size of these patches has a significant impact on the training results of CNN. A lager patch size enables the ability to capture more structural information for the center pixel, while a smaller patch size is beneficial to avoid the inclusion of information irrelevant to the center pixel [63]. Hence, balancing the size of input patch for CNN is crucial and needs to take the image size and spatial resolution into account. The deep learning model is subsequently conducted to extract scene-level features of each pixel by using the CNN since it has exhibited strong discriminative ability in a wide range of computer vision tasks [64], [65]. The detail structure of the CNN constructed in this article is shown in Fig. 2, which contains an input layer, two convolutions layers, two max-pooling layers, two fully connected layers, and a softmax output layer. Such structure cannot only meet the needs of extracting scene-level features, but also avoid low operating efficiency of deep network. It can be defined as follows:

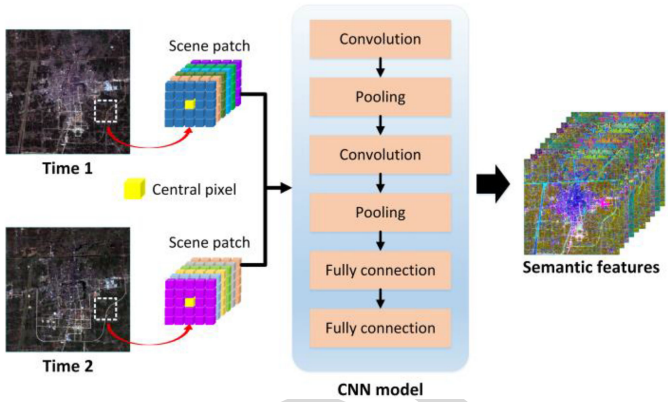



Fig. 2. Structure of the CNN designed for scene feature extraction. The first convolutional layer contains 20 filters with size $3 \times 3$, and the second convolutional layer includes 20 filters with size $2 \times 2$. After each convolutional layer, there is a max-pooling layer with kernel size $2 \times 2$ and a stride of two pixels. The first fully connected layer contains 500 units, and the unit number of the second fully connected layer is the number of change detection categories.

$$L(F_\Theta, D_A) = -\sum_{i=1}^{N}\sum_{k=1}^{K} l\{y_i = k\} \log P(y_i = k|x_i, F_\Theta) \tag{2}$$

where $L$ is the class probability of each pixel in the image by using CNN. $F_\Theta$ is a nonlinear function implemented by CNN with parameter $\Theta$. $N$ and $K$ are the total number of pixels and change categories in the image. $x_i$ and $y_i$ are the 3-D patch of the center pixel $i$ and its label. $l\{\cdots\}$ is the indicator function, and $P(y_i = k|x_i, F_\Theta)$ is the output of the CNN and represents the probability of $x_i$ to have label $k$. After conducting the CNN, the features on the final fully connected layer of the trained CNN model are the highly abstraction representation of the input patch, and can be used for the proposed approach to express

the scene-level information of the pixels in the central image patches [34]. It is worth noting that the spectral reflectance, APs, and object features are extracted from each temporal image and image differencing is conducted to generate the three-level difference features. While the 3-D patches obtained through the stacked bitemporal images are put into the CNN to generate the final scene-level difference features.

### B. Multilevel Feature Fusion

The total dimension of the features obtained from different levels is high. Moreover, some of these features may carry the information relevant for change detection. Thus, a weight parameter $\theta$ is assigned for each feature, $\theta \in [-1, 1]$. The feature is abandoned when its corresponding $\theta$ is less than 0, otherwise the feature is selected. The fitness function is set to the accuracy acquired by SVM with the training set. In order to find the optimal solutions, a feature selection strategy based on an FODPSO [60] is introduced. PSO searches the optimal solutions of fitness function using a swarm of particles. Each particle updates its moving direction according to the best position of itself (personal best) and the best position of the whole swarm (global best) [66], formulated as

$$V_i(t+1) = \omega V_i(t) + c_i r_i (P_p - X_i(t)) + c_2 r_2 (P_g - X_i(t)) \tag{3}$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{4}$$

where $V_i$ is the moving velocity at generation $i$ and $X_i$ is the particle position. $P_p$ denotes personal best and $P_g$ denotes global best. $\omega$, $c$, and $r$ denote the inertia weight, learning factors, and random numbers, respectively.

FODPSO can enhance the ability of traditional PSO based on the idea of running many simultaneous parallel PSO algorithms, each of which is seen as a different swarm on the same test problem. When a search tends to a suboptimal solution, the search in that area is simply discarded, and another area is searched instead. In this approach, at each step, swarms that get better are rewarded (extend particle life or spawn a new descendent), and swarms that stagnate are punished (reduce swarm life or delete particles). Meanwhile, fractional calculus is used to control the convergence rate of the algorithm. In summary, the FODPSO improves the reliability of finding the optimal solution of fitness function and can be used for reducing the relevant features.

### C. Training Sample Selected by Active Learning

Once the feature set to be involved in the task has been defined, a robust change detector should be selected for the supervised change detection step. SVM is chosen thanks to its intrinsic robustness to high-dimensional datasets and ill-posed problems. It has the advantages of superior generalization ability and insensitive value, which is suitable for solving small-sample and nonlinear model change detection problems [67]. However, due to the randomness of the training samples, the trained model is not always applicable for the entire feature set to achieve the optimal change detection result. Active learning, a popular strategy of selecting the most informative samples by querying for labeling in an iterative way, is thereby introduced. A variety of heuristic active learning strategies have been proposed in the machine learning field, such as uncertainty sampling [68], expected model change [69], and estimated error reduction [70]. Since the membership probability of the change categories can be obtained from the supervised change detector, active learning criteria with Best versus Second Best (BvSB) measure can thus be adopted. BvSB measure is specially designed for the multiclass identification problems and can alleviate the issue in which the performance is heavily influenced by small-class probabilities of unimportant classes through measuring the probability difference between the most confused classes, i.e., the first and the second most probable classes. Specifically, this criterion is defined as

$$\mathrm{BvSB}(i) = P_B(i) - P_{\mathrm{SB}}(i) \tag{5}$$

where $P_B(i)$ denotes the best class membership probability of sample $i$, and $P_{\mathrm{SB}}(i)$ is the second-best class membership probability of sample $i$. For this measure, if a sample has a small BvSB value, the classifier is confused with its class membership. Therefore, this kind of sample should be selected and trained in the consequent training iteration for refining CNN and SVM model, and thereby the efficiency will be reduced with the increasing of iterations inevitably.

In summary, the proposed change detection approach can be concluded in Algorithm 1. First, an initialized training patch set $D$ corresponding to a limited number of randomly selected labeled pixels is constructed. Next, the training set $D$ is augmented into a new training set $D_A$, which is used for training the CNN. After multilevel information extraction, the difference features combined with training samples are put into the SVM and the preliminary change detection results can be produced. Then, $b$ the most informative pixels are actively selected based on the class probabilities provided by the SVM and added into the current training set $D$, which is further regarded as a newly training set for the next round. This step together with the previous steps is then implemented iteratively until the stopping criterion is satisfied.

## IV. Experimental Results

### A. Dataset Description

Dataset 1 is the Landsat 5 Thematic Mapper (TM) images acquired on March 17, 2000 and February 6, 2003, covering the Taizhou city, China. The size of the multitemporal images is $400 \times 400$ pixels, with six spectral bands (Bands 1–5 and 7) and a spatial resolution of 30 m. During this period, urban expansion led to massive land cover changes. To quantitatively evaluate the performance of the proposed method, 4408 changed and 18 837 unchanged samples were labeled according to the prior information and detailed visual analysis of the multitemporal images. In detail, the reference data contained 3262 pixels of city expansion, 641 pixels of soil change, 505 pixels of water change, 1300 pixels of stable water, 13 197 pixels of stable vegetation, and 4340 pixels of stable city. Dataset 2 consists of two Sentinel-2 images obtained on February 7, 2016 and January 22, 2019, covering the Nanjing City, China. The bi-images

**Algorithm 1**

**Input:** Training sample set $D$, the number of round $R$, the number of initialized training pixels $a$, and the number of actively selected pixel in each round $b$
**Initialization**: $r = 1$
**While** $r < R$ or stopping criterion is not satisfied do
1: Data augmentation: $D \rightarrow D_A$
2: Pixel-level, neighborhood-level, and object-level feature extraction
3: CNN training ($r = 1$) or fine-tuning ($r > 1$) based on $D_A$ for scene-level feature extraction
4: Feature reduction based on FODPSO algorithm
5: Supervised change detection based on SVM with multi-level features and training samples
6: Calculating class probability of each sample based on SVM
7: Actively selecting additional $b$ pixels via BvSB criterion
8: Supplementing the corresponding patches of the selected pixels into $D$
9: $r = r + 1$
**End while**
**Output**: Final change detection results Y

contain 10 bands (Bands 2–8A, 11–12) and cover 160 × 140 pixels with a spatial resolution of 10 m after preprocessing. Variation of cultivated land and increased impervious surface created significant change during the study period. A total of 1899 changed pixels and 1913 unchanged pixels were labeled by careful visual interpretation for quantitative evaluation. In more detail, 971, 556, 138, 234, 1321, 269, and 323 pixels were labeled as bare land to building, vegetation to building, bare land to vegetation, changed road, stable building, stable bare land, and stable road. Dataset 3 is made up of a pair of bitemporal UAV images with a size of 359 × 537 pixels acquired on May 1, 2012 and May 8, 2014. The images contain three RGB optical bands with a spatial resolution of 2 m. In this area, the changes were mainly caused by urban construction. To assess the change detection results, 28 829 unchanged and 25 841 changed samples were labeled according to careful visual interpretation, including 2964 unchanged samples of water, 2534 unchanged samples of vegetation, 5689 unchanged samples of building, 11 980 unchanged samples of road, 5652 unchanged samples of bare land, 13 158 changed samples of vegetation to building, 7977 changed samples of vegetation to road, 4362 changed samples of bare land to building, and 344 changed samples of bare land to water. Fig. 3 shows the true color composite of each pair of bi-images and their reference change labels.

