FDLdet: A Change Detector Based on Forward Dictionary Learning for Remote Sensing Images

Yuqun Yang, *Student Member, IEEE*, Xu Tang, *Senior Member, IEEE*, Xiangrong Zhang, *Senior Member, IEEE*, Jingjing Ma, *Member, IEEE*, Yiu-Ming Cheung, *Fellow, IEEE*, and Licheng Jiao, *Fellow, IEEE*

***Abstract*— As an important topic in the remote sensing (RS) image processing community, change detection has attracted much attention from researchers, which aims to distinguish land-cover changes in a geographic position. This is a challenging task because the visual representations of land cover captured from RS images at different periods would vary widely and considerably, resulting in significant differences in feature representations. To alleviate this problem, many existing deep-based methods employ the parameter-shared strategy to map RS images into a common feature space for detecting the changes. Although they are feasible, the simple and single visual information learned by deep models is still not sophisticated enough for satisfactory results. To address this problem, we propose a forward dictionary learning (DL) model named forward DL detector (FDLdet) in this article. Besides the common visual features, our FDLdet takes into account the essential information, e.g., element composition and land-cover category, for change detection. FDLdet consists of a feature extractor, a coefficient generator, and a deep dictionary. Specifically, first, the feature extractor is used to extract shared deep features from RS images. Second, the coefficient generator transforms these deep features into word coefficients. Third, words within the deep dictionary are combined by word coefficients to generate the dictionary features with essential information. Finally, the dictionary features are used instead of deep features to detect land-cover changes. Extensive experiments are conducted on two public large-scale datasets, i.e., season-varying change detection (SVCD), Sun Yat-sen University change detection (SYSU-CD), and LEVIR change detection (LEVIR-CD). Experimental results demonstrate the effectiveness of the proposed FDLdet. Our source codes are available at https://github.com/TangXu-Group/FDLdet.**

***Index Terms*— Change detection, deep learning, dictionary learning (DL), remote sensing (RS).**

Manuscript received 8 October 2023; revised 3 April 2024 and 15 June 2024; accepted 24 June 2024. Date of publication 1 July 2024; date of current version 10 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62171332 and Grant 62276197; in part by the Key Research and Development Program of Shaanxi under Grant 2024GX-YBXM-125; in part by the Natural Science Basic Research Program of Shaanxi under Grant 2024JC-YBMS-472; in part by the Shaanxi Province Innovation Capability Support Plan under Grant 2023-CX-TD-09; in part by the NSFC/Research Grants Council (RGC) Joint Research Scheme under Grant N_HKBU214/21; in part by the General Research Fund of RGC under Grant 12201321, Grant 12202622, and Grant 12201323; and in part by the RGC Senior Research Fellow Scheme under Grant SRFS2324-2S02. *(Corresponding author: Xu Tang.)*

Yuqun Yang, Xu Tang, Xiangrong Zhang, Jingjing Ma, and Licheng Jiao are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710126, China (e-mail: tangxu128@gmail.com).

Yiu-Ming Cheung is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, SAR, China.

Digital Object Identifier 10.1109/TGRS.2024.3421664

I. Introduction

CHANGE detection of remote sensing (RS) image aims to identify changes in land covers between multitemporal RS images covering the same geographic areas [1], [2]. It has been applied to many practical applications, such as disaster monitoring [3], resource survey [4], and urban planning [5]. In recent years, with the development of sensor technologies, the resolution of RS images has improved significantly. The land-cover information within high-resolution RS (HRRS) images is delicate and sufficient, which simultaneously brings opportunities and challenges for RS change detection. On the one hand, the precise and elegant land-cover contents can provide valuable clues for researchers to make their decisions. On the other hand, owing to this detailed information, the problem of the same objects with various appearances has become particularly prominent, which obviously raises the difficulty of change detection. For example, the same forest within HRRS images produced at different times may render diverse states. Its distribution is sparse in spring, autumn, and winter but is dense in summer. Also, its color will change from green to yellow and gray when the season varies from spring to winter. Therefore, exploring the invariant characteristics of land covers to eliminate the negative influences caused by the "visual appearance gap" (see Fig. 1), i.e., the different appearance of the same land covers in multitemporal HRRS images, is one of the crucial problems in HRRS image change detection.

