This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

SPCNet: A Subpixel Convolution-based Change Detection Network for Hyperspectral Images with Different Spatial Resolutions

Lifeng Wang, Liguo Wang, Heng Wang, Xiaoyi Wang, and Lorenzo Bruzzone

***Abstract*— The very high spectral resolution in hyperspectral images (HSIs) offers an opportunity to detect subtle land-cover changes. However, the availability of HSIs acquired from different platforms requires the development of change detection (CD) methods capable of processing HSIs with different spatial resolutions. In this paper, we propose a general end-to-end subpixel convolution-based residual network (SPCNet) to accomplish the CD task between high spatial resolution (HR) and low spatial resolution (LR) HSIs. To effectively tackle the resolution matching issue, a super resolution (SR) block with an efficient subpixel convolution layer is introduced to upscale the LR feature maps into HR maps. The subpixel convolution layer can fully explore the subpixel context information by learning an array of upscaling filters. Moreover, the designed SPC module is embedded into the LR branch to generate more discriminative representations. More importantly, the SPC module as a plug-and-play unit has the potential to be embedded into other baseline networks to enhance the feature learning capability. Experimental results on four HSI datasets demonstrate the effectiveness of the proposed SPCNet.**

***Index Terms*— Change detection, deep learning, multiscale images, residual network, subpixel convolution, hyperspectral images, remote sensing.**

I. INTRODUCTION

HSIs are valuable data sources to monitor global changes for Earth observation [1]. CD methods can recognize the differences in multitemporal HSIs acquired over the same geographical area at multiple times [2], [3]. They can be extensively exploited in many research activities, such as forestry and agriculture monitoring [4], [5], land sprawl dynamic analysis [6], [7], [8] and natural disaster damage assessment [9], [10], [11].With the successful launch of satellites such as NASA Earth Observation-1 (EO-1), the Chinese Gaofen-5 satellites, and the Italian Space Agency PRISMA, the availability of HSIs at a global scale has significantly increased. Accordingly, CD in HSIs has received increasing attention in recent years.

Over the past few decades, many conventional methods have been proposed to detect changes in bitemporal HSIs. In general, these methods can be divided into three main categories: algebra-based methods, transformation-based methods, and post-classification comparison methods. Typical algebra-based methods, such as image subtraction [12], image ratio [13], and change vector analysis (CVA) [14], are sensitive to radiation and noise, which compromises the CD accuracy. The representative transformation-based methods include iteratively reweighed multivariate alteration detection (IR-MAD) [15], slow feature analysis (SFA) [16], and principal component analysis (PCA) [17]. Although transformation-based methods can transform high-dimensional HSIs into other feature spaces and thus reduce the redundancy, they have a drawback in the need to select appropriate threshold values to detect land-cover changes. With regard to post-classification comparison methods [18], [19], the class maps generated by the classification of two images independently are compared pixel-by-pixel to obtain the changed regions [20], [21], [22]. A common drawback of post-classification comparison methods is that the CD result is directly affected by the propagation of classification errors on both images.

Recently, deep learning-based methods have been introduced for addressing the HSI task [23]. In [24], a Caps-TripleGAN framework is proposed by exploring the 1-D structure triple generative adversarial network (TripleGAN) for sample generation and integrating CapsNet for HSI classification. Wang et al. [25] proposed a unified multiscale learning framework, which contains a multiscale spatial-channel attention mechanism and a multiscale shuffle block to improve the problem of land-cover map distortion in HSI classification. Wang et al. [26] proposed an SSA-SiamNet method to adaptively extract features from input patch pairs of HSIs for CD. In [27], a multilevel encoder-decoder attention network is proposed, which fuses the hierarchical features from all of the convolutional layers to detect changes in the HSIs. In [28], a three-direction CNN (TDSSC) that contains a change tensor that decomposes along the spectral direction and two spatial directions, is developed for HSI CD. In [29], the recurrent three-dimensional (3D) fully convolutional network (Re3FCN) merges the superiorities of a 3D fully convolutional network and a convolutional long short-term memory. It is developed to extract joint spectral-spatial-temporal features. Wang et al. [30] proposed a general end-to-end two-dimensional-CNN framework (GETNET), which uses a mixed-affinity matrix as the input to detect binary changes in HSIs. Nonetheless, the abovementioned methods were developed for bitemporal HSIs with the same spatial resolution and do not sufficiently consider the use of subpixel context information.

Manuscript received XXX. This work was supported by the National Natural Science Foundation of China under Grants 62071084. (Corresponding author: Liguo Wang)

Lifeng Wang, Liguo Wang, Heng Wang, and Xiaoyi Wang are with the College of Information and Communication Engineering, Harbin Engineering University, 145 Nantong Avenue, Harbin 150000, China (e-mail: wanglifeng_2016 @163.com).

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

In general, conventional and deep learning-based CD methods are applied to bitemporal HSIs acquired by the same sensors with the same spatial resolution. However, they are not adequate to address CD when HSIs acquired by different sensors with different spatial resolutions are considered [31]. For example, this case occurs if an HR HSI of a certain region is available at time T1 but only an LR HSI of the same area is available at time T2. In these cases, it is necessary to design CD methods that can effectively analyze bitemporal HSIs with different spatial resolutions. These methods should address two main challenges: spatial resolution matching and design of specific CD algorithms.

To address the spatial resolution matching issue, the most intuitive method is downsampling the HR HSI to make LR images, or interpolating the LR HSI to the HR HSI to obtain two HSIs with the same spatial resolution. Then, common CD methods can be applied to detect changes [31]. However, on the one hand, the downsampling step leads to a lack of detailed spatial information. On the other hand, common interpolation methods, such as linear, bilinear, and bicubic, are not able to bring additional spatial information to solve the reconstruction problem. Therefore, it is necessary to develop information reconstruction strategies that make full use of subpixel information to detect land cover changes in bitemporal HSIs with different spatial resolutions.

Different from the abovementioned interpolation methods, subpixel-based methods have been explored to accomplish the CD task for remote sensing images. For example, Liu et al. [32] proposed an SR-based CD network that combines adversarial learning and stacked attention modules for CD on high-resolution images with different spatial resolutions. In [33], a novel approach that combines the subpixel mapping based on a backpropagation neural network and soft classification was proposed to address subpixel land-cover change detection for multispectral images with different spatial resolutions. In [34], an iterative method to detect changes for multispectral images at different spatial scales was proposed. It involves endmember estimation, spectral unmixing, land-cover fraction change detection, and super resolution land-cover mapping. However, when considering the problem of CD in bitemporal HSIs, to the best of our knowledge, the case of images with different spatial resolutions has seldom been considered.

Based on the abovementioned analysis, we propose a subpixel convolution-based residual network (SPCNet) for CD in bitemporal HSIs with different spatial resolutions. In greater detail, the main contributions of this paper are as follows:

1) A general end-to-end SPCNet framework is proposed to solve the CD problem in bitemporal HSIs with different spatial resolutions, which can fully extract the subpixel information from LR HSIs.
2) An SR block with a subpixel convolution operator, which learns an array of upscaling filters, is introduced to upscale the LR feature maps into HR images. It fully exploits the subpixel context information and generates more discriminative representations.
3) The SPC module is designed as a plug-and-play unit and has the potential to be embedded into other baseline networks to enhance the feature learning capability in other HSI processing tasks.

The remainder of this paper is organized as follows. Section II describes the proposed SPCNet method in detail. Section III reports the experimental results of the proposed method. Section IV discusses further issues and draws the conclusions of the paper.

II. Methodology

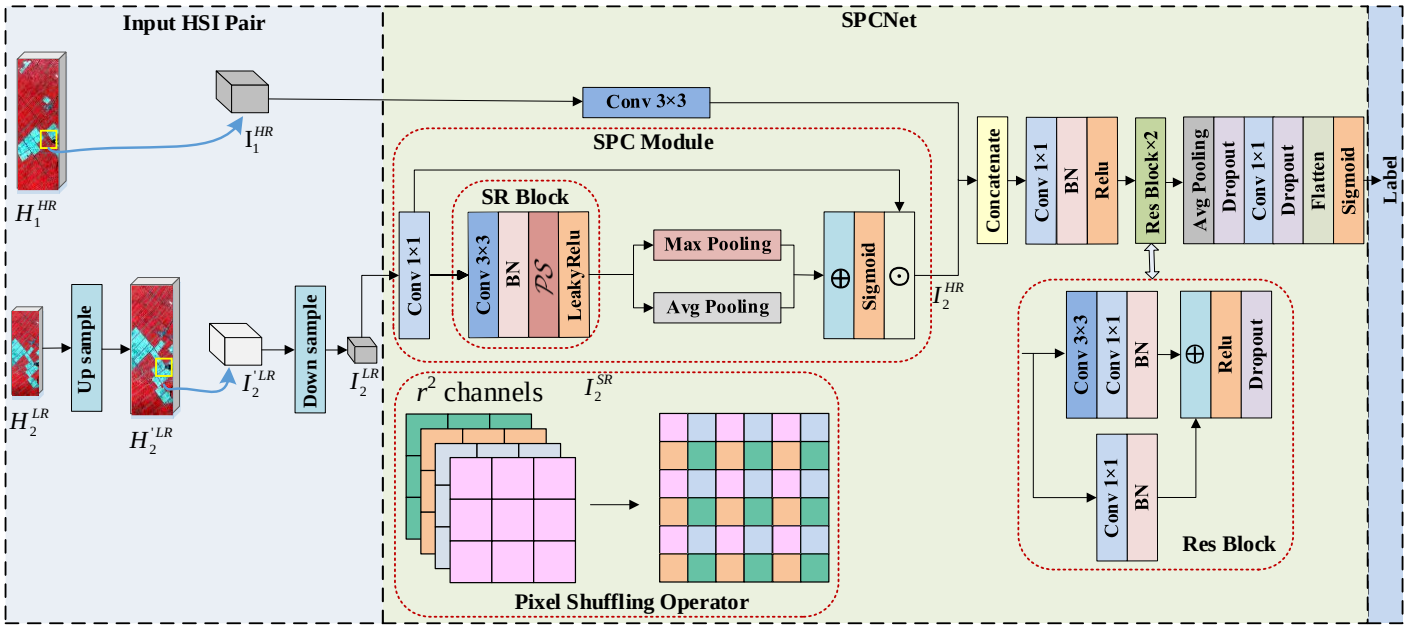


Fig. 1. Overview of the proposed SPCNet for CD on HSIs with different spatial resolutions.