Fundamental image preprocessing, including radiometric calibration, orthorectification, and coregistration, was performed to reduce discrepancies between bitemporal images. Radiometric calibration was carried out to eliminate radiance or reflectance differences caused by the digitalization process of the remote sensing systems [2]. Orthorectification was used to remove relief displacement for the images on different dates [71]. Coregistration was essential to ensure the bitemporal image pixels or objects in the same location can be compared [72]. For ensuring high accuracy of the results, bitemporal images were coregistered to a root-mean-square error of less than 0.5 pixels.

### B. Parameter Settings

First, to extract object features for the proposed approach, image segmentation based on FNEA was conducted with the software Definiens eCognition Developer Version 9.0. ESP embedded in the software, which was proposed in [61], [62] to optimal scale segmentation parameters, was used in this research. For the other parameters, the color weight was set as 0.8, the shape weight was set as 0.2, and the smoothness and compactness weights were both set as 0.5 according to the trial and error tests [73]. After image segmentation, 18 object features were extracted for each pixel in the segments, which are listed in Table I.

Second, for calculating the neighborhood-level features in the proposed method, four different attributes with defined ranges of thresholds were considered to generate the APs according to the automatic scheme in [41] and prior knowledge of the datasets [74]:

$$\lambda_s = \frac{\mu}{100} \times \{\sigma_{\min}, \sigma_{\min} + \delta_s, \sigma_{\min} + 2\delta_s, \ldots, +\sigma_{\max}\} \quad (6)$$

$$\lambda_a = \frac{1000}{\varphi} \times \{a_{\min}, a_{\min} + \delta_a, a_{\min} + 2\delta_a, \ldots, +a_{\max}\} \quad (7)$$

$$\lambda_d = [5, 10, 15, \ldots, 100] \quad (8)$$

$$\lambda_i = [0.24, 0.28, 0.32, \ldots, 1.00] \quad (9)$$

where $\mu$ was the mean of the bands of bitemporal images and $\sigma_{\min}$, $\sigma_{\max}$, and $\delta_s$ were 0.15, 3, and 0.15, respectively. $\varphi$ was the spatial resolution of the input data and $a_{\min}$, $a_{\max}$, and $\delta_a$ were 0.075, 1.5, and 0.075, respectively.

Third, for the CNN model, the number of initialized training samples in the first iteration has an influence on the final result since the subsequent active sample selection steps partly rely on the performance of initialized trained CNN and SVM. Hence, in order to prove the capability of the proposed approach in handling the task with limited labels, the initial training samples were set within the range between 0.1% and 1% of the reference pixels for each category in the experiments. Then, the same number of additional training samples were actively selected and added in from the remaining reference labels in each consequent iteration. Moreover, we empirically obtained the parameter settings of the CNN structure, which is shown in the caption of Fig. 2. Some other parameters were also empirically set. For example, the batch size was 50, the learning rate was 0.001, and the scale parameter was 1.

In feature reduction process, as is proposed in [60], the overall accuracy of SVM over validation samples was used as the fitness value for FODPSO feature selection method. Meanwhile, the parameters of SVM (i.e., penalty parameter and Gamma), which is used as the change detector for the final extracted multilevel

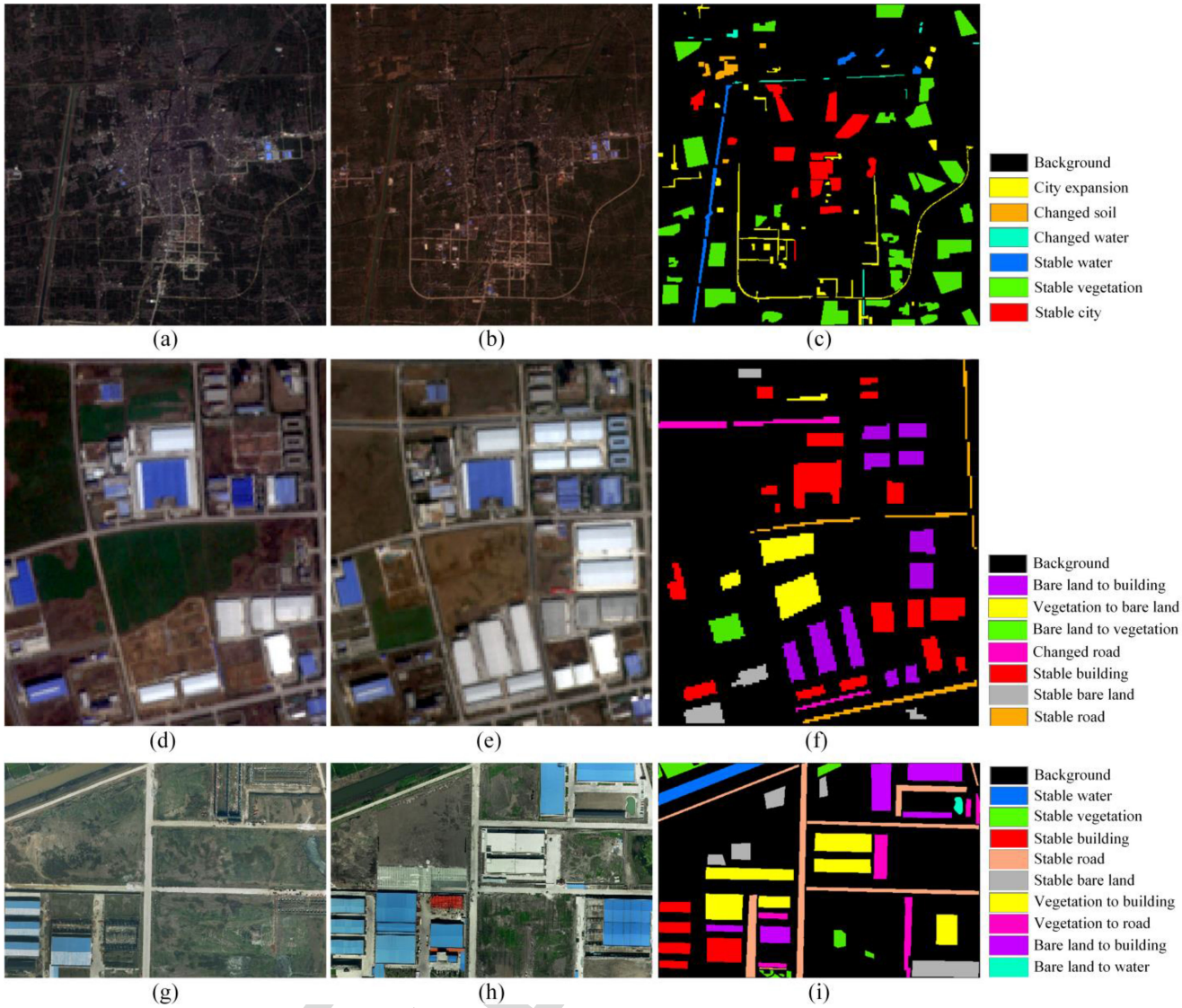



Fig. 3. True color composite bitemporal images of three datasets and their reference change map: (a) and (b) are Landsat 5 images in Taizhou acquired in 2000 and 2003; (d) and (e) are Sentinel-2 images in Nanjing acquired in 2016 and 2019; (g) and (h) are UAV images in Gaoyou acquired in 2012 and 2014; (c), (f), and (i) are the reference change map of the three datasets.

TABLE I
OBJECT FEATURES USED IN THE PROPOSED APPROACH AND THE COMPARATIVE OBJECT-BASED METHODS

| Type | No. | Feature names |
|---|---|---|
| Spectrum | 6 | Mean of all the spectral bands, brightness, and maxdiff. |
| Shape | 4 | Length-width ratio, compactness, density, and shape index. |
| Texture | 8 | Mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation derived from GLCM. |

features, were selected based on the PSO automatically. The Radial Basis Function (RBF) kernel was used for SVM in the proposed approach. To evaluate the performance of the proposed approach in each iteration, the reference samples were first divided into two parts, one was selected as the training sample to train the CNN and SVM model, the remainder was used as the test samples to examine the results of the first iteration. After assessment of the current iteration, these test samples became the candidate pool, in which the most informative samples were actively selected for training and the remaining were used as the new test samples in the next round until the iteration stops.