In an early stage, the conventional machine-learning-based technologies (such as principal component analysis and mixture parameter estimation) with hand-crafted features were popular in HRRS image change detection as they are easy to implement and stable in behavior [6], [7], [8]. However, their performance cannot meet what we expected because the low-level features are unable to describe the complex contents within HRRS images. Then, learning-based feature extraction methods appeared and became prevalent in HRRS image change detection. One popular feature learning approach is dictionary learning (DL) [9]. Benefiting from its learning strategy, the representation capacity of the features obtained by DL is stronger. Both the shallow information (e.g., color and textual) and the essential clues of various land covers within HRRS images can be captured. Therefore, DL achieves successes in many HRRS applications [10], [11], as well as HRRS change detection [12], [13]. Nevertheless, the inadequate generalization

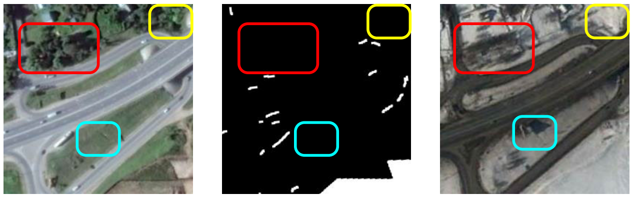


Fig. 1. Sample illustrating the concept of "visual appearance gap." Here, images *A* and *B* denote the two temporal HRRS images, while the change ground truth depicts changes between them. It is evident that even though most areas in both images *A* and *B* are not labeled as changes, they still display a significant difference in appearance. For instance, the red/yellow/blue boxes correspond to the reasons of "leaf fall"/"snow cover"/"shadow variation," respectively.

limits its performance in multitemporal HRRS image change detection.

With the development of deep learning [14], [15], many deep-based HRRS image change detection methods have been proposed [16], [17], [18]. Owing to the strong capacity of nonlinear fitness, the methods developed based on deep convolutional neural network (DCNN) have dominated the HRRS change detection community [19]. Most of them are habituated to embedding the multitemporal HRRS images into a common feature space so that the relationships between diverse land covers corresponding to various HRRS images can be measured directly. In other words, they always input multitemporal HRRS images into the same or parameter-shared DCNNs to extract features and generate the change maps [20], [21], [22]. Despite their well performance, these DCNN-based methods have not reached the satisfactory stage. Due to the stacked structure and hierarchical learning, DCNN-based methods can capture visual and semantic information from HRRS images. These deep features can support diverse HRRS tasks [23], [24], [25], [26], [27], [28]. However, when we apply them to complete HRRS change detection, the false alarm rates of change maps may be relatively high [29], because the visual and semantic clues cannot effectively narrow the "visual appearance gap." Therefore, supplementing the essential knowledge to the deep features is key to improving the accuracy of deep-based change detection models.

Most recently, the deep DL emerged [30], [31], [32]. Combining deep learning and DL, the obtained feature representation contains visual, semantic, and essential information, which benefits HRRS image change detection. However, some inherent disadvantages influence deep DL's efficiency. First, complicated reverse reconstruction and repeated coefficient training are required to generate word coefficients, even in the inference stage. Second, the sparsity constraint in word coefficient production may result in information loss. These two shortcomings limit the current deep DL technologies' performance for HRRS image change detection tasks.