Fig.1 shows an overview of the proposed end-to-end CD framework based on the SPCNet to detect land-cover changes in HSIs with different spatial resolutions. First, modified residual blocks are stacked to form an improved ResNet, which is the

baseline to detect land-cover changes. Second, a pixel shuffling operator is embedded to form the SR block, which upscales the LR feature maps into HR maps. Third, the SPC module is designed to generate more discriminative subpixel-based representations from LR feature maps. Finally, the subpixel-based features of the LR HSI and the pixel-based features of the HR HSI are concatenated and given as input to the modified ResNet.

The four main parts of the proposed SPCNet method, i.e., the SR block, the SPC module, the Res block and ResNet, as well as the loss function and the optimization strategy, are introduced in the next subsections.

A. SR Block

The two HSIs with different spatial resolutions acquired at times T1 and T2 are represented as H_1^{HR} and H_2^{LR}, respectively. Let us assume that the same number b of spectral bands is associated with both H_1^{HR} and H_2^{LR}. The spatial sizes of H_1^{HR} and H_2^{LR} are $h_1 \times w_1$ and $h_2 \times w_2$, respectively, where $h_2 \le h_1$ and $w_2 \le w_1$. H_2^{LR} is interpolated as follows:

$$H_2^{'LR} = up(H_2^{LR}) \tag{1}$$

where $up(\cdot)$ is the bicubic interpolation operator and the size of $H_2^{'LR}$ is $h_1 \times w_1 \times b$. To make full use of the spatial context information, a set of patches from H_1^{HR} and $H_2^{'LR}$ is constructed by combining a center pixel and its s surrounding pixels, thus producing patches I_1^{HR} and $I_2^{'LR}$ with size $s \times s \times b$. To reduce the interference of redundant information caused by the interpolation strategy and the related processing, $I_2^{'LR}$ is transformed as follows:

$$I_2^{LR} = down(I_2^{'LR}) \tag{2}$$

where $down(\cdot)$ is downsampling operation, which exploits the same strategy as that used in the interpolation operation to restore the original spectral information as much as possible. The patch size of I_2^{LR} is $d \times d \times b$ and $d \le s$.

To fully extract subpixel context information, a subpixel convolution layer is introduced to upscale the LR feature maps $\mathbf{I}_2^{LR}$ into the HR output features into $\mathbf{I}_2^{SR}$ [35][35]. The subpixel convolution layer can be expressed as follows:

$$\mathbf{I}_2^{'SR} = f^L(\mathbf{I}_2^{LR}) = \mathcal{PS}(W_L * f^{L-1}(\mathbf{I}_2^{LR}) + b_L) \tag{3}$$

where f^L denotes the L-th network layer, and W_L, b_L are learnable network weights and biases, respectively. $\mathcal{PS}$ represents a periodic shuffling operator, which rearranges the elements of a $d \times d \times c \times r^2$ tensor to a tensor with $rd \times rd \times c$ shape. Moreover, r is an upscale factor that is set to 2 for experiments in Section III. Mathematically, the $\mathcal{PS}$ operation can be described as follows:

$$\begin{aligned}\mathcal{PS}(T)_{x,y,c_i} = T\lfloor x/r \rfloor, \lfloor y/r \rfloor, c_i \cdot r \cdot \mathrm{mod}(y,r) \\ + c_i \cdot \mathrm{mod}(x,r)\end{aligned} \tag{4}$$

These patterns are periodically activated during the convolution of the filter across the patch depending on different subpixel locations: $\mathrm{mod}(x,r)$, $\mathrm{mod}(y,r)$, where x, y are the output pixel coordinates in the HR space. $\mathrm{mod}(\cdot)$ represents the modulo operator, which returns the remainder of dividing two numbers, and c_i is the i-th filter channel.

In the proposed method, the SR block is designed to extract effective subpixel features and includes a convolutional layer, a batch normalization (BN) operation, a $\mathcal{PS}$ operation, and a LeakyReLU activation function. The specific arrangement of the SR block is shown in Fig. 1 and can be calculated as:

$$\mathbf{I}_2^{SR} = \sigma(\mathcal{PS}(g(f(\mathbf{I}_2^{LR})))) \tag{5}$$

where $\sigma(\cdot)$ represents the LeakyReLU activation function, and $g(\cdot)$ denotes the transform of the convolution and BN operations.

B. SPC Module

The proposed SPC module is arranged according to Fig. 1. First, the intermediate subpixel features are summarized as follows:

$$\mathbf{I}_2^{SR} = \mathcal{F}(f(\mathbf{I}_2^{LR})) \tag{6}$$

where $f(\cdot)$ represents the convolution operation, and $\mathcal{F}(\cdot)$ is the series of operations of the SR block. Then, the subpixel-based features $\mathbf{I}_2^{SR}$ can be transformed as follows:

$$\mathbf{I}_2^{HR} = f(\mathbf{I}_2^{LR}) \cdot \sigma(MaxPool(\mathbf{I}_2^{SR}) + AvgPool(\mathbf{I}_2^{SR})) \tag{7}$$

where $MaxPool(\cdot)$ and $AvgPool(\cdot)$ represent the max-pooling and avg-pooling operations, respectively, and "$\cdot$" denotes elementwise multiplication. Note that this designed SPC module with auxiliary subpixel context information can be extended directly to the other frameworks.

C. Res Block and ResNet

ResNet has been introduced to robustly improve the performance of CNNs with a very deep architecture. It is an extension of the CNN with shortcut connections that promotes the propagation of gradients. Many residual units are stacked to form deep residual networks to avoid gradient extinction [36].

In our approach, the residual unit is modified specifically for the HSI CD task. Instead of the identity shortcut connection present in the traditional residual unit, in the designed residual block a connection with 1×1 convolutional and BN operations is employed. More specifically, a dropout layer is adopted in the modified residual unit to avoid the overfitting problem. The improved residual unit can be described as follows:

$$h^{l+1} = G(\mathrm{g}(f(h^l))) + \mathrm{g}(h^l)) \tag{8}$$

where h^l and h^{l+1} are the input and output features of the lth unit, respectively. $G(\cdot)$ represents the ReLU activation function and dropout operation.

As shown in Fig. 1, the subpixel-based features of the LR branch and the pixel-based features of the HR branch are concatenated to be given as input to the designed ResNet. In this study, the CD problem is modeled as a binary classification task. Thus, a fully connected layer, including two filters and a sigmoid activation function, is adopted as the end of the proposed architecture. It provides as outputs a binary label that indicates whether the pixel has changed or not.

D. Loss Function and Optimization

An appropriate loss function is crucial to optimizing the designed network during model training. The cross entropy function can measure the dissimilarity between the ground-truth labels and the predicted probabilities. Therefore, binary cross entropy is commonly utilized for model training in binary classification [37]. Accordingly, the binary cross entropy is adopted as the loss function of the proposed SPCNet method. It is defined as follows:

$$Loss_{BCE} = -\frac{1}{M}\sum_{i=1}^{M}[y_i \times \log(\hat{y}_i) + (1-y_i) \times \log(1-\hat{y}_i)] \quad (9)$$

where M represents the number of patch pairs in a mini-batch during the model training. y_i and $\hat{y}_i$ denote the ground-truth label and the predicted probability for the i-th patch pair, respectively. The labels of the unchanged and changed patch pairs are set to $y_i = 0$ and $y_i = 1$, respectively. Stochastic gradient descent and backpropagation are applied to update the parameters in the designed network.

III. Experiments

A. Description of Datasets

To assess the performance of the proposed SPCNet method, four HSI datasets were used in the experiments: Farmland, River, Hermiston, and Bay Area. The first three datasets were acquired by using the Hyperion sensor onboard the EO-1 satellite. The Hyperion sensor covers wavelengths that range from 0.4 to 2.5 μm with 242 spectral bands. It provides HSIs with a spectral resolution of approximately 10 nm and a spatial resolution of approximately 30 m. The Bay Area dataset was acquired by an Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) from NASA's Jet Propulsion Laboratory (JPL). AVIRIS covers the spectral range from 0.4 to 2.5 μm with 224 spectral bands. The AVIRIS HSIs have a spatial resolution of 20 m and a spectral resolution of 10 nm.

The first dataset, named Farmland, covers a farmland area in Yancheng, Jiangsu Province, China. The bitemporal HSIs were acquired on May 3, 2006, and April 23, 2007. The spatial extent is 450×140 pixels, and 155 bands are selected for CD after removing noisy bands. Crop rotation represents the main changes in this dataset.

The second dataset, named River, contains a river region in Jiangsu Province, China. The bitemporal HSIs were acquired on May 3, 2013, and December 31, 2013. The spatial size is 463×241 pixels, and 198 bands are considered after the removal of noisy bands. The changes in the dataset are mainly related to the removal of sediment from the river.

The third dataset, named Hermiston, encompasses an irrigated agricultural field in Hermiston City, Umatilla Country, Oregon, USA. The pair of HSIs was acquired on May 1, 2004, and May 8, 2007. The images have a size of 307×241 pixels and contain 154 bands after noisy band removal. The changes in this dataset are caused mainly by the regulation of irrigation areas.