### C. Comparative Methods

To prove the effectiveness and superiority of the proposed method, some popular and advanced supervised change detection methods were conducted for comparison. First, change detection based on contextual information (referred as CBCD hereafter) presented in [18] was implemented, which can better utilize the spatial features of the adjacent pixels (neighborhood level) to distinguish the difference of bitemporal images. The

MPs and GLCMs for each band were extracted and used as the input. The second method was based on segmented images (OBCD), in which various types of object-based features are extracted to reduce the salt and pepper noise caused by outliers and improve the change detection performance [27]. The third method introduced active learning to extend OBCD (referred as OBAL hereafter) similar to the way presented in [75]. Additionally, two CNN-based change detection methods were also conducted. One method utilized AlexNet, which is a widely used model and has been proven effective for most remote sensing tasks [76]–[78], as the pretrained network for unsupervised scene-level feature extraction, and then identified the change class by SVM [79]. The other one adopted the Siamese CNN presented in [80], which is an advanced network suitable for multi-input tasks, to achieve the change detection results. It is a type of network that uses two or more identical subnetworks that have the same architecture and share the same parameters and weights, and is typically used in the applications that involve finding the relationship between two comparable things, such as change detection. Additionally, the other contrastive methods were carried out though feature combination strategies from the aforementioned methods, including the change detection based on contextual and object features (COBCD), and contextual, object ,and scene features from AlexNet (COAlexCD). In addition, since all the features used in the comparative and the proposed methods were extracted based on the original pixel value of the spectral reflectance, the method using original bands (referred as Referee hereafter) was implemented as benchmark to highlight the advantages of the proposed approach as well as the other methods simultaneously. The change detector in the listed methods was all conducted by SVM with RBF kernel. For accuracy assessment, the training samples were randomly selected from the reference labels, and the remaining were used as the test samples.

### D. Experimental Results

After conducting the proposed change detection approach and the comparative methods for the aforementioned three datasets, their performances are displayed as follows.

*1) Landsat 5 Dataset:* Five rounds of iteration were conducted in the proposed approach. In each round, four level features were extracted with the parameters mentioned before. The CNN trained in each round contains 60 epochs, in which 0.5% of reference labels were randomly selected as the initial training samples in the first round for CNN model training and then for SVM training. Next, the same number training samples were actively selected and added in each subsequent round to repeat the steps until the iteration stops. The scene size was set as $8 \times 8$ pixels for the high-level semantic feature extraction. For the neighborhood-based contextual features in CBCD, COBCD, and COAlexCD, the GLCM features (e.g., mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment, and correlation) with a window size of $3 \times 3$ pixels and multiscale MPs (e.g., opening and closing) though the disk-shaped structuring element with radius of 2, 4, and 5 pixels for each spectral band were extracted, respectively. The object features in OBCD, COBCD, and COAlexCD were extracted by the same method as the object-level features in the proposed approach. For the two CNN models in AlexCD and SiamCD, an $8 \times 8$ pixel-window (the same size with the proposed approach) patch centered on each pixel for each band was also used to extract the features. Note that the experimental results of all the listed methods were achieved by the mean of 10 Monte Carlo runs to reduce the uncertainty caused by random initialization. The accuracies of the final round of the proposed approach and the other methods with the same number of total training samples are reported in Table II.

From Table II, it can be seen that the proposed approach using multilevel feature strategy outperformed the other listed methods in terms of the highest OA, which is 3.45%, 3.72%, 3.72%, 3.81%, 4.40%, 4.66%, 5.36%, 7.81%, and 10.72% higher than COAlexCD, COBCD, SiamCD, OBAL, OBCD, CBCD, AlexCD, and Referee methods, respectively. More specifically, it has improved the accuracies from 0.16% to 4.87% compared with the best results of all the other methods in each change category except changed soil type, which means that the proposed approach is not only more accurately but also more comprehensively improved the change detection results. From the perspective of error analysis for binary change detection, the proposed approach achieved least commission, omission, and total error with 1.19%, 0.14%, and 0.25%, respectively.

The change detection maps of different methods are shown in Fig. 4, where the following results can be noticed. First, the change map by using the Referee method was discrete with more missed and false alarms, which show up in the form of salt and pepper noise. Second, for the results of CBCD, structuring elements with different sizes were selected according to the ground truth, so that it can detect the difference of surface information with different neighborhood scales. As a result, the salt and pepper noise was reduced, as well as the changed and unchanged categories were more clearly distinguished. The OBCD searched optimal segmentation scales considering the entire images and used multiple object features to describe their change information, which generated the changing patterns similar to real ground objects. For example, the category of urban expansion (new buildings and roads) in Fig. 4(c) is more continuous and regular than those in Fig. 4(b). After introducing active learning to OBCD, some misclassifications were corrected, thanks to refined samples in OBAL. The change class in AlexCD [Fig. 4(f)] and SiamCD [Fig. 4(g)] considered the entire scene centered on each pixel, which made full use of the information in a scene and prevented the results from being affected by abnormal pixel values such as shadows and clouds. However, the detection of the change in the broken area is their weakness. For example, there were some misclassifications of stable vegetation, stable and changed water in the north of the study area. With the combination of different level features such as COBCD [Fig. 4(e)] and COAlexCD [Fig. 4(h)], the change detector obtained a better model considering all these features through training process, so that the boundaries of each change and stable category were more clear and regular. Their false and missed alarms were reduced accordingly. Fig. 4(i)–(m) shows the proposed results with 1 to 5 iterations. Since it combined

TABLE II
CHANGE DETECTION ACCURACIES OF DIFFERENT METHODS FOR THE LANDSAT 5 DATASET

| Accuracy | Referee | CBCD | OBCD | OBAL | COBCD | AlexCD | SiamCD | COAlexCD | Proposed |
|---|---|---|---|---|---|---|---|---|---|
| Ch-1 [a] | 92.18% | 89.89% | 89.96% | 95.87% | 94.95% | 85.10% | 94.33% | 94.85% | 99.03% |
| Ch-2 | 66.23% | 72.95% | 76.48% | 69.09% | 90.48% | 86.71% | 94.57% | 78.90% | 90.42% |
| Ch-3 | 82.69% | 72.31% | 59.88% | 90.48% | 78.48% | 39.75% | 68.98% | 80.78% | 93.06% |
| Un-1 | 68.35% | 96.42% | 96.10% | 99.35% | 96.98% | 84.85% | 99.37% | 96.93% | 99.53% |
| Un-2 | 93.84% | 98.09% | 97.86% | 97.73% | 98.04% | 96.35% | 97.98% | 98.06% | 99.96% |
| Un-3 | 79.56% | 88.82% | 94.30% | 88.48% | 90.83% | 89.79% | 90.55% | 92.10% | 99.17% |
| OA [b] | 88.53% | 93.89% | 94.59% | 94.95% | 95.53% | 91.44% | 95.44% | 95.53% | 99.25% |
| Kappa | 0.8108 | 0.8992 | 0.9115 | 0.9165 | 0.9270 | 0.8586 | 0.9257 | 0.9273 | 0.9878 |
| CE [c] | 6.48% | 8.58% | 12.45% | 3.26% | 4.90% | 15.42% | 4.66% | 4.42% | 1.19% |
| OE [d] | 2.86% | 0.74% | 3.91% | 1.53% | 1.02% | 7.50% | 2.83% | 0.68% | 0.14% |
| TA [e] | 1.75% | 1.74% | 3.04% | 0.87% | 1.11% | 4.21% | 1.41% | 0.96% | 0.25% |

[a] Accuracy of multiple change detection categories: Ch-1 = City expansion, Ch-2 = Changed soil, Ch-3 = Changed water; Un-1 = Stable water, Un-2 = Stable vegetation, Un-3 = Stable city; [b] Overall accuracy; [c] Commission error; [d] Omission error; [e] Total error; CE, OE, and TA are the indicators for binary change detection results.