To overcome the above limitations, we propose a forward DL detector (FDLdet) for HRRS image change detection. On the one hand, considering the targets of the HRRS image change detection task (i.e., judging whether each pixel is changed or not), FDLdet replaces the complicated reverse image reconstruction by narrowing the differences between the predicted change maps and ground truth. In this way, the efficiency of FDLdet can be improved significantly. On the other hand, taking the characteristics of HRRS images (e.g., diverse land covers with various types and scales) into account, instead of learning the sparse word coefficients at the pixel level, FDLdet develops the regional word coefficients to explore the land-cover essences deeply. In this fashion, the information loss issue can be mitigated. At the same time, the shape and edge knowledge, which is crucial for HRRS change detection, can also be captured. Specifically, for two temporal HRRS images, we first use a feature extractor to learn the useful visual features. Then, these visual features are input into a coefficient generator to generate regional word coefficients. Next, a deep DL algorithm is proposed to construct dictionary features (which are rich in essential information) according to regional word coefficients. Finally, the obtained dictionary features of two temporal HRRS images will be concatenated and fed into the follow-up detector to produce the change map.

The main three contributions of this article are summarized as follows.

1) A forward DL-based HRRS image change detection framework (FDLdet) is proposed, which can effectively alleviate the negative influence caused by the "visual appearance gap" and achieve accurate change detection results in an end-to-end manner.
2) We propose a regional word coefficient learning strategy and a simple yet effective dictionary feature algorithm. Along with DCNN, the diverse land covers' visual, semantic, and essential features can be fully explored for change detection.
3) The comprehensive experiments are conducted in three large-scale public change detection datasets to demonstrate the effectiveness of the proposed FDLdet.

The remainder of this article is organized as follows. Section II briefly reviews the HRRS change detection methods based on DL and DCNN. Section III discusses the details of the proposed FDLdet. The experiments and results analyses are shown in Section IV. Finally, we draw a conclusion in Section V.

II. Related Work

A. Dictionary-Learning-Based Methods

The target of HRRS image change detection is to accurately distinguish the change information of land covers by analyzing the complex contents of HRRS images. To this end, the essential knowledge of various land covers should be exploited carefully. Fortunately, DL is a suitable technique for this goal. Therefore, many notable DL-based HRRS change detection methods were proposed. As early as 2014, DL was applied to HRRS change detection successfully [33], in which a tree-structured dictionary is learned to generate robust distance for accurately distinguishing changed land covers. Then, Li et al. [34] applied coupled DL (CDL) to HRRS change detection tasks. The authors leveraged two coupled dictionaries to model the relationships between patches corresponding to difference images (DIs) and change maps so that credible detection results can be generated. Another CDL-based method was proposed in [13] for multisource HRRS image change detection. The designed coupled dictionaries map multisource HRRS images into a high-dimensional feature space and find their differences. Also, an iterative optimization algorithm is proposed to update the atoms within coupled dictionaries to guarantee that the useful information hidden in different HRRS images can be explored. Lu et al. [12] introduced a joint DL method to detect the changes from multitemporal HRRS images. The change detection task is converted into a dictionary reconstruction issue using the unchanged prior information. Furthermore, DL was integrated with graph technology to carry out the HRRS change detection task [35]. The process involves utilizing DL to construct the similarity graph matrix, which is then employed to generate sparse DIs and the change map. As a robust low-rank learning tool, DL was also utilized to construct low-rank representation (LLR) in [36] for finding the representative pixels within DIs so that favorable HRRS change maps can be obtained. Similarly, a nonlocal low-rank model was proposed [37], which groups similar patches into a nonlocal window for conducting LLR learning, then adopts the two-level clustering to predict the change map.

Although the above-mentioned methods have succeeded in their application domains, they all depend on the grid HRRS image patches. In other words, each word of a dictionary links a regular rectangular image patch. This would decrease their performance more or less. To solve this limitation, several strategies were proposed. For instance, a deformable DL algorithm was proposed in [38] for HRRS image change detection. Instead of the simple patches, the flexible patches are used to construct the dictionary, which enhances the precision and robustness of the change detection results. Moreover, Yu et al. [39] presented a joint-related DL to address HRRS image change detection, where the concept of dictionary correlation is embedded to ensure that the initial dictionary is uniform and change detection results are stable. A deep DL-based change detection model was developed in [40], in which the weighted collaborative representation is adopted to enhance the model robustness for regular image patches and invisible outliers.