The fourth dataset, named Bay Area, covers an area in the San Francisco Bay Area, California, USA. The bitemporal HSIs were acquired in 2013 and 2015. The image size is 600×500 pixels and the number of spectral bands is 224.

The bitemporal HSIs have the same spatial resolution in the above four datasets. However, bitemporal HSIs with different spatial resolutions are needed for subsequent experiments. Therefore, we first downsample the original HR image in T2 to simulate the corresponding LR version of the T2 image. The scale factor is the spatial resolution ratio between the LR and HR image pairs. The original HR T1 image and the corresponding LR T2 image form the multiresolution dataset. Examples of true-color compositions of three spectral channels of images with different spatial resolutions for the four datasets are shown in Fig. 2. The use of these downsampled images has some limitations but allows us to accurately assess the performance of the proposed SPCNet and compare it with the performances of other techniques.

B. Experimental Setup

The optimizer is based on the root mean square propagation (RMSprop). The number of total epochs is 200, and the initial learning rate is set to 0.001. The batch size is 64 for the Farmland and Bay Area datasets and 128 for the River and Hermiston datasets. The learning rate decays by a factor of 10 every 50 epochs. An NVIDIA GeForce RTX3090 GPU was employed to conduct the experiments. The proposed SPCNet was implemented on the TensorFlow and Keras framework.

For a fair comparison, we randomly select 5% of the samples from the Farmland, River, and Hermiston datasets and 1% of the samples from the ground references of the Bay Area dataset as training sets. The remaining samples are used as test sets. In the experiments, all of the input data were standardized to zero mean values with unit variance, and 10 independent tests were performed to avoid biased estimation.

To quantitatively assess the detection performance, we adopted the overall accuracy (OA), kappa coefficient (Kappa), precision (Pr), recall (Re), and F1-score (F1) as evaluation metrics [26].

C. Parameter Setting

In the proposed SPCNet method, several hyperparameters can affect the training model and, further, the CD results. They include the kernel numbers of convolutional filters, the patch size, the training sample proportion, and the scale factor. In this section, these hyperparameters are analyzed in detail. The impact of a certain parameter was explored by fixing the values of the other parameters.

1) Impact of the kernel numbers

The kernel numbers of the convolutional filters are related to the representation capacity and the computational cost of the SPCNet method. In the proposed SPCNet, the kernel number in each convolutional layer is the same except for the SR block. Different kernel numbers (i.e., 4, 8, 16, 24, 32, 64) were examined in preliminary experiments. As shown in Fig. 3(a), as the kernel numbers increase from 4 to 32, the OA values also increase in all four datasets. Compared with 32 kernels, 64 kernels increase the model complexity, but the accuracy is not significantly improved. Thus, the kernel number is fixed to 32 for all four datasets.

2) Impact of the patch size

The size of the input patch pair is related to the amount of spatial information surrounding the center pixel. To evaluate the influence of different patch sizes on the SPCNet, we considered

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

a set of four patch sizes: 5×5, 7×7, 9×9, and 11×11. As shown in Fig. 3(b), as the patch size increases, the OA value decreases gradually in all of the datasets except for the Bay Area dataset. In our study, the patch size was set to 5×5 in all datasets.

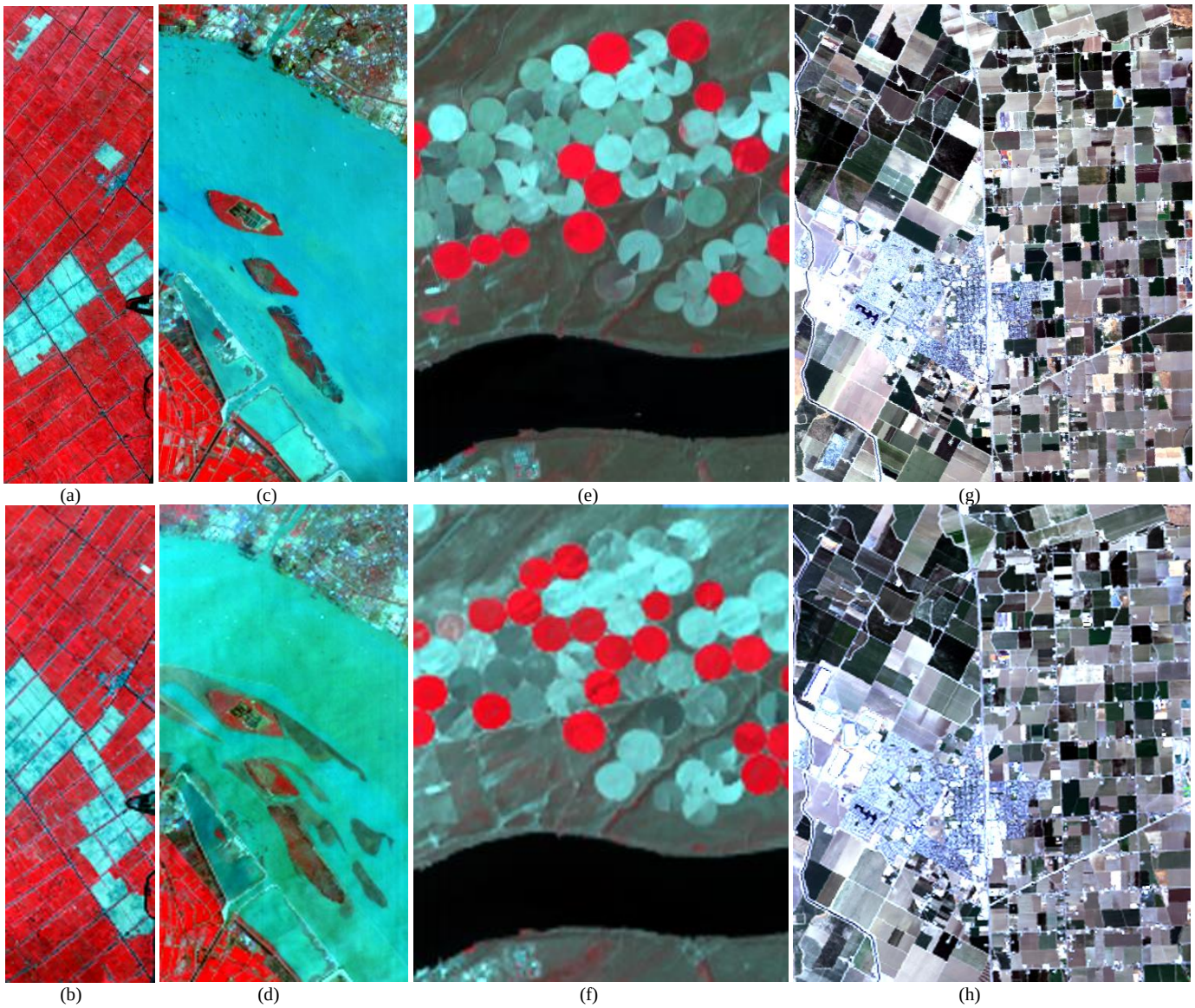


Fig. 2. The four HSI datasets used in the experiments (bands 33, 22, and 11 as RGB). (a) HR Farmland image at time T1 and (b) LR Farmland image at time T2. (c) HR River image at time T1 and (d) LR River image at time T2. (e) HR Hermiston image at time T1 and (f) LR Hermiston image at time T2. (g) HR Bay Area image at time T1 and (h) LR Bay Area image at time T2. The scale factor of the LR image is set to 2.

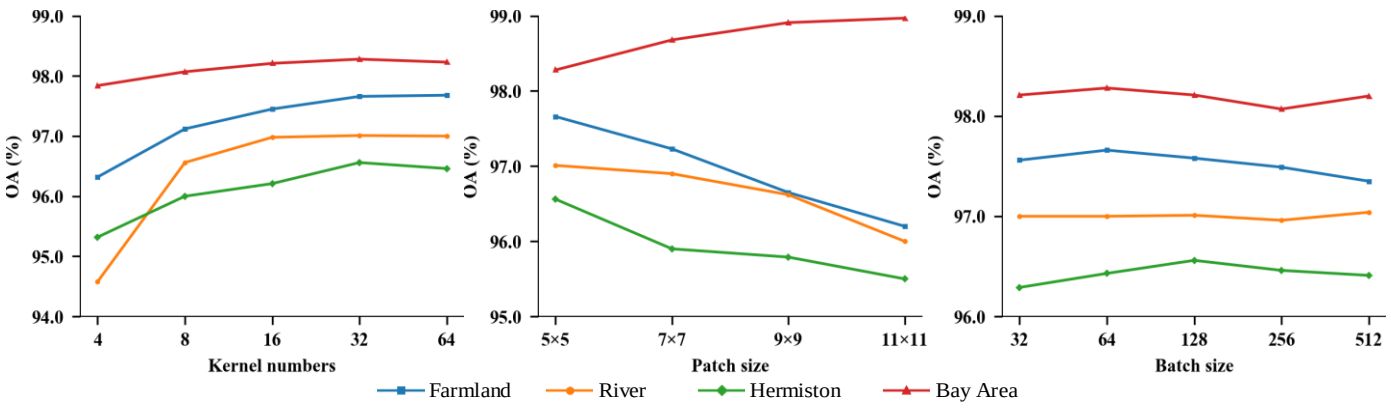


Fig. 3. OA values obtained by the SPCNet with different hyperparameter values for the four datasets: (a) the kernel numbers, (b) the patch size, (c) the batch size.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