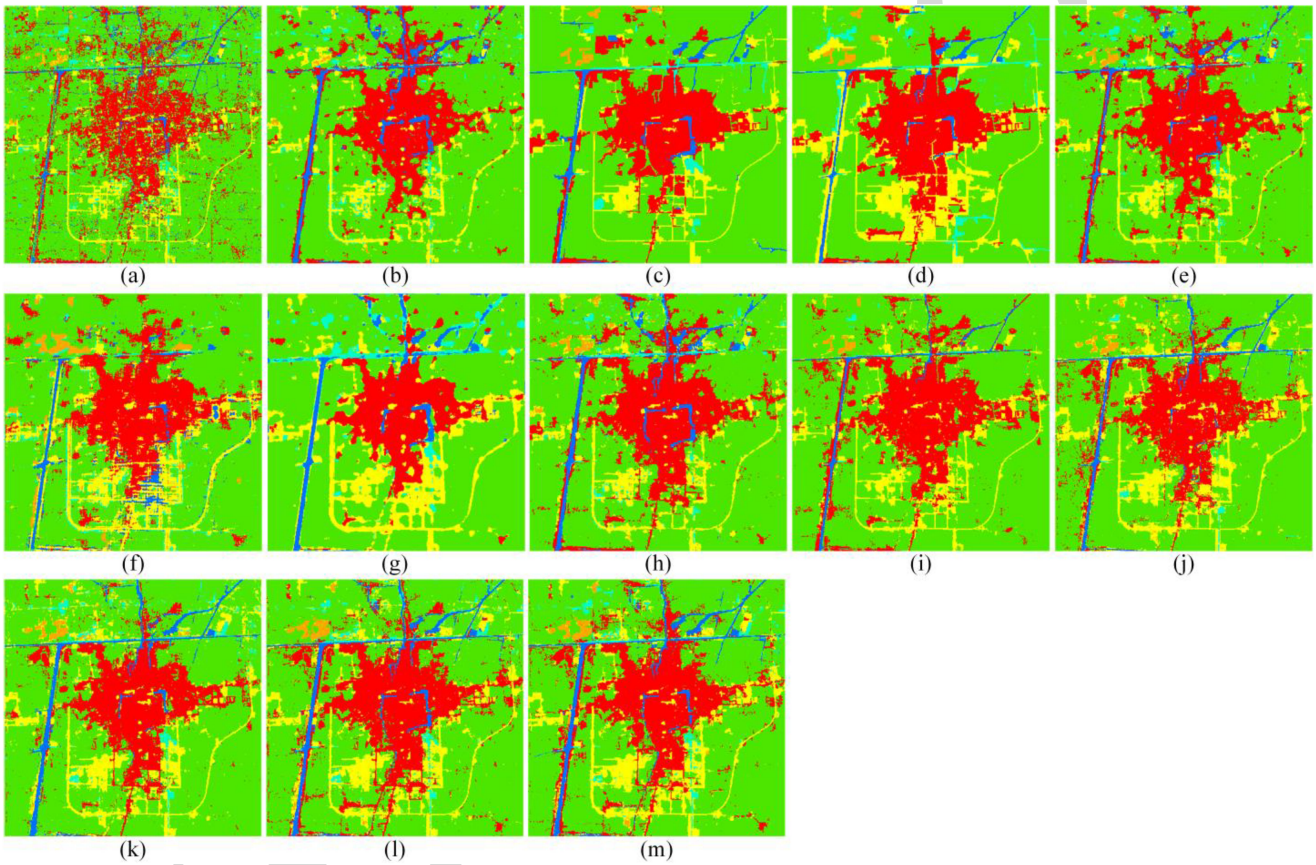



Fig. 4. Change detection maps of Landsat 5 dataset with different methods: (a)–(h) are the results of Referee, CBCD, OBCD, OBAL, COBCD, AlexCD, SiamCD, and COAlexCD; (i)–(m) are the results of the proposed approach with iteration from 1 to 5.

multilevel information to comprehensively characterize the image difference between bitemporal images, it achieved a good result at initial round. As the iteration progressed, the models of CNN and SVM were both continuously optimized with active selected samples, resulting in significant enhancements to some categories. For example, the identification of stable water in the central city and changed soil in the southwest of the study area have continuously been improved as the iteration increased.

Furthermore, to evaluate the performance of the proposed approach with iteration increasing, additional experiments with corresponding training samples were also conducted. Specifically, we conducted five rounds for the proposed method with active learning. Accordingly, the same number of training samples (0.5%, 1.0%, 1.5%, 2.0%, and 2.5% of the reference labels) as each iteration in proposed approach was used in the other contrastive methods. The results are displayed in Fig. 5. It shows that the proposed approach has achieved the best OA among all the methods regardless of how many training samples were used, which is due to its comprehensive characterization of change information with different levels. The results of the COBCD and COAlexCD methods performed slightly worse, with the

TABLE III
CHANGE DETECTION ACCURACIES OF DIFFERENT METHODS FOR THE SENTINEL-2 DATASET

| Accuracy | Referee | CBCD | OBCD | OBAL | COBCD | AlexCD | SiamCD | COAlexCD | Proposed |
|---|---|---|---|---|---|---|---|---|---|
| Ch-1 [a] | 99.89% | 99.89% | 99.57% | 99.80% | 99.89% | 99.98% | 99.88% | 99.89% | 99.98% |
| Ch-2 | 99.82% | 99.08% | 99.27% | 99.82% | 99.92% | 99.82% | 99.84% | 99.81% | 99.96% |
| Ch-3 | 96.18% | 99.23% | 99.84% | 99.92% | 99.96% | 99.84% | 99.96% | 99.90% | 99.92% |
| Ch-4 | 84.55% | 75.89% | 67.11% | 79.07% | 77.23% | 78.03% | 76.44% | 83.11% | 99.56% |
| Un-1 | 98.35% | 93.32% | 95.05% | 94.62% | 97.64% | 96.55% | 99.92% | 96.53% | 99.92% |
| Un-2 | 96.69% | 85.38% | 90.91% | 93.57% | 79.92% | 99.61% | 99.61% | 99.61% | 99.96% |
| Un-3 | 27.01% | 98.71% | 95.39% | 91.72% | 93.27% | 90.45% | 93.23% | 97.04% | 96.78% |
| OA [b] | 91.83% | 94.90% | 95.04% | 95.78% | 95.80% | 96.59% | 97.95% | 97.44% | 99.70% |
| Kappa | 0.8923 | 0.9347 | 0.9360 | 0.9456 | 0.9457 | 0.9562 | 0.9735 | 0.9670 | 99.61% |
| CE [c] | 1.80% | 1.59% | 2.61% | 0.49% | 0.93% | 1.37% | 2.90% | 1.90% | 0.05% |
| OE [d] | 0.06% | 0.11% | 0.06% | 0.22% | 0.28% | 1.37% | 0.67% | 0.01% | 0.01% |
| TA [e] | 0.93% | 0.84% | 1.34% | 0.35% | 0.60% | 1.36% | 1.77% | 0.95% | 0.03% |

[a] Accuracy of multiple change detection categories: Ch-1 = Bare land to building, Ch-2 = Vegetation to building, Ch-3 = Bare land to vegetation, Ch-4 = Changed road, Un-1 = Stable building, Un-2 = Stable bare land, Un-3 = Stable road; [b] Overall accuracy; [c] Commission error; [d] Omission error; [e] Total error; CE, OE, and TA are the indicators for binary change detection results.

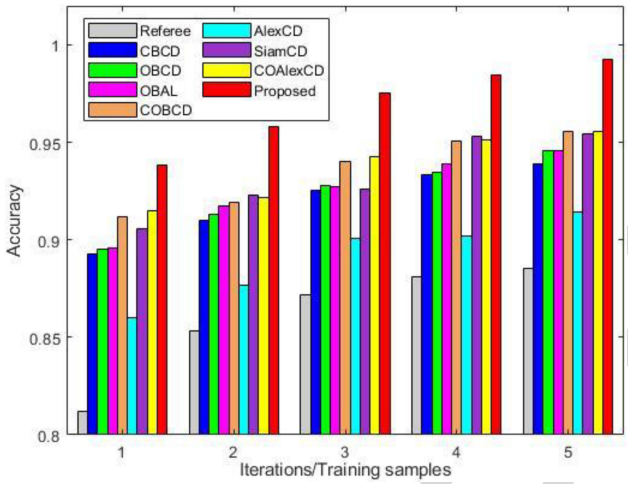



Fig. 5. Comparison of different change detection results of Landsat 5 dataset with increasing number of training samples.

average accuracy of 3.25% and 3.42% lower than the proposed in different amount of training samples. The performance of the SiamCD, OBAL, OBCD, CBCD, and AlexCD methods were mediocre, with the average accuracy of 3.73%, 4.39%, 4.65%, 4.98%, and 7.90% lower than the proposed approach. The accuracy of the Referee method is the most unsatisfactory from beginning to the end with the average accuracy 10.92% lower than the proposed due to the use of only simple pixel-based surface reflectance. It can also be seen from the figures that the accuracy advantage of the proposed approach suddenly increased from the second iteration, indicating that its performances were further improved with the optimized samples from the second iteration by active learninig.

*2) Sentinel-2 Dataset:* This experiment contained four rounds of iterations, in which 50 epochs and 0.1% of the reference labels were used for initial training of the CNN and SVM. In the consequent iterations, additional actively selected 0.1% samples were put in to form the new set for training. The size of scene patch was set as $10 \times 10$ pixels, while the moving window and the radius of disk-shaped structuring element for GLCM and MPs were $5 \times 5$ pixels and [4], [5], [6]. For controlling variables, the size of the input patches in AlexCD and SiamCD was also set as $5 \times 5$. The results of these methods were all achieved based on the average of 10 Monte Carlo runs to avoid random errors.

Based on the quantitative analysis results in Table III, it can be concluded that the proposed approach achieved highest OA of 99.70% and Kappa of 0.9961 thanks to its comprehensive feature description and active sample selection. Three contrastive methods containing the CNN model, such as SiamCD, COAlexCD, and AlexCD, also obtained excellent results with precisions of 97.95%, 97.44%, and 96.59%, since the scene-level feature was highly efficacious for characterizing surface change information in this dataset. The results of COBCD, OBAL, OBCD, and CBCD were close and slightly inferior, which were 95.80%, 95.78%, 95.04%, and 94.90%, respectively. In these methods, although the scale setting of the segmentation and connection area met acquirements of most ground objects in the study area, there were still a few land cover changes that were difficult to take into account. The performance of the Referee method was much worse than the others because it only considered the spectral information of the pixels and ignored their spatial relationship. From the perspective of individual change class, the proposed approach also achieved the best performance except for the category of stable road. Meanwhile, it also obtained the least commission error (0.05%), omission error (0.01%), and total error (0.03%) in identification of change and nonchange.