B. Deep-Based Methods

In recent years, with the help of deep learning, especially the DCNNs, many deep-based methods have been proposed to accomplish HRRS image change detection [41], [42], [43].

As a pixel-level task, change detection can be regarded as a two-class semantic segmentation mission. Thus, UNet [44] and fully convolutional network (FCN) [45] are two popular structures in deep-based methods. For example, in [21], three FCN-based models fully convolutional-early fusion (FC-EF), FC-siamese-difference (FC-Siam-diff), and FC-siamese-concatenation (FC-Siam-conc) were proposed for HRRS image change detection. Here, FC-EF is a single encoder–decoder network that first concatenates the multitemporal HRRS images and then leverages FCN to generate the change maps. While FC-Siam-diff and FC-Siam-conc have two encoders, they receive bitemporal HRRS images. Along with the different skip connection schemes and a unique decoder, FC-Siam-diff and FC-Siam-conc can predict the land-cover changes. Another similar work can be found in [20]. Besides the FCN backbone, the authors introduce the pyramid pooling and skip connection to improve the HRRS image change detection performance. Based on UNet, an end-to-end HRRS change detection framework was proposed [46]. It combines the contributions of co-registered image pairs, global and fine-grained information extraction, and a specific fusion strategy to produce superb change maps.

To further improve the HRRS image change detection results, the visual attention mechanism [47] is adopted by scholars, which can alleviate the negative influence caused by noise, pseudo-changes, etc. A deeply supervised attention metric-based network (DSAMNet) was proposed in [48], in which the convolutional block attention module is used to generate the discriminative features for change detection. Cheng et al. [49] developed a deliver improved separability network to generate the change maps with low false alarms. In this network, channel and spatial attention are embedded to highlight the semantic and positional information within HRRS images so that the learned features can be refined to ensure the change maps' quality. Similarly, Li et al. [50] proposed a densely attentive refinement network for HRRS image change detection. This network contains a hybrid attention module and a recurrent refinement module, which aim to explore the contextual information and refine change maps. By combining them, favorable change detection results can be obtained.

Although the behavior of the above methods is assertive, they ignore the valuable spatial–temporal dependency information, which is also essential for HRRS image change detection. To fill this gap, Chen et al. [51] propose a bitemporal image transformer (BIT). With the help of self-attention, BIT is good at modeling contextual clues in the spatial–temporal domain effectively and efficiently. Consequently, the change maps are improved markedly. In view of the low capacity of the transformer in capturing low-level details, UNet and Transformer are combined to construct a new change detection framework in [52] named TransUNetCD. By solving the problems of feature redundancy and context loss, TransUNetCD can precisely find the changes in bitemporal HRRS images.

III. Proposed Method

The framework of FDLdet is shown in Fig. 2. It consists of a feature extractor, a coefficient generator, and a deep dictionary. The feature extractor aims to learn visual and semantic features from bitemporal HRRS images. The coefficient generator pays attention to transforming these image features into word coefficients considering regional information hidden in HRRS images. Then, the deep dictionary can generate discriminative sentences (i.e., dictionary features) for describing bitemporal HRRS images based on these word coefficients. Finally, the produced sentences are utilized to identify land-cover changes.