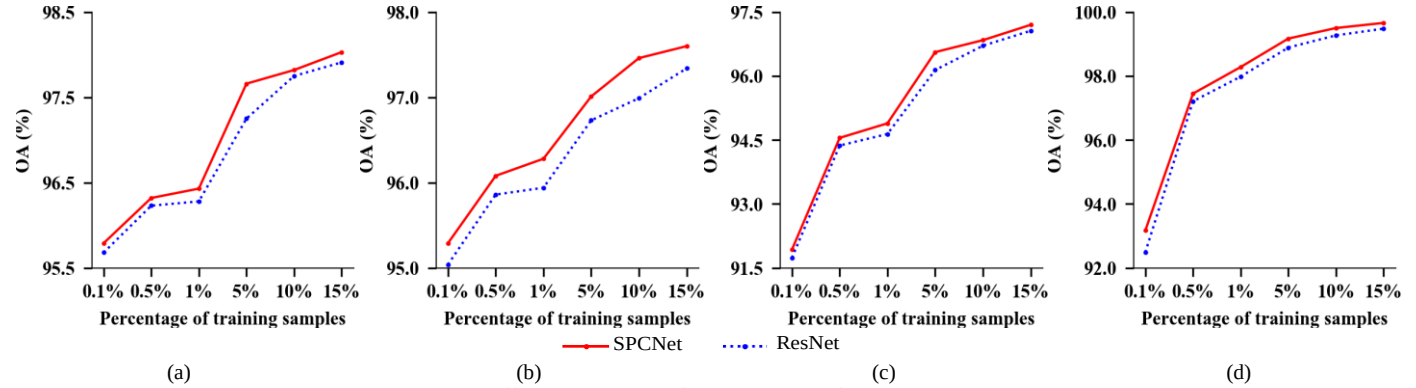


Fig. 4. OA values obtained by ResNet and SPCNet with different proportions of training samples for (a) the Farmland dataset, (b) the River dataset, (c) the Hermiston dataset, and (d) the Bay Area dataset.

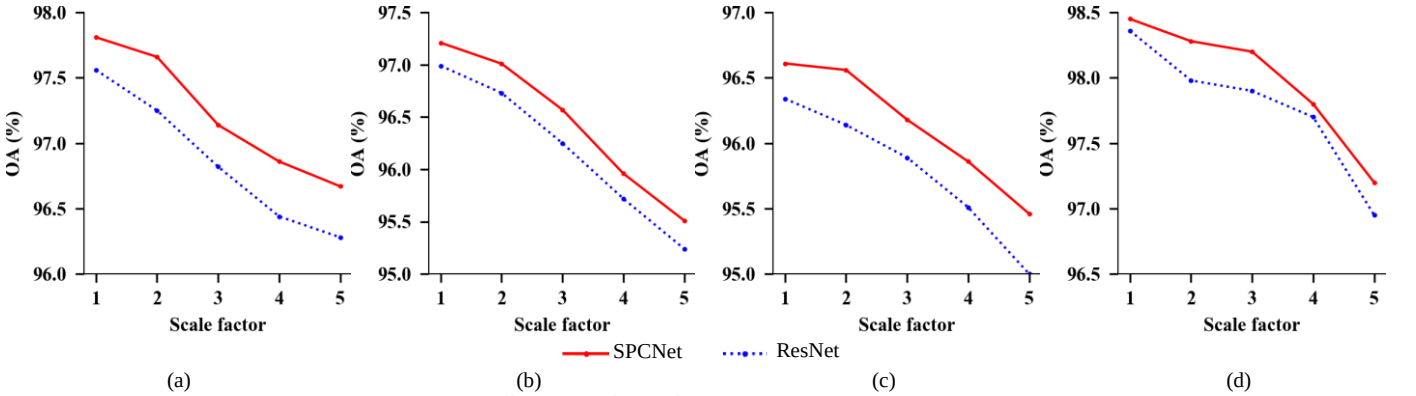


Fig. 5. OA values obtained by ResNet and SPCNet with different scale factors for (a) the Farmland dataset, (b) the River dataset, (c) the Hermiston dataset, and (d) the Bay Area dataset.

3) Impact of the batch size

To explore the influence of the patch size on the proposed SPCNet, a set of batch sizes (32, 64, 128, 256, 512) were examined. From Fig. 3(c), one can observe that the OA values do not change significantly under different batch sizes, which suggests that the proposed SPCNet is robust to the batch size. The optimal batch size is 64 for the Farmland and Bay Area datasets and 128 for the River and Hermiston datasets.

4) Impact of the proportion of training samples

To demonstrate the generality of the proposed SPCNet method, we examined different proportions of training samples (0.1%, 0.5%, 1%, 5%, 10%, and 15%) for ResNet and SPCNet in all four datasets. As shown in Fig.4, the OA values of ResNet and SPCNet increased by increasing the proportion from 0.1% to 15% for the four datasets. In general, the proposed SPCNet method improves the OA value compared with ResNet.

5) Impact of the scale factor

To evaluate the effectiveness of the proposed SPCNet method, we considered a set of scale factors (1, 2, 3, 4, and 5) between the spatial resolutions of T1 and T2 images. From Fig. 5, one can see that as the scale factor increases, due to the decrease in the spatial resolution of the HSI at T2 time, the OA value decreases gradually in all datasets. As expected, the proposed SPCNet method is effective in detecting changes from bitemporal HSIs with the same spatial resolution (the scale factor is set to 1).

D. Comparison with Other Methods

To verify the effectiveness of the proposed SPCNet method, we compare it with eleven benchmark methods, which include CVA [38], SVM [39], SiamNet [26], SSA-SiamNet [26], GETNET [30], Sub-2DCNN [40], Con-2DCNN [40][40], Con-3DCNN [41], Sub-RSSAN [42], Con-RSSAN [42], and ResNet. In SVM, the radial basis function kernel was selected in the four datasets. The input images for the Sub-2DCNN and the Sub-RSSAN are the result of subtracting the HR image and the interpolated LR image. For Con-based methods (i.e., SVM, Con-2DCNN, Con-3DCNN, Con-RSSAN, and ResNet), the HR image and the interpolated LR image are concatenated and given as input to the model. For deep learning-based methods, the specific parameter settings refer to [26].

1) Quantitative comparison

Table 1 reports the CD accuracy of different methods on the four datasets. By analyzing the table, we can derive the four following observations.

First, the CVA method provides the smallest accuracy for the four datasets, especially for the River and Bay Area datasets. Compared with supervised approaches, CVA is a classic unsupervised method. It does not require any training samples that prevent us from accurately distinguishing various changes.

Second, the accuracies of deep learning-based methods are superior to that of the SVM method on all datasets except for the Farmland dataset. The reason is that different from the SVM that exploits the spectral information, deep learning-based methods can make full use of the spatial-spectral context information and generate more discriminative features. This process facilitates the detection of land-cover changes.

Third, compared with the SiamNet and GETNET methods, other deep learning methods provide metric values with lower standard deviations and are more centralized around the means. This finding indicates the stability of the algorithm.

Finally, the proposed SPCNet enhances the performance of ResNet on all datasets. More precisely, the increases in OAs for the Farmland, River, Hermiston, and Bay Area datasets are 0.41%, 0.28%, 0.42%, and 0.30%, respectively. The reason for these improvements is that the proposed SPCNet extracts more discriminative features through the subpixel convolution operation, which improves the CD accuracy.

2) Qualitative comparison

Ground-reference maps for the four datasets are shown in Panels (m) of Figs. 6-9, where the white, black, and gray parts describe changed, unchanged, and unknown pixels, respectively. Moreover, Figs. 6 to 9 show the CD maps obtained by different methods on the four datasets for a comparative qualitative comparison. Areas that have the main differences in the maps are indicated by red squares and circles. Overall, the CVA maps contain more noise than those of the other supervised methods, which indicates the importance of the training information in detecting changes. Generally, compared with the SVM method, the CD maps of deep learning-based techniques are much more similar to the ground-reference map, which confirms the effectiveness of using the spectral-spatial context information. Furthermore, the CD map produced by the proposed SPCNet method is the closest to the ground-reference maps on all datasets. This aspect is consistent with the quantitative results in Table 1.

3) Computational cost

A number of trainable parameters and floating-point operations (FLOPs) are used to describe the computational cost of the deep learning-based methods. Table 2 lists the computational costs of ten deep learning-based methods. Obviously, GETNET requires the largest number of parameters and FLOPs among the nine methods for the four datasets. The reason is that the input data of GETNET are large patches with size $c \times c \times 1$. Overall, the modified ResNet involves the lowest number of parameters and FLOPs among all methods, which indicates the superiority of the improved ResNet as a baseline. Furthermore, the proposed SPCNet, which improves the CD accuracy, shows an acceptable slight increase in the model complexity.