The maps of change detection with different methods are displayed in Fig. 6. The Referee obtained the worst result with much salt and pepper noise. CBCD achieved a better result with the ground object being much more continuous. With the segmentation-based features introduced, the edges of the objects were much more regular and closer to reality in OBCD, OBAL, and COBCD, especially for those changes related to artificial constructions such as buildings and roads. In the result of SiamCD, the contours of the objects were smoother and the boundaries of complex edges were blurred due to the consideration of scene-level information with multiple inputs in the Siamese network. Compared with former methods, AlexNet has enhanced the ability of distinguishing the categories between

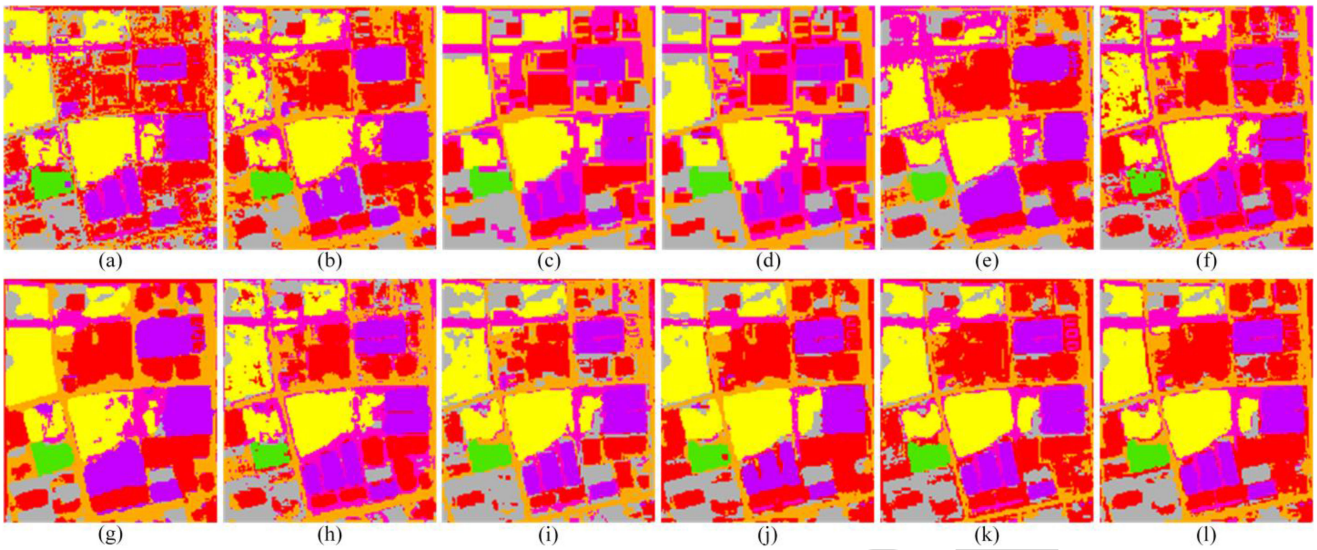



Fig. 6. Change detection map of Sentinel-2 dataset with different methods: (a)–(h) are the results of Referee, CBCD, OBCD, OBAL, COBCD, AlexCD, SiamCD, and COAlexCD; (i)–(l) are the results of the proposed approach with iteration from 1 to 4.

artificial constructions such as the stable building, stable road, and changed road in the northeast of the study area. But it made a few misclassifications between stable building and the transformation from vegetation to bare land in the upper left farmland. However, after the introduction of neighborhood and object features in COAlexCD, this confusion has been significantly improved. The proposed approach achieved a good result at initial stage thanks to the multilevel features. But there were some misclassifications between stable bare land and stable building in the southwest corner, and stable road and stable building in the northeast corner. As iterations increased, additional representative training samples were added and most errors were finally corrected.

To evaluate the performance of the methods under the condition of different labels, the results of four rounds of the proposed approach and the same number of training samples (1%, 2%, 3%, 4%, and 5% of the reference samples) to the other methods were achieved and are displayed in Fig. 7. It can be noticed that the proposed approach outperformed the other methods in all situations, closely followed by SiamCD, COAlexCD, and AlexCD, which were 0.88%, 1.23%, and 1.40% lower in average. The accuracies of COBCD, OBAL, CBCD, and OBCD were moderate, which were 2.48%, 2.72%, 3.10%, and 3.55% lower than the proposed in average. The accuracy of Referee was 5.21% less than the proposed approach on average, which was the worst of all methods. In summary, owing to the combination of multilevel information and optimized training samples chosen by active learning, the proposed change detection approach achieved the best results under different sample conditions.

*3) UAV Dataset:* Four rounds of change detection were conducted in the proposed approach, in which four level features were extracted. The constructed CNN model in the first round contains 80 epochs and 0.2% of the reference labels for extracting scene features. These training samples were also used as the input for SVM in the same round. By that analogy, the

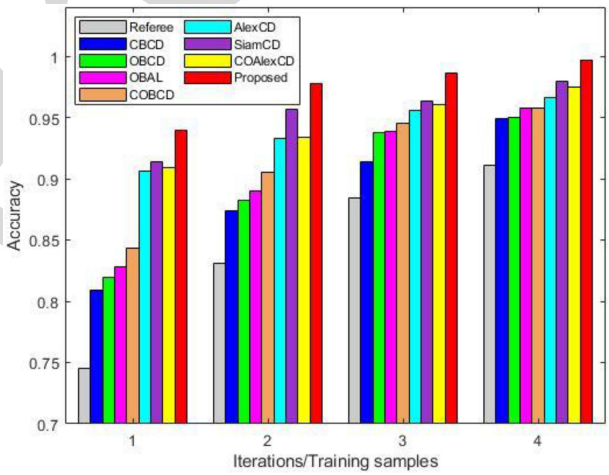



Fig. 7. Comparison of different change detection results of Sentinel-2 dataset with increasing number of training samples.

subsequent four rounds for CNN training also contained 80 epochs, in which an additional 0.2% of the reference labels were actively selected and added in to enlarge the training sample set when each additional round was conducted. The window size was set as 15 × 15 pixels for the scene-level feature extraction. For contextual features, the eight GLCM features were extracted by 7 × 7 pixels moving window for each band, and multiscale opening and closing MPs were extracted though the disk-shaped structuring element with radius of 5, 6, and 7 pixels for each band. To control the variables, the scene size was also set as 15 × 15 pixels for extracting scene-level features in AlexCD and SiamCD. The final results were given as the average over 10 runs for all of the methods to avoid random errors. Accuracies are achieved and listed in Table IV.

Based on Table IV, it can be observed that the proposed approach achieved best OA with 99.35%, followed by the

TABLE IV
CHANGE DETECTION ACCURACIES OF DIFFERENT METHODS FOR THE UAV DATASET

| Accuracy | Referee | CBCD | OBCD | OBAL | COBCD | AlexCD | SiamCD | COAlexCD | Proposed |
|---|---|---|---|---|---|---|---|---|---|
| Un-1 [a] | 99.32% | 99.86% | 98.54% | 99.25% | 99.93% | 98.64% | 93.96% | 98.88% | 99.98% |
| Un-2 | 29.49% | 96.04% | 70.90% | 97.93% | 98.25% | 90.41% | 82.70% | 99.56% | 99.48% |
| Un-3 | 80.89% | 97.36% | 91.64% | 99.81% | 97.22% | 98.12% | 98.46% | 99.31% | 99.96% |
| Un-4 | 90.34% | 95.77% | 93.36% | 95.95% | 95.81% | 97.04% | 96.92% | 97.68% | 99.83% |
| Un-5 | 90.22% | 90.93% | 96.04% | 95.97% | 91.13% | 94.56% | 69.65% | 95.63% | 99.59% |
| Ch-1 | 87.75% | 87.90% | 91.18% | 93.05% | 97.41% | 92.26% | 94.12% | 95.75% | 99.86% |
| Ch-2 | 76.97% | 88.30% | 98.46% | 87.52% | 92.90% | 83.90% | 90.29% | 93.82% | 99.92% |
| Ch-3 | 47.60% | 65.98% | 81.59% | 78.42% | 83.37% | 71.37% | 32.36% | 83.02% | 95.28% |
| Ch-4 | 80.88% | 81.36% | 89.41% | 96.11% | 71.13% | 44.97% | 85.71% | 88.72% | 78.53% |
| OA [b] | 80.97% | 90.21% | 91.95% | 93.31% | 94.62% | 91.23% | 86.56% | 95.53% | 99.35% |
| Kappa | 0.7703 | 0.8832 | 0.9042 | 0.9203 | 0.9357 | 0.8957 | 0.8386 | 0.9467 | 0.9923 |
| CE [c] | 5.64% | 4.35% | 2.72% | 1.73% | 1.14% | 4.24% | 3.05% | 2.49% | 0.52% |
| OE [d] | 5.38% | 2.01% | 2.13% | 1.79% | 1.90% | 2.47% | 8.18% | 1.22% | 0.12% |
| TA [e] | 5.20% | 2.98% | 2.29% | 1.66% | 1.45% | 3.15% | 5.52% | 1.75% | 0.30% |

[a] Accuracy of multiple change detection categories: Un-1 = Stable water, Un-2 = Stable vegetation, Un-3 = Stable building, Un-4 = Stable road, Un-5 = Stable bare land, Ch-1 = Vegetation to building, Ch-2 = Vegetation to road, Ch-3 = Bare land to building, Ch-4 = Bare land to water. [b] Overall accuracy. [c] Commission error. [d] Omission error. [e] Total error. CE, OE, and TA are the indicators for binary change detection results.