A. Preliminary of Dictionary Learning

Suppose there is an HRRS image $\mathbf{X}$ with N patches, the process of DL is defined as

$$\arg\min_{c_i,\mathbf{D}} \sum_{i=1}^{N} ||x_i - \mathbf{D}\cdot c_i||_2^2 + \lambda \sum_{i=1}^{N} ||c_i||_1 \tag{1}$$

where $x_i \in \mathbb{R}^l$ represents ith image patch, $\mathbf{D} \in \mathbb{R}^{l\times n}$ denotes the dictionary with n word vectors, $c_i \in \mathbb{R}^n$ indicates ith word coefficient corresponding to x_i, λ implies the optimization weight that is used to balance the accuracy of image reconstruction and the sparsity of word coefficients, and $||\cdot||_1$ and $||\cdot||_2$ are the $L1$ and $L2$ norm, respectively.

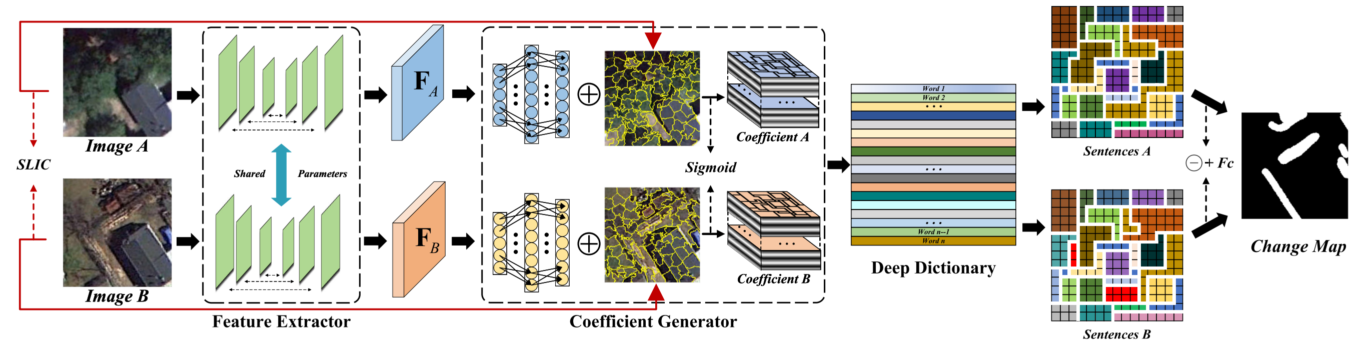


Fig. 2. Overall framework of the proposed FDLdet. First, two temporal HRRS images *Image A* and *Image B* are inputted into feature extractor to extract the image features $\mathbf{F}_A$ and $\mathbf{F}_B$. Second, the coefficient generator receives both image features and regional representations of two HRRS images to generate the regional word coefficients *Coefficient A* and *Coefficient B*. Here, the regional representations are generated by utilizing the superpixel segmentation algorithm simple linear iterative cluster (SLIC). Third, *Coefficient A* and *Coefficient B* are utilized to select words from deep dictionary for constructing the dictionary features *Sentences A* and *Sentences B*. Finally, after utilizing fully connected (FC) layer to analyze these distances adaptively, the change map can be generated by measuring the distances between these two *Sentences*.

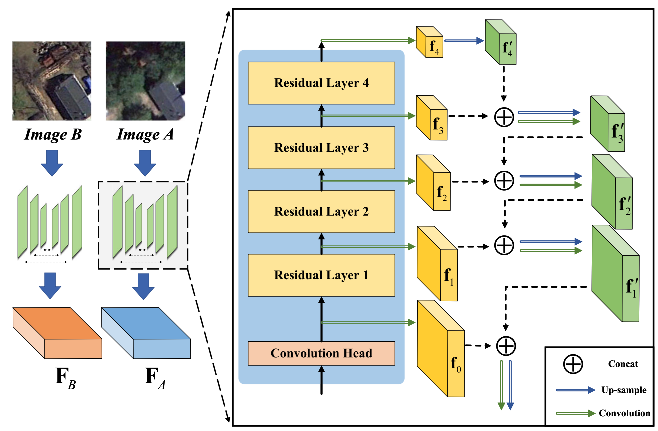


Fig. 3. Feature extractor of the proposed FDLdet. It consists of a ResNet-based backbone (marked by the background of blue color) and a feature fusion process, which receives two temporal HRRS images *Image A* and *Image B*, and outputs their image features $\mathbf{F}_A$ and $\mathbf{F}_B$.