Table 1 Accuracy of different CD methods on the four considered datasets
(Bold values indicate the highest accuracies for each variable and each dataset)

Dataset	Method	Metrics				
		OA (%)	Kappa×100	F1 (%)	Pr (%)	Re (%)
Farmland	CVA	96.18	90.83	93.54	91.78	95.37
	SVM	97.73	94.49	96.09	96.34	95.84
	SiamNet	96.98±0.29	92.65±0.76	94.77±0.56	95.34±0.37	94.20±1.39
	SSA-SiamNet	97.04±0.36	92.80±0.68	94.87±0.49	95.45±0.24	94.30±0.76
	GETNET	97.56±0.16	94.05±0.42	95.77±0.62	96.38±0.44	95.16±0.51
	Sub-2DCNN	**97.74±0.04**	**94.51±0.09**	**96.10±0.14**	96.07±0.22	**96.14±0.15**
	Con-2DCNN	97.58±0.06	94.13±0.14	95.83±0.10	95.84±0.19	95.83±0.13
	Con-3DCNN	97.59±0.08	94.13±0.25	95.82±0.37	96.54±0.31	95.11±0.23
	Sub-RSSAN	97.44±0.04	93.78±0.08	95.58±0.05	95.64±0.08	95.53±0.06
	Con-RSSAN	97.50±0.05	93.92±0.12	95.67±0.09	96.28±0.12	95.08±0.13
	ResNet	97.25±0.07	93.31±0.18	95.24±0.13	95.60±0.21	94.89±0.27
	Proposed SPCNet	97.66±0.05	94.30±0.14	95.94±0.09	**96.55±0.22**	95.34±0.31
River	CVA	82.96	39.79	47.45	32.42	88.52
	SVM	96.31	73.81	75.77	88.15	66.44
	SiamNet	96.57±0.25	76.86±0.19	78.72±0.25	85.06±0.51	73.26±0.41
	SSA-SiamNet	96.70±0.31	78.88±0.25	80.68±0.43	82.17±0.57	79.24±0.62
	GETNET	96.50±0.26	76.68±0.61	78.57±0.53	84.05±0.49	73.76±0.95
	Sub-2DCNN	96.65±0.04	78.50±0.30	80.33±0.28	82.01±0.11	78.72±0.50
	Con-2DCNN	96.61±0.04	78.48±0.17	80.34±0.15	81.08±0.41	79.61±0.14
	Con-3DCNN	96.57±0.09	77.26±0.59	79.12±0.62	83.93±0.72	74.84±0.89
	Sub-RSSAN	96.88±0.04	79.88±0.24	81.58±0.22	83.61±0.56	79.66±0.52
	Con-RSSAN	96.68±0.04	78.37±0.20	80.18±0.18	83.30±0.44	77.29±0.33
	ResNet	96.73±0.04	79.14±0.44	80.93±0.42	82.01±0.59	**79.88±1.25**
	Proposed SPCNet	**97.01±0.09**	**80.29±0.93**	**81.91±0.89**	**86.43±1.62**	77.93±0.25
Hermiston	CVA	92.94	77.48	81.70	98.22	69.94
	SVM	95.65	87.04	89.79	95.35	84.84
	SiamNet	95.90±0.24	88.25±0.56	90.90±0.32	90.59±0.66	91.21±0.35
	SSA-SiamNet	96.22±0.34	89.02±0.43	91.45±0.43	93.18±0.71	89.77±0.52
	GETNET	95.58±0.27	87.50±0.41	90.37±028	88.75±0.96	92.05±0.89
	Sub-2DCNN	96.17±0.04	88.89±0.12	91.35±0.09	92.81±0.30	89.94±0.21
	Con-2DCNN	96.48±0.45	89.93±0.13	92.20±0.11	92.22±0.21	**92.17±0.30**
	Con-3DCNN	96.10±0.41	88.68±0.58	91.19±0.32	92.87±0.51	89.56±0.46
	Sub-RSSAN	95.67±0.04	87.48±0.10	90.26±0.08	91.53±0.20	89.03±0.18
	Con-RSSAN	95.85±0.06	88.00±0.18	90.67±0.15	91.97±0.13	89.40±0.31
	ResNet	96.14±0.05	88.86±0.26	91.34±0.19	92.23±0.48	90.47±0.46
	Proposed SPCNet	**96.56±0.04**	**90.05±0.13**	**92.26±0.10**	**93.51±0.44**	91.06±0.05
Bay Area	CVA	82.15	64.81	80.81	94.99	70.32
	SVM	94.78	89.53	95.07	96.02	94.14
	SiamNet	97.53±0.74	95.03±0.81	97.71±0.65	96.55±0.63	**98.90±0.95**
	SSA-SiamNet	97.62±0.69	95.21±0.96	97.76±0.57	98.08±0.71	97.45±0.83

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

	GETNET	93.85±0.56	87.65±0.75	94.26±0.43	94.18±0.53	94.33±0.76
	Sub-2DCNN	96.95±0.18	93.87±0.22	97.12±0.19	97.97±0.21	96.28±0.27
	Con-2DCNN	97.58±0.13	95.13±0.25	97.72±0.12	98.13±0.17	97.32±0.12
	Con-3DCNN	97.08±0.28	94.15±0.55	97.24±0.35	98.32±0.61	96.19±0.43
	Sub-RSSAN	96.91±0.27	93.80±0.42	97.08±0.42	98.11±0.51	96.07±0.59
	Con-RSSAN	97.31±0.33	94.60±0.64	97.45±0.38	98.62±0.47	96.32±0.75
	ResNet	97.98±0.48	95.96±0.97	98.11±0.45	98.60±0.45	97.62±0.58
	Proposed SPCNet	**98.28±0.12**	**96.55±0.24**	**98.39±0.11**	**98.92±0.24**	97.85±0.10

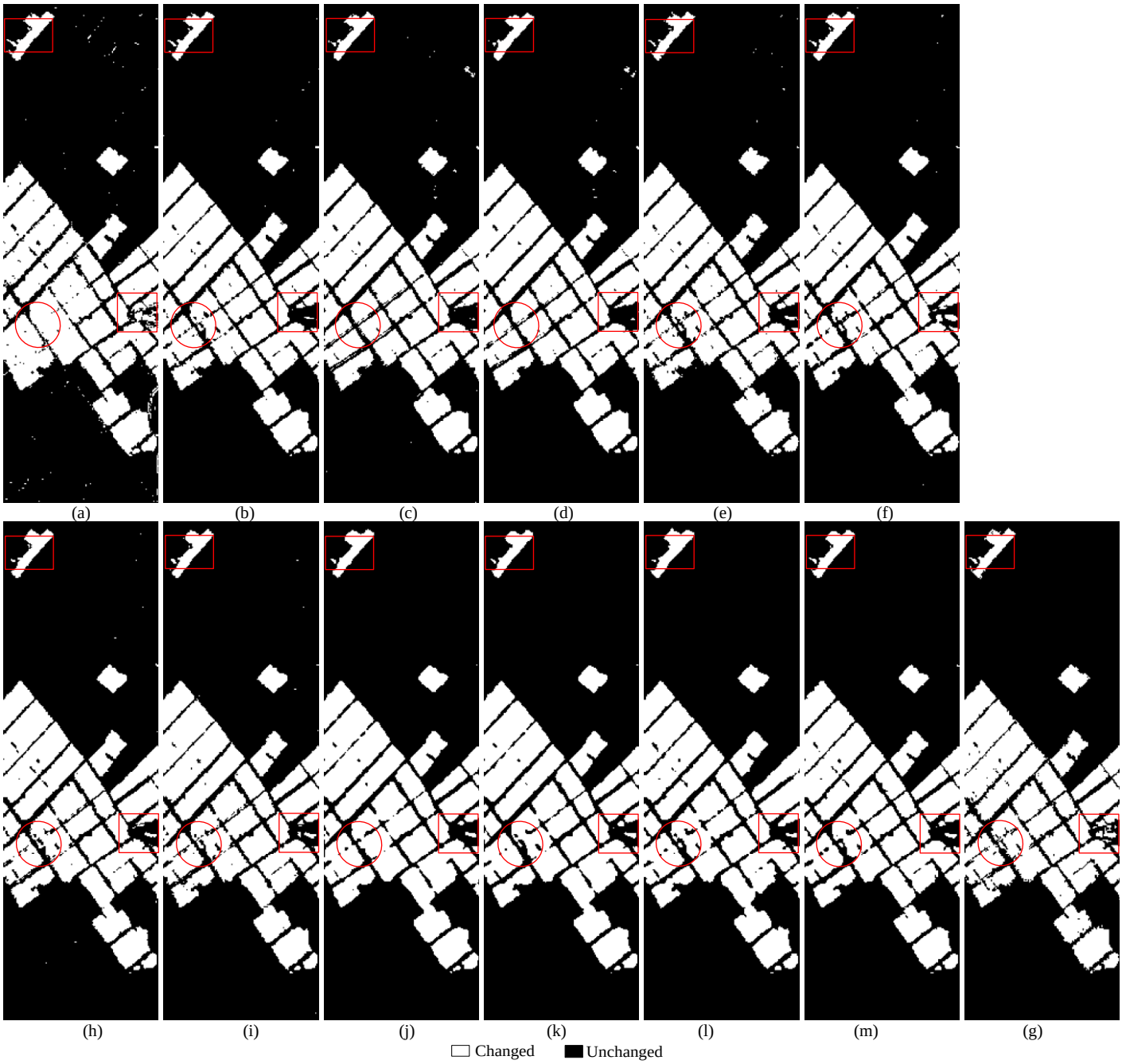


Fig. 6. CD maps produced by different methods on the Farmland dataset: (a) CVA, (b) SVM, (c) SiamNet, (d) SSA-SiamNet, (e) GETNET, (f) Sub-2DCNN, (g) Con-2DCNN, (h) Con-3DCNN, (i) Sub-RSSAN, (j) Con-RSSAN, (k) ResNet, (l) Proposed SPCNet, and (m) Ground-reference map.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