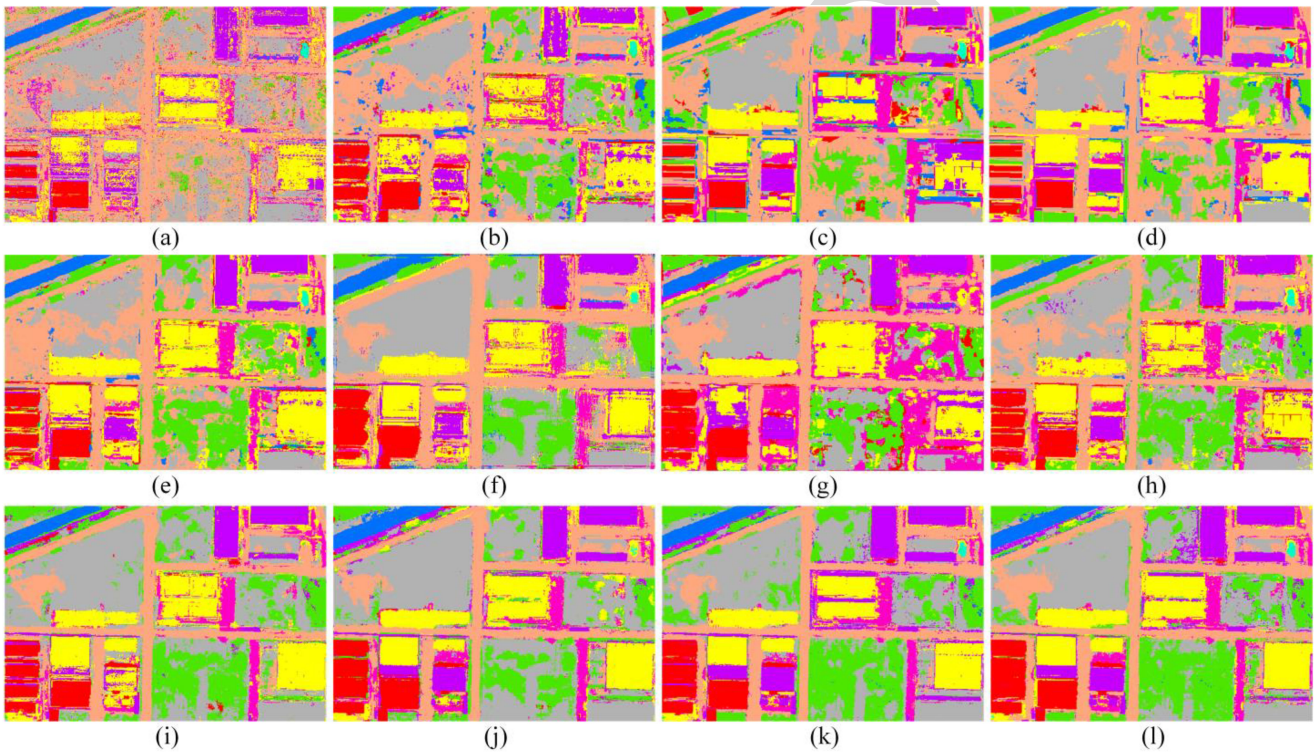



Fig. 8. Change detection map of UAV dataset with different methods: (a)–(h) are the results of Referee, CBCD, OBCD, OBAL, COBCD, AlexCD, SiamCD, and COAlexCD; (i)–(l) are the results of the proposed approach with iteration from 1 to 4.

COAlexCD (95.53%), COBCD (94.62%), OBAL (93.31%), OBCD (91.95%), AlexCD (91.23%), CBCD (90.21%), SiamCD (86.56%), and the Referee method (80.59%). From the aspect of individual class accuracy, the proposed approach achieved the best class accuracy for most classes except Un-2 (stable vegetation) and Ch-4 (transformation from bare land to water). It also achieved the least commission error, omission error, and total error for binary change detection results with 0.52%, 0.12%, and 0.30%. In summary, the proposed approach was not only effective in distinguishing between change and nonchange, but also outstanding in identifying specific change categories.

The change maps of the listed methods are displayed in Fig. 8. The results based on the Referee method were not good enough due to the confusion of stable road, stable bare land, and stable vegetation (sparse grassland) when only RGB bands were available. Thus, the distribution of change categories was discrete with significant salt and pepper noise. The accuracy of most categories in the CBCD has been improved, in which the distinction between stable road and stable vegetation was the

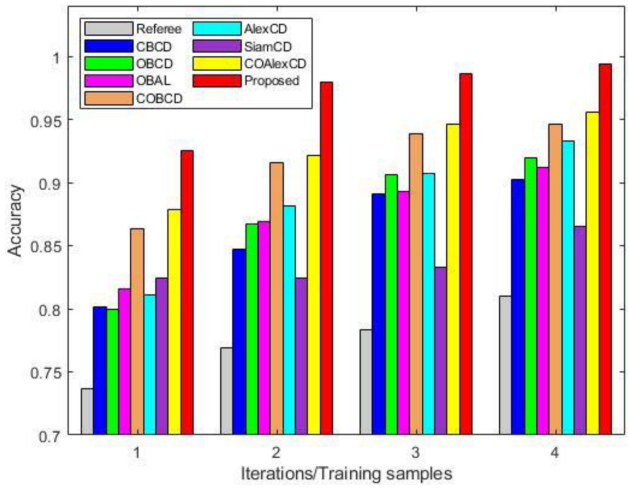



Fig. 9. Comparison of different change detection results of UAV dataset with increasing number of training samples.

most obvious. The salt and pepper noise of the entire result was also reduced. However, there was still some room for improvement in recognizing the difference between stable bare land and stable road. Compared with the methods of Referee and CBCD, the change map of OBCD and OBAL not only enhanced the accuracy of the result, but also made their shape more similar to the ground object. The result of AlexCD and SiamCD obtained better visual inspections although their accuracies were not as high as the OBCD and OBAL. In particular, contours of the changed buildings and roads were more clear and more accurate than the previous ones. By integrating the multiple features, the results of COBCD and COAlexCD have been enhanced, in which the confusion between roads and bare land has been distinctly reduced, and the discrimination among the types of stable and newly added buildings has been improved. The maps generated by the proposed approach with different iterations achieve better results in most categories, especially the categories related to the urban facilities with complex structures such as stable and changed buildings and roads, which reflected the comprehensiveness of multilevel features and the effectiveness of introducing active learning. Additionally, as the iterations increased, the results of change detection got even better.

In order to explore the performance under conditions with different degrees of prior knowledge, experiments with increasing training samples were conducted. According to the amount of initial and additional selected training samples in the proposed approach, the corresponding number of training samples (0.2%, 0.4%, 0.6%, 0.8%, and 1.0% of the reference pixels) was randomly selected for all the contrastive methods. Their results were obtained and displayed in Fig. 9. Similar to the previous results, the proposed approach achieved the best OA with regardless of how many the training samples were followed by the COAlexCD, COBCD, OBAL, OBCD, AlexCD, CBCD, and SiamCD, which were 4.57%, 5.51%, 8.78%, 9.79%, 9.85%, 11.07%, and 13.14% lower than the proposed approach in average. Also, the basic RGB bands of the original images were obviously not sufficient for describing the different changed and unchanged categories when they were directly used as the input, resulting in the worst performance among all methods (19.63% lower than the proposed approach in average). In addition, since the active selected samples were added in, the accuracy advantage of the proposed approach was further expanded after the second iteration.

## V. Discussion

### A. Sensitivity of Scale Parameter

The scale of scene patch is the key parameter for CNN feature extraction in the proposed approach. It determines the range which is centered on a pixel to extract its high-level semantic features. Therefore, as part of the multi-level features, it will affect the final change detection results to a certain extent. To evaluate its influence, various window sizes for cropping patches to extract scene features for these three datasets were carried out. Their effects are depicted in Fig. 10. In the first and second rounds of the proposed approach, the size of scene patch had a significant impact on the results. It can be seen that when the window size was less than or equal to $8 \times 8$ pixels (Landsat 5 dataset), $10 \times 10$ (Sentinel-2 dataset), and $15 \times 15$ pixels (UAV dataset), the accuracy improved as the scale increases. Conversely, when the window size was greater than or equal to $8 \times 8$ pixels (Landsat 5 dataset), $10 \times 10$ (Sentinel-2 dataset), and $15 \times 15$ pixels (UAV dataset), the accuracy began to decrease as the scale increased. Since the active learning selected the most informative samples for training as the round increased, the accuracy gap between different scenes reduced. That is to say, the accuracy of the overall change detection results gradually converged. Although different scenes still had a certain impact on the result in this condition, it was too small and could be even ignored when sufficient rounds were conducted. Therefore, choosing the appropriate scene scale and the number of iterations is crucial to effectively and efficiently enhancing the accuracy of the proposed approach.

### B. Time Consumption

To evaluate the efficiency of the proposed change detection approach, computational time of the aforementioned methods in all three datasets were recorded and listed in Table V. All the experiments were carried out using MATLAB R2018a on Intel (R) Core (TM) i7-6700 PC machine with 3.4 GHz of CPU and 16 GB of RAM. It can be seen from the table that the Referee method using spectral bands took the least time. CBCD took a few more seconds due to the contextual feature extraction process. Since obtaining object-based features needs to conduct image segmentation, additional tens of seconds are needed for OBCD, OBAL, and COBCD compared with the previous two methods (e.g., computational time of image segmentation for Landsat 5, Sentinel-2, and UAV datasets: 22.67, 27.41, and 24.58 s). AlexCD contained 25 layers and the features were extracted based on the last fully connected layer "fc8," whose dimension was 1000 (e.g., computational time of feature extraction in AlexNet for Landsat 5, Sentinel-2, and UAV datasets: 10045.36, 1880.58, and 24839.39 s). As a consequence, much