Generally, the elements within $\mathbf{D}$ are initialized with random vectorized image patches, and c_i is defined and initialized randomly for each item in $\mathbf{D}$. By optimizing $\mathbf{D}$ and c_i alternately according to (1), the HRRS image $\mathbf{X}$ can be reconstructed under the sparse constraint. This way, the optimal dictionary $\mathbf{D}$ contains essential information about the HRRS image, and $\mathbf{A}$ composed by the optimized c_i can be considered as features learned from the HRRS image $\mathbf{X}$. If there are several serial dictionaries, the obtained $\mathbf{A}$ of the previous dictionary can be deemed as $\mathbf{X}$ of the next dictionary. After the multi-DL, more essential information of HRRS images can be studied.

B. Feature Extractor

The visual and semantic features are important for HRRS image change detection. To ensure their discrimination, we develop the feature extractor in this section, and its framework is shown in Fig. 3. To meet the demands of change detection, there are two feature extractors in FDLdet to process bitemporal HRRS images (*Image A* and *Image B*) with the size of $c \times h \times w$, where c, h, and w denote the number of channels, heights, and widths. Since two extractors have the same structure and share the same parameters, we only explain the extractor corresponding to *Image A* for clarity.

Specifically, considering the strong capacity of feature learning, ResNet [53] is adopted as the feature extraction backbone to extract multilayer features from *Image A*. ResNet mainly consists of a convolution head and four residual layers. After inputting *Image A* into ResNet, five image features $\mathbf{f}_0$, $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$, and $\mathbf{f}_4$ can be learned by the convolution head and four residual layers, respectively. Their sizes are $64 \times (h/2) \times (w/2)$, $64 \times (h/4) \times (w/4)$, $128 \times (h/8) \times (w/8)$, $256 \times (h/16) \times (w/16)$, and $512 \times (h/32) \times (w/32)$. Owing to the different learning stages, those five features incorporate various information, such as color, texture, semantics, and multiscale clues. To integrate them, a feature fusion procedure is proposed to hierarchically combine them for describing the complex contents within the HRRS image comprehensively. This fusion procedure is formulated as follows:

$$\begin{aligned}\mathbf{f}'_4 &= \mathrm{Upsample}(\mathbf{f}_4)\\ \mathbf{f}'_i &= \mathrm{Conv}\big(\mathrm{Upsample}\big(\mathbf{f}'_{i+1} \oplus \mathbf{f}_i\big)\big), \quad i = 3, 2, 1\\ \mathbf{F} &= \mathrm{Conv}\big(\mathrm{Upsample}\big(\mathbf{f}'_1 \oplus \mathbf{f}_0\big)\big)\end{aligned} \tag{2}$$

where Upsample$(\cdot)$ denotes the bilinear interpolation upsampling, Conv$(\cdot)$ implies the 1×1 convolutional operation, and $\oplus$ represents the feature concatenation in the channel dimension. Along these lines, the learned five features $\mathbf{f}_0$, $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$, and $\mathbf{f}_4$ would be fused to feature $\mathbf{F}_A \in \mathbb{R}^{N_c \times w \times h}$, which involves rich visual and semantic knowledge. Here, N_c means the number of channels. Similarly, when we input *Image B* into the feature extractor, its feature representation $\mathbf{F}_B \in \mathbb{R}^{N_c \times w \times h}$ can be obtained.

It is worth noting that $\mathbf{f}_0$ is learned by only one convolution layer (convolution head), so it hardly contains much-advanced information. However, it still is adopted in the feature fusion procedure since the low-level information of land covers within $\mathbf{f}_0$ is also important for change detection tasks.