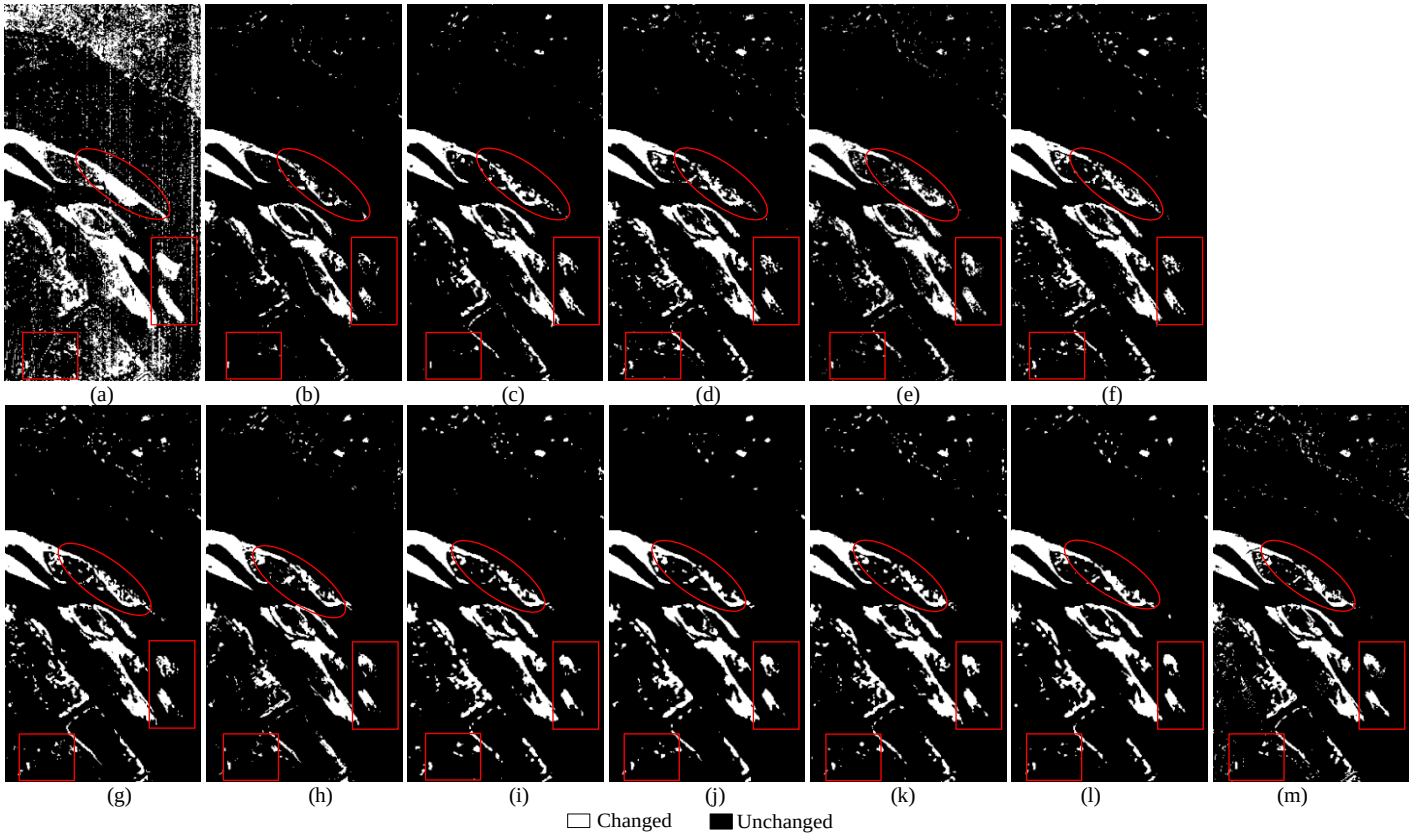
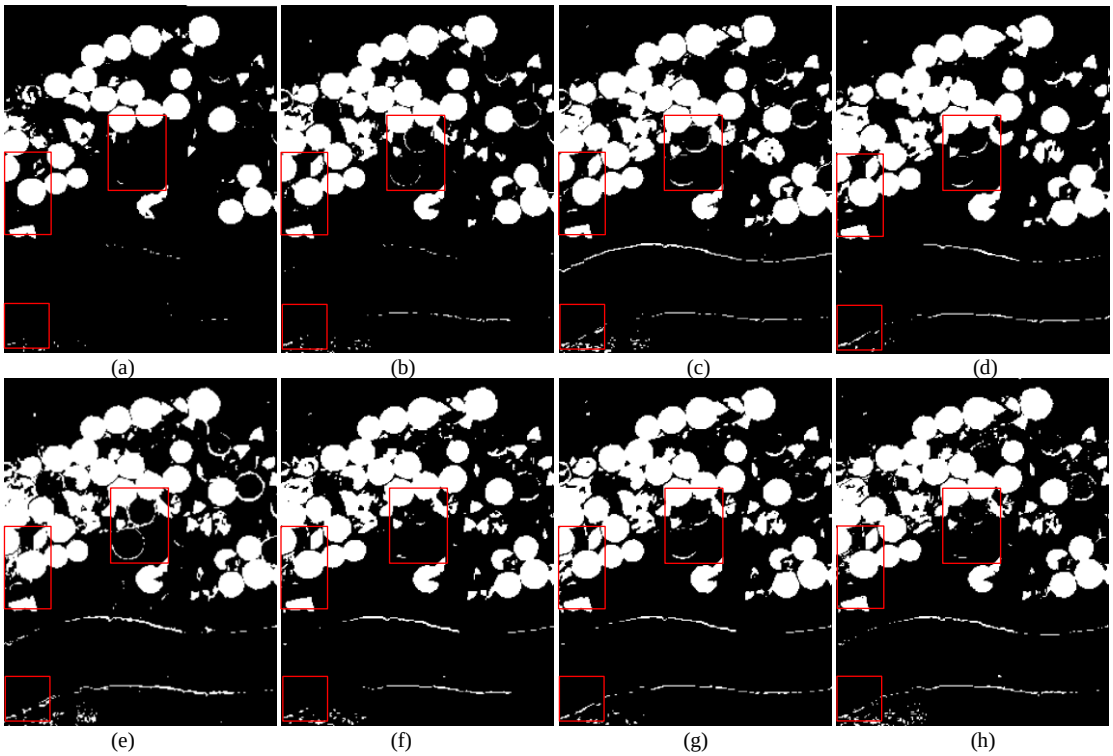


Fig. 7. CD maps produced by different methods on the River dataset: (a) CVA, (b) SVM, (c) SiamNet, (d) SSA-SiamNet, (e) GETNET, (f) Sub-2DCNN, (g) Con-2DCNN, (h) Con-3DCNN, (i) Sub-RSSAN, (j) Con-RSSAN, (k) ResNet, (l) Proposed SPCNet, and (m) Ground-reference map.



This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

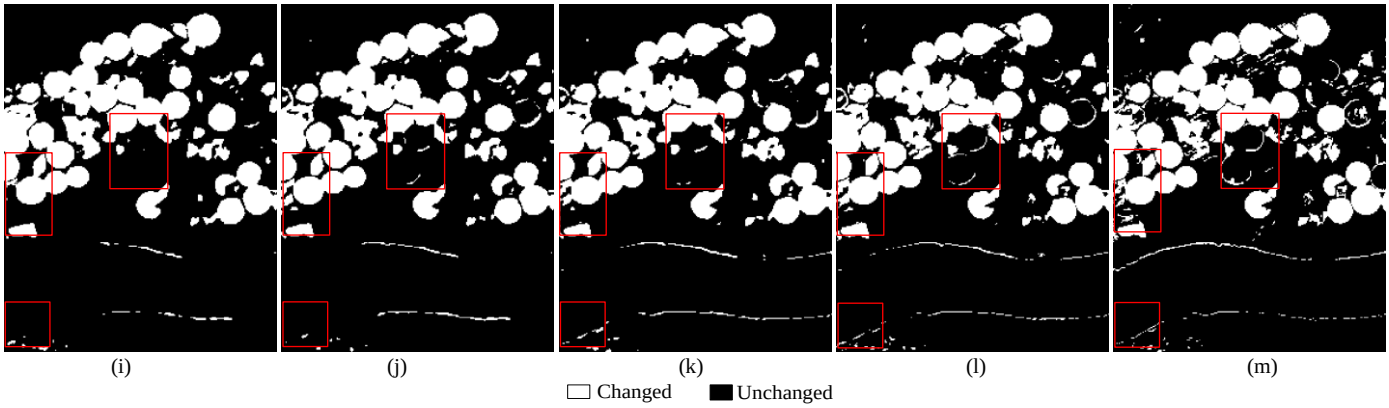


Fig. 8. CD maps produced by different methods on the Hermiston dataset: (a) CVA, (b) SVM, (c) SiamNet, (d) SSA-SiamNet, (e) GETNET, (f) Sub-2DCNN, (g) Con-2DCNN, (h) Con-3DCNN, (i) Sub-RSSAN, (j) Con-RSSAN, (k) ResNet, (l) Proposed SPCNet, and (m) Ground-reference map.