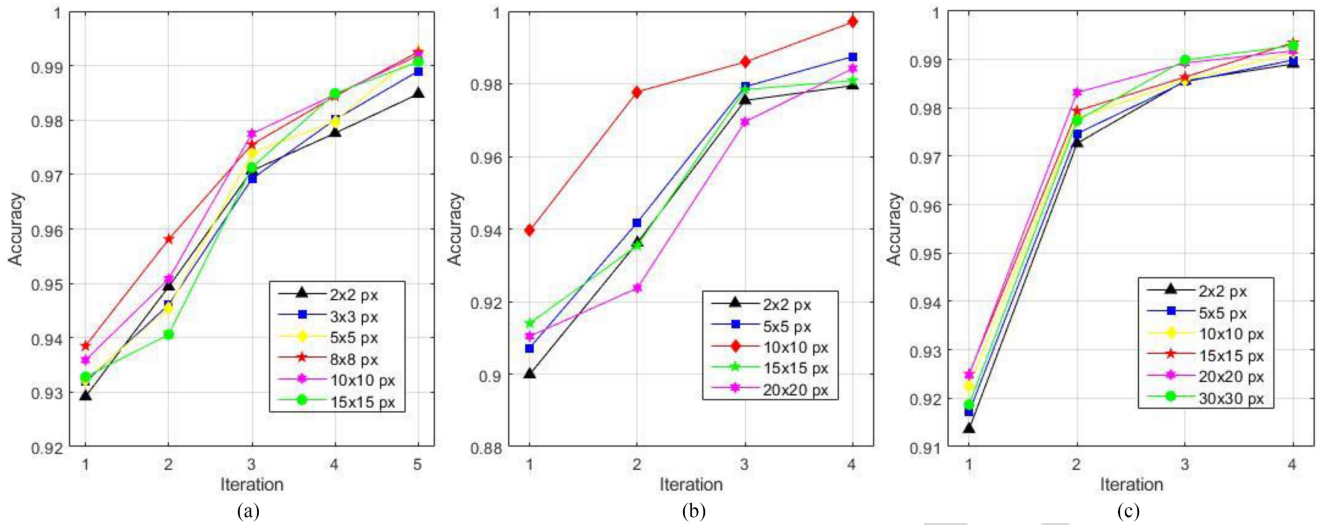



Fig. 10. Comparison of change detection accuracies conducting different iterations in the proposed approach with various scene scales. (a) Landsat 5 dataset. (b) Sentinel-2 dataset. (c) UAV dataset.

TABLE V
COMPARISON OF TIME CONSUMPTION WITH DIFFERENT CHANGE DETECTION METHODS

| Case | Ite/Samp | Referee | CBCD | OBCD | OBAL | COBCD | AlexCD | SiamCD | COAlexCD | Proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| Taizhou | 1 / 0.5% | 1.98 s | 5.79 s | 25.22 s | 25.40 s | 28.87 s | 10095.77 s | 1185.48 s | 10115.22 s | 786.07 s |
| | 2 / 1.0% | 2.49 s | 7.54 s | 25.85 s | 26.76 s | 31.92 s | 10137.64 s | 1301.13 s | 10180.99 s | 1574.02 s |
| | 3 / 1.5% | 3.27 s | 9.82 s | 26.53 s | 28.31 s | 36.02 s | 10185.89 s | 1422.27 s | 10245.04 s | 2705.60 s |
| | 4 / 2.0% | 3.79 s | 14.83 s | 27.46 s | 30.06 s | 38.95 s | 10232.64 s | 1531.83 s | 10362.09 s | 3747.83 s |
| | 5 / 2.5% | 4.56 s | 22.76 s | 28.30 s | 31.78 s | 46.06 s | 10288.86 s | 1678.65 s | 10415.54 s | 4773.66 s |
| Nanjing | 1 / 1.0% | 1.39 s | 2.36 s | 28.92 s | 28.96 s | 29.69 s | 1887.41 s | 184.14 s | 1916.33 s | 105.90 s |
| | 2 / 2.0% | 1.54 s | 2.94 s | 29.03 s | 29.68 s | 30.36 s | 1891.86 s | 229.62 s | 1923.80 s | 212.48 s |
| | 3 / 3.0% | 1.71 s | 3.69 s | 29.19 s | 30.23 s | 30.87 s | 1894.92 s | 274.98 s | 1931.35 s | 343.77 s |
| | 4 / 4.0% | 1.87 s | 4.27 s | 29.29 s | 31.02 s | 31.75 s | 1905.20 s | 323.16 s | 1940.99 s | 478.10 s |
| Gaoyou | 1 / 0.2% | 2.69 s | 5.73 s | 28.33 s | 28.47 s | 31.35 s | 24914.95 s | 1542.69 s | 24948.46 s | 850.11 s |
| | 2 / 0.4% | 3.31 s | 7.52 s | 28.87 s | 30.05 s | 33.90 s | 24967.63 s | 1680.42 s | 25018.17 s | 1936.66 s |
| | 3 / 0.6% | 3.92 s | 9.78 s | 29.75 s | 32.36 s | 37.65 s | 25021.62 s | 1798.50 s | 25090.67 s | 2832.93 s |
| | 4 / 0.8% | 4.73 s | 13.05 s | 30.51 s | 34.69 s | 40.42 s | 25091.07 s | 1936.02 s | 25154.77 s | 4289.27 s |

time was spent. On top of that, COAlexCD also led to longer running time due to this reason. SiamCD consisted of multi-stream networks, thereby the process of parameter optimizing was also time-consuming. Although the proposed approach utilized multilevel features (including high-level semantic features), the structure of the constructed CNN was simple, plus the dimension of the semantic features was much less than that obtained from "fc8," which made it require less run time than the AlexCD and COAlexCD. Compared with SiamCD, it was more efficient under the condition of fewer iterations. As the number of iterations increases, multiple independent training of the CNN and SVM models led to a gradual decrease in efficiency.

## VI. CONCLUSION

A novel change detection approach based on low-level to high-level features with limited labels was proposed in this article to better address the change detection task. The experimental results of three datasets with spatial resolutions from medium to high (30, 10, and 2 m) proved its superiority and universality than those state-of-art methods in multispectral image change detection, demonstrating that multilevel change information could comprehensively characterize change categories in different scales. In addition, the introduction of active learning in the proposed approach could not only select the most informative samples for training the change detector model, but also iteratively optimize the scene-level features simultaneously to improve the change detection results. The qualitative and quantitative results indicated that the proposed approach outperforms the most widely used change detection methods with better overall accuracy, kappa coefficient, commission, omission, and total error, providing better visual effect with more clear boundaries and shapes close to the reality. Additionally, due to the fact that stopping criterion of the optimization iteration can be manually customized, it is feasible to improve the precision degree of the final result by increasing the number of iterations/running time or reducing the convergence condition. From these points of views, the proposed approach has the potential of being an effective and efficient way for change detection. The future work of this research would focus on extending the supervised framework to an unsupervised framework. In addition, we hope to design more useful and efficient active learning measures to select suitable samples for change detection.

## ACKNOWLEDGMENT

The authors would like to thank Matthew Senyshen from Queen's University for his advice to improve presentation of the manuscript. The first author would like to express his gratitude to the China Scholarship Council for supporting his stay at Queen's University, Kingston, ON, Canada. Last but not least, they would like to thank the Associate Editor who handled their article and the two anonymous Reviewers for providing truly insightful comments and suggestions that were significantly helpful for them to improve the quality of their article. This work was conducted while the first author was a Visiting Ph.D. student with the Department of Geography and Planning, Queen's University.

## REFERENCES

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

[2] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004.

[3] Z. Zhu, "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 370–384, 2017.

[4] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review ArticleDigital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.

[5] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.

[6] M. Reba and K. C. Seto, "A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111739.

[7] V. Akbari, A. P. Doulgeris, and T. Eltoft, "Monitoring glacier changes using multitemporal multipolarization SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3729–3741, Jun. 2014.

[8] M. Wulder, J. White, F. Alvarez, T. Han, J. Rogan, and B. Hawkes, "Characterizing boreal forest wildfire with multi-temporal Landsat and LIDAR data," *Remote Sens. Environ.*, vol. 113, no. 7, pp. 1540–1555, 2009.

[9] Z. Zhu, C. E. Woodcock, and P. Olofsson, "Continuous monitoring of forest disturbance using all available Landsat imagery," *Remote Sens. Environ.*, vol. 122, pp. 75–91, 2012.

[10] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4363–4378, Aug. 2015.

[11] Z. Li, W. Shi, P. Lu, L. Yan, Q. Wang, and Z. Miao, "Landslide mapping from aerial photographs using change detection-based Markov random field," *Remote Sens. Environ.*, vol. 187, pp. 76–90, 2016.

[12] L. S. Guild, W. B. Cohen, and J. B. Kauffman, "Detection of deforestation and land conversion in Rondonia, Brazil using change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 4, pp. 731–750, 2004.

[13] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection with kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1026–1030, Nov. 2012.

[14] F. Pacifici, F. Del Frate, C. Solimini, and W. J. Emery, "An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 9, pp. 2940–2952, Sep. 2007.

[15] P. Serra, X. Pons, and D. Sauri, "Post-classification change detection with data from different sensors: Some accuracy considerations," *Int. J. Remote Sens.*, vol. 24, no. 16, pp. 3311–3340, 2003.

[16] L. Wan, Y. Xiang, and H. You, "A post-classification comparison method for SAR and optical images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.

[17] P. Kempeneers, F. Sedano, P. Strobl, D. O. McInerney, and J. San-Miguel-Ayanz, "Increasing robustness of postclassification change detection using time series of land cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3327–3339, Sep. 2012.

[18] 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 Observ. Geoinf.*, vol. 20, pp. 77–85, 2013.

[19] K. Chen, Z. Zhou, C. Huo, X. Sun, and K. Fu, "A semisupervised context-sensitive change detection technique via Gaussian process," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 236–240, Mar. 2013.