C. Coefficient Generator

Before introducing the proposed coefficient generator, let us review how human beings understand an HRRS image. Given an HRRS image, scholars will analyze its contents first, and then they will select some words to describe the key contents. Here, the key contents indicate the salient objects or regions within the HRRS image, and they share similar low-level

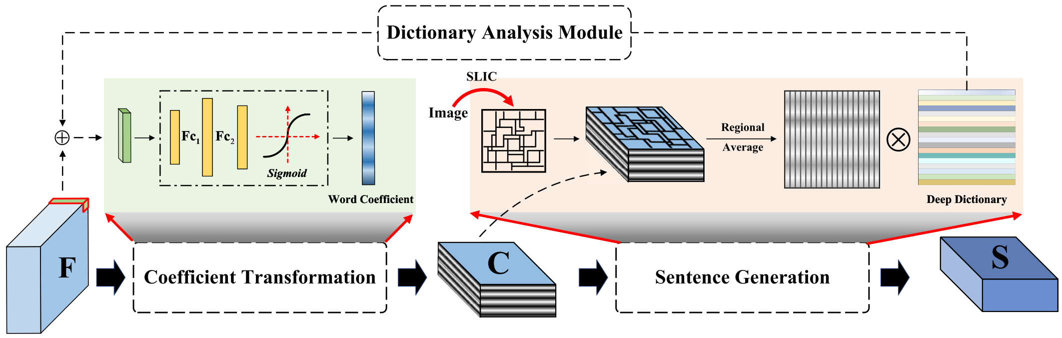


Fig. 4. Flowchart of the forward DL. After obtaining the image feature **F**, each of the feature pixels is inputted into coefficient transformation module to generate word coefficient **C**. Then, the segmentation results generated by segmenting HRRS image with SLIC and word coefficients **C** are inputted into sentence generation module to generate sentences **S**. The dictionary analysis module is employed to help the generation of word coefficients by preliminarily understanding deep dictionary.

visual features or high-level semantical clues. For example, "Lawn" can describe the object with green color and specific texture (low-level visual features), and "Airport" can be used to depict a region with "Airplane," "Runway," and "Building" (high-level clues). Next, an HRRS image can be represented by a sentence with these preferred words. Compared with the feature vectors, the above expression's manner matches human cognition habits. Also, the obtained sentence can portray the HRRS image's key contents directly.

If we follow the above-mentioned human habits to address our change detection tasks, the first problem to be solved is to select the proper words for describing HRRS images. In other words, assume that there is a dictionary $\mathbf{D}$ with sufficient words. What we have to do is to find the suitable words from $\mathbf{D}$ according to advisable word coefficients. To this end, we develop our coefficient generator based on the visual and semantic features ($\mathbf{F}_A$ and $\mathbf{F}_B$). The obtained word coefficients are rich in pixel- and region-level information. For clarity, we use $\mathbf{F}$ to unify $\mathbf{F}_A$ and $\mathbf{F}_B$ to explain our coefficient generator as their procedures are the same to each other. The flowchart of the coefficient generator is shown in Fig. 4. When we input the visual and semantic feature $\mathbf{F}$, a coefficient transformation module will be applied to generate the word coefficients $\mathbf{C}$ at the pixel level. Then, a regional information mining strategy is developed to reinforce $\mathbf{C}$ at the region level. Finally, the enriched $\mathbf{C}$ can be used to select valuable words from the dictionary $\mathbf{D}$. Note that, to optimize the coefficients and dictionary at the same time, the dictionary $\mathbf{D}$ will be analyzed deeply, and the analyzed results will be added to $\mathbf{F}$. The details of dictionary analysis are discussed in Section III-D.