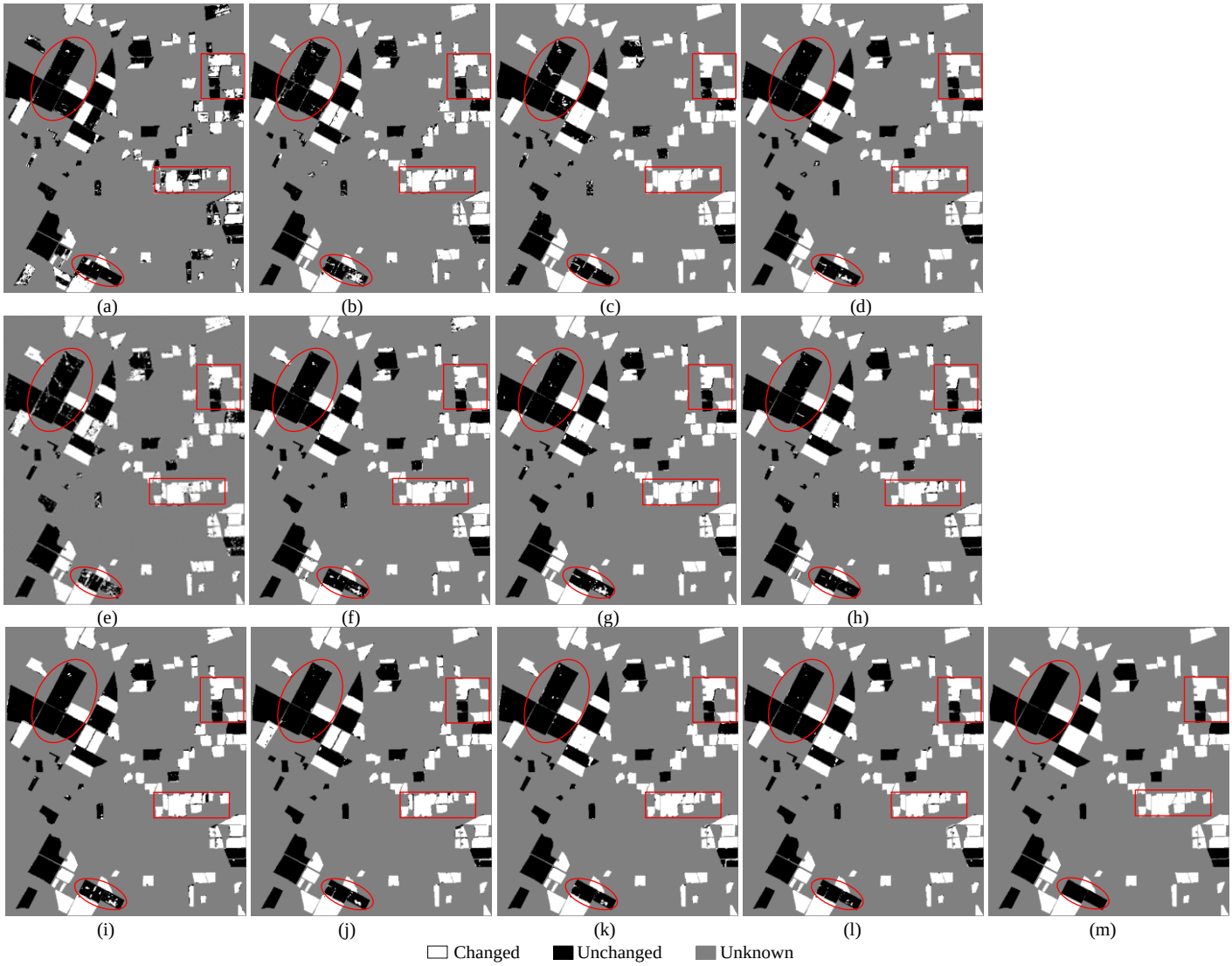


Fig. 9. CD maps produced by different methods on the Bay Area dataset: (a) CVA, (b) SVM, (c) SiamNet, (d) SSA-SiamNet, (e) GETNET, (f) Sub-2DCNN, (g) Con-2DCNN, (h) Con-3DCNN, (i) Sub-RSSAN, (j) Con-RSSAN, (k) ResNet, (l) Proposed SPCNet, and (m) Ground-reference map.

Table 2 Computational costs of different methods on the four considered datasets
(Bold values indicate the lowest computational cost results for each variable and each dataset)

Dataset		Method									
		SiamNet	SSA-SiamNet	GETNET	Sub-2DCNN	Con-2DCNN	Con-3DCNN	Sub-RSSAN	Con-RSSAN	ResNet	SPCNet
Farmland	Parameters(K)	87.84	88.60	93420	55.43	73.13	155.07	49.25	103.05	**34.47**	80.42
	FLOPs (M)	**39.36**	39.60	55080	52.81	104.24	1130	41.25	83.63	88.98	90.88

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

River	Parameters (K)	106.42	107.18	154180	48.94	91.71	159.19	63.65	135.14	**37.22**	94.18
	FLOPs (M)	134.15	134.80	90910	100.13	132.77	1470	**52.87**	107.72	186.75	213.47
Hermiston	Parameters (K)	110.15	111.52	92180	55.14	99.49	72.07	37.17	79.53	**34.40**	80.10
	FLOPs (M)	104.95	105.60	54350	52.48	103.57	1120	**40.99**	83.09	177.74	181.03
Bay Area	Parameters (K)	166.08	167.32	198270	75.30	139.81	277.03	72.68	156.25	**38.88**	102.50
	FLOPs (M)	151.40	152.05	116910	75.71	150.02	1670	**59.98**	122.72	96.04	116.32

E. Ablation Analysis

To verify the effectiveness of the proposed SPCNet method, we implemented an ablation study in which the hyperparameters were set according to Section III-C.

1) Analysis with and without downsampling strategies

To investigate the importance of using the downsampling strategy in the proposed SPCNet method, we carried out experiments without using it. To ensure the feature size matching of the method without the downsampling strategy, a convolution layer with kernel size 3×3 is used at the T2 branch. As shown in Fig. 10, the CD results of the SPCNet with and without the downsampling strategy are relatively similar for the Farmland and River datasets. In contrast, the accuracy of the proposed SPCNet with the downsampling strategy is superior to that of without it for the Hermiston and Bay Area datasets. From Table 3, one can see that the computational costs in terms of the training parameters and FLOPs with the downsampling operation are smaller than those without, which confirms the superiority of the proposed method with the downsampling operation.

2) Analysis of different downsampling strategies

Different downsampling strategies, including bilinear, nearest neighbor, and bicubic are considered to analyze their influence on the CD accuracy in the proposed SPCNet method. As shown in Table 4, the CD results with the three downsampling strategies are close in the four datasets. This finding illustrates the robustness of the proposed SPCNet to the downsampling strategy. Note that the bicubic strategy is employed in the proposed SPCNet method.

3) Analysis of the pooling strategy

We explored the influence of different pooling strategies on the CD results of the proposed SPC module. Four potential pooling strategies were considered in this paper, which were the individual average pooling operator (Avg), individual max pooling operator (Max), sum of average pooling and max pooling operators (Avg+Max), and concatenated average pooling and max pooling operators ([Avg, Max]). The strides of MaxPool and AvgPool operators are set to 1. When different pooling strategies are used in the proposed SPCNet, all of the other configurations are kept fixed. As shown in Table 5, the performances obtained with different pooling strategies are relatively close, which indicates that the SPC module is robust to the pooling strategy. Overall, Avg+Max is the most suitable for adoption in the proposed SPCNet.

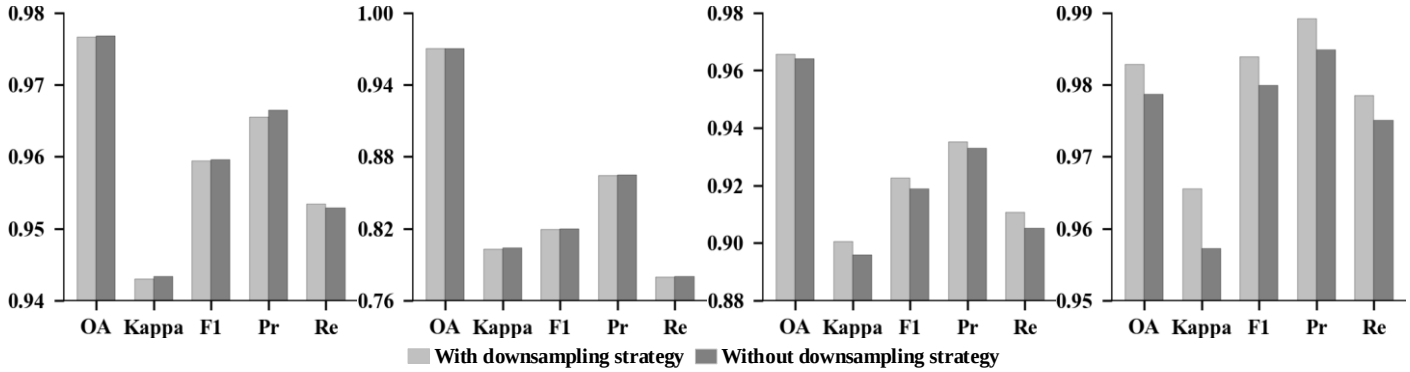


Fig. 10. Accuracy of the proposed SPCNet with and without downsampling strategies for the (a) Farmland dataset, (b) River dataset, (c) Hermiston dataset, and (d) Bay Area dataset.

Table 3 Computational costs of the proposed SPCNet with and without downsampling strategies

Downsampling strategies	Farmland		River		Hermiston		Bay Area	
	Parameters (K)	FLOPs (M)	Parameters (K)	FLOPs (M)	Parameters (K)	FLOPs (M)	Parameters (K)	FLOPs (M)
With	80.42	90.88	94.18	213.47	80.10	181.03	102.50	116.32
Without	121.12	137.77	145.89	332.61	120.55	274.22	160.87	183.56

Table 4 Results obtained by the proposed SPCNet with different downsampling strategies
(Bold values indicate the most accurate results for each variable and each dataset)

Downsampling strategies	Farmland			River			Hermiston			Bay Area		
	OA (%)	Kappa ×100	F1 (%)	OA (%)	Kappa ×100	F1 (%)	OA (%)	Kappa ×100	F1 (%)	OA (%)	Kappa ×100	F1 (%)
Bilinear	97.63	94.24	95.91	96.97	**80.39**	**82.04**	96.41	89.67	91.98	98.20	96.39	98.31
Nearest Neighbor	97.50	93.93	95.69	**97.03**	80.38	82.00	96.43	89.66	91.95	98.24	96.47	98.35
Bicubic	**97.66**	**94.30**	**95.94**	97.01	80.29	81.91	**96.56**	**90.05**	**92.26**	**98.28**	**96.55**	**98.39**

Table 5 Results obtained by the proposed SPCNet with different pooling strategies
(Bold values indicate the most accurate results for each variable and each dataset)