[20] L. An, M. Li, P. Zhang, Y. Wu, L. Jia, and W. Song, "Discriminative random fields based on maximum entropy principle for semisupervised SAR image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3395–3404, Aug. 2016.

[21] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jun. 2009.

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

[23] N. Falco, M. Dalla Mura, F. Bovolo, J. A. Benediktsson, and L. Bruzzone, "Change detection in VHR images based on morphological attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, May 2013.

[24] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multiscale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, Sep. 2017.

[25] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[26] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, 2013.

[27] X. Wang, S. Liu, P. Du, H. Liang, J. Xia, and Y. Li, "Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 276.

[28] G. Chen, G. J. Hay, L. M. Carvalho, and M. A. Wulder, "Object-based change detection," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4434–4457, 2012.

[29] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state-of-the-art," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[30] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.

[31] M. Volpi and D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, Feb. 2017.

[32] Y. Liu, B. Fan, L. Wang, J. Bai, S. Xiang, and C. Pan, "Semantic labeling in very high resolution images via a self-cascaded convolutional neural network," *ISPRS J.Photogramm. Remote Sens.*, vol. 145, pp. 78–95, 2018.

[33] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[34] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.

[35] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[36] E. Li, J. Xia, P. Du, C. Lin, and A. Samat, "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5653–5665, Oct. 2017.

[37] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[38] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[39] 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.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[40] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[41] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.
[42] M. Pedergnana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.
[43] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.
[44] E. Aptoula, M. C. Ozdemir, and B. Yanikoglu, "Deep learning with attribute profiles for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1970–1974, Dec. 2016.
[45] A. Taghipour and H. Ghassemian, "Hyperspectral anomaly detection using attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1136–1140, Jul. 2017.
[46] M. Imani, "Attribute profile based target detection using collaborative and sparse representation," *Neurocomputing*, vol. 313, pp. 364–376, 2018.
[47] X. Wang, P. Du, S. Liu, G. Lu, and X. Gao, "Change detection in high-resolution images based on feature importance and ensemble method," *Arabian J. Geosci.*, vol. 12, no. 14, 2019, Art. no. 446.
[48] C. Kwan, B. Ayhan, J. Larkin, L. Kwan, S. Bernabé, and A. Plaza, "Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (EMAPs)," *Remote Sens.*, vol. 11, no. 20, 2019, Art. no. 2377.
[49] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[50] T. Blaschke *et al.*, "Geographic object-based image analysis—Towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 180–191, 2014.
[51] G. J. Hay, T. Blaschke, D. J. Marceau, and A. Bouchard, "A comparison of three image-object methods for the multiscale analysis of landscape structure," *ISPRS J. Photogramm. Remote Sens.*, vol. 57, nos. 5/6, pp. 327–345, 2003.
[52] J.-M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: A stepwise optimization approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 2, pp. 150–163, Feb. 1989.
[53] Q. Yu and D. A. Clausi, "IRGS: Image segmentation using edge penalties and region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2126–2139, Dec. 2008.
[54] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[55] B. Desclée, P. Bogaert, and P. Defourny, "Forest change detection by statistical object-based method," *Remote Sens. Environ.*, vol. 102, nos. 1/2, pp. 1–11, 2006.
[56] T. Su, "Efficient paddy field mapping using Landsat-8 imagery and object-based image analysis based on advanced fractel net evolution approach," *GISci. Remote Sens.*, vol. 54, no. 3, pp. 354–380, 2017.
[57] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, 2017.
[58] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
[59] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state-of-the-art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[60] P. Ghamisi, M. S. Couceiro, and J. A. Benediktsson, "A novel feature selection approach based on FODPSO and SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2935–2947, May 2015.
[61] L. Drăguţ, D. Tiede, and S. R. Levick, "ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *Int. J. Geographical Inf. Sci.*, vol. 24, no. 6, pp. 859–871, 2010.
[62] L. Drăguţ, O. Csillik, C. Eisank, and D. Tiede, "Automated parameterisation for multi-scale image segmentation on multiple layers," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 119–127, 2014.
[63] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
[64] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[65] X. Fu, J. Huang, X. Ding, Y. Liao, and J. J. I. T. o. I. P. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017.
[66] Y. Del Valle, G. K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, and R. G. Harley, "Particle swarm optimization: Basic concepts, variants and applications in power systems," *IEEE Trans. Evol. Comput.*, vol. 12, no. 2, pp. 171–195, Apr. 2008.
[67] G. Mountrakis, J. Im, C. J. I. J. o. P. Ogole, and R. Sensing, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
[68] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *Int. J. Comput. Vis.*, vol. 113, no. 2, pp. 113–127, 2015.
[69] W. Cai, M. Zhang, and Y. Zhang, "Batch mode active learning for regression with expected model change," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1668–1681, Jul. 2017.
[70] N. Roy and A. McCallum, "Toward optimal active learning through Monte Carlo estimation of error reduction," in *Proc. Int. Conf. Mach. Learn.*, 2001, pp. 441–448.
[71] J. Im, J. Jensen, and J. Tullis, "Object-based change detection using correlation image analysis and image segmentation," *Int. J. Remote Sens.*, vol. 29, no. 2, pp. 399–423, 2008.
[72] J. R. Townshend, C. O. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 1054–1060, Sep. 1992.
[73] J. Chen, J. Xia, P. Du, and J. Chanussot, "Combining rotation forest and multiscale segmentation for the classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4060–4072, Sep. 2016.
[74] C. Zhu, S. Zhang, J. Plaza, J. Li, and A. Plaza, "Impervious surface extraction from multispectral images via morphological attribute profiles based on spectral analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4775–4790, Dec. 2018.
[75] N. Débonnaire, A. Stumpf, and A. Puissant, "Spatio-temporal clustering and active learning for change classification in satellite image time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3642–3650, Aug. 2016.
[76] M. Rezaee, M. Mahdianpari, Y. Zhang, and B. Salehi, "Deep convolutional neural network for complex wetland classification using optical remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3030–3039, Sep. 2018.
[77] E. Li, A. Samat, W. Liu, C. Lin, and X. Bai, "High-resolution imagery classification based on different levels of information," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2916.
[78] K. Nogueira, O. A. Penatti, and J. A. Dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.
[79] A. M. El Amin, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images," in *Proc. 1st Int. Workshop Pattern Recognit.*, 2016, Art. no. 100110W.
[80] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.

**Xin Wang** received the B.S. degree in geographic information system from Lanzhou University, Lanzhou, China, in 2016. He is currently working toward the Ph.D. degree in cartography and geographic information system at Nanjing University, Nanjing, China.

He is currently a Visiting Ph.D. Student with the Department of Geography and Planning, Queen's University, Kingston, ON, Canada. His research interests include multitemporal image processing, change detection, signal processing, and pattern recognition.

**Peijun Du** (Senior Member, IEEE) received the Ph.D. degree from China University of Mining and Technology, Xuzhou, China, in 2001.

He is currently a Professor of Remote Sensing and Geographic Information Science with Nanjing University, Nanjing, China. He was a Senior Visiting Scholar with the University of Nottingham, Nottingham, U.K. and Grenoble Institute of Technology, Grenoble, France. He has authored/coauthored more than 120 articles in international peer-reviewed journals, with more than 60 published in IEEE TGRS/JSTARS/GRSL and ISPRS *Journal of Photogrammetry and Remote Sensing*. His research interests include advanced image processing and machine-learning techniques, including support vector machine, ensemble learning, sparse representation, and active learning, for remote sensing image classification, change detection, environmental, and urban remote sensing.

**Dongmei Chen** received the B.A. degree in geography from Peking University, Beijing, China, the master's degree in GIS and remote sensing application from the Institute of Remote Sensing Application, Chinese Academic of Science, Beijing, China, and the Ph.D. degree in geography from the Joint Doctoral Program, San Diego State University, San Diego, CA, USA, and the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2001.

She is currently a Professor with Queen's University, Kingston, ON, Canada. Her research interests include the understanding and modeling of interactions between human activities and the physical environment by using GIS and remote sensing techniques and spatial modeling approaches from local to regional scales.

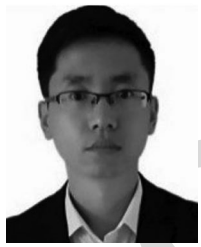
**Sicong Liu** (Member, IEEE) received the B.S. degree in geographical information system and the M.E. degree in photogrammetry and remote sensing from China University of Mining and Technology, Xuzhou, China, in 2009 and 2011, respectively, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, in 2015.

He is currently an Assistant Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His research interests include multitemporal remote sensing data analysis, change detection, multispectral and hyperspectral remote sensing, signal processing, and pattern recognition.

Dr. Liu was the recipient (ranked as third place) of the Paper Contest of the 2014 IEEE GRSS Data Fusion Contest. He is the Technical Cochair of the 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp 2019). He has been the Session Chair for many international conferences such as the International Geoscience and Remote Sensing Symposium. He is also a Referee for more than 20 international journals.

**Wei Zhang** received the M.S. degree in geodesy and survey engineering from China University of Mining & Technology, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in cartography and geographic information system at the School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing, China.

His research interests include image classification, machine learning, and heavy metal pollution monitoring with hyperspectral remote sensing.



**Erzhu Li** received the M.S. degree in photogrammetry and remote sensing from China University of Mining and Technology, Xuzhou, China, in 2014, and the Ph.D. degree in cartography and geographic information system from Nanjing University, Nanjing, China, in 2017.

He is currently a Lecturer and a Researcher with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, China. His research interests include high-resolution image processing and computer vision in urban remote sensing applications.