Pooling strategies	Farmland			River			Hermiston			Bay Area		
	OA (%)	Kappa ×100	F1 (%)	OA (%)	Kappa ×100	F1 (%)	OA (%)	Kappa ×100	F1 (%)	OA (%)	Kappa ×100	F1 (%)
Avg	97.56	94.05	95.77	96.89	79.96	81.66	96.45	89.75	92.03	98.30	96.58	98.40
Max	97.65	94.29	95.95	96.90	79.73	81.42	96.33	89.38	91.74	**98.38**	**96.74**	**98.48**
Avg+Max	**97.66**	94.30	95.94	97.01	80.29	81.91	**96.56**	**90.05**	**92.26**	98.28	96.55	98.39
[Avg, Max]	97.65	**94.31**	**95.96**	**97.03**	**80.85**	**82.48**	96.32	89.39	91.76	98.31	96.61	98.42

IV. Discussion And Conclusions

In this paper, we have proposed a general end-to-end SPCNet method to accomplish the CD task in bitemporal HSIs with different spatial resolutions. In the proposed SPCNet method, a modified ResNet is used as a baseline to detect land-cover changes. The subpixel convolution layer can fully extract the subpixel context information by learning an array of upscaling filters. Then, the designed SR block that contains an efficient subpixel convolution layer is introduced to upscale the LR feature maps into HR feature maps. Furthermore, the designed SPC module, which integrates the SR block and other related operators, is embedded into the LR branch to generate more discriminative representations. Compared with other benchmark methods, the improvement in terms of the CD accuracy obtained by the proposed SPCNet is mainly due to the application of the designed SPC module that can generate more discriminative representations of the subpixel context information. The proposed method represents a general end-to-end CD framework for bitemporal HSIs with different spatial resolutions.

As shown in Fig. 3(b), the HSIs acquired by the EO-1 satellite. (i.e., the Farmland, River, and Hermiston datasets) show a similar trend between the CD accuracy and the patch size. However, this trend is different for the Bay Area dataset acquired by AVIRIS. The apparent difference between the EO-1 satellite and AVIRIS is the spatial resolution. Therefore, it is worthwhile to further investigate the relation between increases in the CD results and the spatial resolution. The proposed SPCNet method has the potential to accomplish the CD task with other data sources, such as MSI, VHR images, and SAR images. In these cases, the hyperparameters, such as the kernel number, the batch size, and the patch size should be properly determined. For example, unlike HSI, MSI and SAR images have few spectral bands, and thus, it is reasonable to reduce the kernel numbers. For VHR images, the patch size can be increased to acquire more correlation with the surrounding pixels.

The designed SPC module can be extended as a plug-and-play unit to other baseline networks and other HSI processing tasks, such as heterogeneous image fusion [43], hyperspectral unmixing [44], and subpixel mapping [45]. For different tasks, a key parameter is the scale factor for different resolution remote sensing images. In contrast, it would be interesting in future research to explore the relation between the number of SR blocks in the SPC module and the scale factor.

It is worthwhile to note that the experiments carried out on the proposed SPCNet method used synthetic LR images. This approach is a limitation of the current experimental setup. As a future development, we plan to test the proposed method on real bitemporal images that have different spatial resolutions. Moreover, we will explore the generalization of the method to multiple CD. In addition, the proposed SPCNet is a supervised method.

References

[1] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: current techniques, applications, and challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 140–158, 2019.

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

[3] L. Ding, H. Guo, S. Liu, L. Mou, J. Zhang, and L. Bruzzone, "Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2022.3154390.

[4] W. H. Maes and K. Steppe, "Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture," *Trends in Plant Science*, vol. 24, no. 2, pp. 152–164, Feb 2019.

[5] H. Yin, D. Pflugmacher, A. Li, Z. Li, and P. Hostert, "Land use and land cover change in Inner Mongolia-understanding the effects of China's re-vegetation programs," *Remote Sensing of Environment*, vol. 204, pp. 918–930, 2018.

[6] A. Schneider, "Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach," *Remote Sensing of Environment*, vol. 124, pp. 689–704, 2012.

[7] X. Li, P. Gong, and L. Liang, "A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data," *Remote Sensing of Environment*, vol. 166, pp. 78–90, 2015.

[8] X.-P. Song, J. O. Sexton, C. Huang, S. Channan, and J. R. Townshend, "Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover," *Remote Sensing of Environment*, vol. 175, pp. 1–13, 2016.

[9] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: application to tsunami-damage assessment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1658–1670, 2007.

[10] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, 2010.

[11] L. Dong and J. Shan, "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 84, pp. 85–99, 2013.

[12] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *International Journal of Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 2010.

[13] R. J. Radke, S. Andra, O. Al-Kofahi and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.

[14] F. Bovolo, S. Marchesi and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2196–2212, 2012.

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

[16] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2858–2874, May 2014.

[17] J. S. Deng, K. Wang, Y. H. Deng, and G. Qi, "PCA-based landuse change detection and analysis using multitemporal and multisensor satellite data," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4823–4838, 2008.

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

[19] W. Sun et al., "A label similarity probability filter for hyperspectral image post classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6897–6905, 2021.

[20] L. Bruzzone, S.B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 858-867, 1997.

[21] B. Demir, F. Bovolo, L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active learning based compound classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1930-1941, 2012.

[22] F. Bovolo, L. Bruzzone, "The time variable in data fusion: a change detection perspective," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 8-26, 2015.

[23] N. Wambugu et al., "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 105, 2021.

[24] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-TripleGAN: GAN-assisted CapsNet for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7232-7245, 2019.

[25] X. Wang, K. Tan, P. Du, C. Pan, and J. Ding, "A unified multiscale learning framework for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-19, 2022.

[26] L. Wang, L. Wang, Q. Wang and P. M. Atkinson, "SSA-SiamNet: spectral–spatial-wise attention-based siamese network for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[27] J. Qu, S. Hou, W. Dong, Y. Li, and W. Xie, "A multilevel encoder–decoder attention network for change detection in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[28] T. Zhan et al., "TDSSC: A three-directions spectral–spatial convolution neural network for hyperspectral image change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 377–388, 2021.

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

[30] 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 3–13, 2019.

[31] K. Zhang, M. Wang, and S. Yang, "Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1363–1371, 2017.

[32] M. Liu, Q. Shi, A. Marinoni, D. He, X. Liu and L. Zhang, "Super-resolution-based change detection network with stacked attention module for images with different resolutions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[33] K. Wu, Q. Du, Y. Wang, and Y. Yang, "Supervised sub-pixel mapping for change detection from remotely sensed images with different resolutions," *Remote Sensing*, vol. 9, no. 3, 2017.

[34] X. Li, F. Ling, G. M. Foody, and Y. Du, "A super resolution land-cover change detection method using remotely sensed images with different spatial resolutions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3822–3841, 2016.

[35] W. Shi, J. Caballero, F. Huszár, et al., "Real-time single image and video super-resolution using an efficient," in *Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[37] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of Operations Research*, vol. 134, no. 1, pp. 19–67, 2005.

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

[39] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2070–2082, 2008.

[40] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[41] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.

[42] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral-spatial attention network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2020.

[43] K. Zhang, M. Wang, S. Yang and L. Jiao, "Spatial–spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1030–1040, 2018.

[44] L. Wang, D. Liu, and Q. Wang, "Geometric method of fully constrained least squares linear spectral mixture analysis", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 6, pp. 3558–3566, 2013.

[45] P. Wang, L. Wang, H. Leung, and G. Zhang, "Super-resolution mapping based on spatial–spectral correlation for spectral imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2256–2268, 2021.

Lifeng Wang received the B.S. and M.S. degree from the Northeast Agricultural University, Harbin, China, in 2014 and 2017, respectively.

She is currently pursuing the Ph.D. degree in the College of Information and Communication Engineering Harbin Engineering University, Harbin, China.

Her research interests include hyperspectral image change detection and deep learning.

Liguo Wang received his M.S. degree in natural science and the Ph.D. degree in engineering from the Harbin Institute of Technology (HIT) in 2002 and 2006, respectively, and held postdoctoral research position from 2006 to 2008 in the College of Information and Communications Engineering, Harbin Engineering University, where he is currently a professor.

His research interests are remote sensing image processing and machine learning. He has published four books, 35 patents, and more than 230 papers in journals and conference proceedings.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3189188

Heng Wang received the B.S. degree in Computer and Communication Engineering from the Zhengzhou University of Light Industry, Zhengzhou, China, in 2015, and the M.S. degree in information and communication engineering from Harbin Engineering University, Harbin, China, in 2019. He is pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin China.

His research interests include computer vision, deep learning, remote sensing and deep generative models.

Xiaoyi Wang received the M.A. degree from Heilongjiang University, Harbin, China, in 2018. She is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin.

Her research interests are remote-sensing image fusion and hyperspectral target detection.

Lorenzo Bruzzone received the Laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, and digital communications.

Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory (https://rslab.disi.unitn.it/) in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and SAR, signal processing, machine learning and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the Principal Investigator of many research projects. Among the others, he is currently the Principal Investigator of the Radar for icy Moon exploration (RIME) instrument in the framework of the JUpiter ICy moons Explorer (JUICE) mission of the European Space Agency (ESA) and the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA.

He is the author (or coauthor) of 294 scientific publications in referred international journals (221 in IEEE journals), more than 340 papers in conference proceedings, and 22 book chapters. He is editor/co-editor of 18 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven from the total number of citations (more than 40000) and the value of the h-index (92) (source: Google Scholar). He was invited as keynote speaker in more than 40 international conferences and workshops. Since 2009 he has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS), where since 2019 he is Vice-President for Professional Activities. Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since that he was recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards and the 2019 WHISPER Outstanding Paper Award. Dr. Bruzzone was a Guest Co-Editor of many Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003 he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the founder of the IEEE Geoscience and Remote Sensing Magazine for which he has been Editor-in-Chief between 2013-2017.Currently he is an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing. He has been Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society between 2012-2016. He is a Fellow of IEEE.