In detail, for $\mathbf{F} \in \mathbb{R}^{N_c \times w \times h}$, we first use two FC layers to transform it into word coefficients $\mathbf{C} \in \mathbb{R}^{N_w \times w \times h}$, where N_w indicates the volume of words within $\mathbf{D}$. Then, a sigmoid function is applied to normalize the values of the elements within $\mathbf{C}$ into an interval of $(0, 1)$. The above process can be formulated as

$$\mathbf{C} = \text{sigmoid}(\mathcal{F}_{\text{fc}}(\mathbf{F})) \tag{3}$$

where sigmoid$(\cdot)$ and $\mathcal{F}_{\text{fc}}(\cdot)$ indicate the sigmoid function and nonlinear mapping function of FC layers. Although the current word coefficients can help us pick useful words from the dictionary, they only involve pixel-level information. This leads to a potential problem, i.e., some feature pixels corresponding to similar image characteristics may not be transformed to similar word coefficients. It is adverse to distinguishing the land-cover changes accurately, especially the edges of change regions. To overcome this limitation and further improve $\mathbf{C}$, the "word region" (WR) is developed here. Particularly, assume that the HRRS image has been over-segmented into N_r superpixels by the SLIC algorithm [54]. Therefore, $\mathbf{C}$ can be divided into WRs $\{\mathbf{c}_1, \ldots, \mathbf{c}_{N_r}\}, \mathbf{c}_i \in \mathbb{R}^{N_w \times N_{f_i}}$ according to these superpixels first, where N_{f_i} means the number of pixels within ith superpixel. Then, the regional coefficients $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{N_r}\}, \mathbf{v}_i \in \mathbb{R}^{N_w}$ will be generated for WRs by averaging the word coefficients within each WR. Finally, the regional coefficients will be embedded to $\mathbf{C}$ by

$$\mathbf{C} = \mathbf{C} + \mathbf{V} = \left\{\mathbf{c}_1 + \mathbf{v}_1, \ldots, \mathbf{c}_{N_r} + \mathbf{v}_{N_r}\right\}$$
$$\mathbf{c}_i + \mathbf{v}_i = \left\{\mathbf{c}_{i_{[:,1]}} \oplus \mathbf{v}_i, \ldots, \mathbf{c}_{i_{[:,N_{f_i}]}} \oplus \mathbf{v}_i\right\} \tag{4}$$

where $\mathbf{c}_{i_{[:,j]}} \in \mathbb{R}^{N_w}$ and $\oplus$ indicates element-wise addition operation. Consequently, the current word coefficients enclose the pixel- and region-level information.

Based on the above method, we can get $\mathbf{C}_A$ and $\mathbf{C}_B$ for $\mathbf{F}_A$ and $\mathbf{F}_B$. By multiplying the dictionary with them (i.e., $\mathbf{D} \cdot \mathbf{C}_A$ and $\mathbf{D} \cdot \mathbf{C}_B$), the appropriate sentences can be produced to represent *Image A* and *Image B*.

D. Deep Dictionary

In Section III-C, we suppose there is a dictionary $\mathbf{D}$ with sufficient words. Along with the coefficients $\mathbf{C}$, suitable words can be selected from $\mathbf{D}$ to form a sentence for describing the HRRS images. In this section, we will explain how to get the appropriate $\mathbf{D}$. To enhance the effectiveness of words for our change detection tasks, we develop a simple analysis method for $\mathbf{D}$ and the analyzed results will be added to $\mathbf{F}$ (see Fig. 4). Such that the dictionary and word coefficients can be optimized simultaneously.

The flowchart of the dictionary analysis module is shown in Fig. 5. Assume that the N_w words' length is l, so that the size of $\mathbf{D}$ is $l \times N_w$. First, the word vectors within the dictionary are averaged into the word value $\mathbf{w} = [w_1, \ldots, w_{N_w}]^T$ for exploring word characteristics and representing various words simply. Second, two FC layers with the weight of $N_w \to 2 \cdot N_w \to N_c$ are employed to capture the relations between different words and output the relation vector $\mathbf{r} = [r_1, \ldots, r_{N_c}]^T$. Finally, the relation vector $\mathbf{r}$ is added to each feature pixel within $\mathbf{F}$.

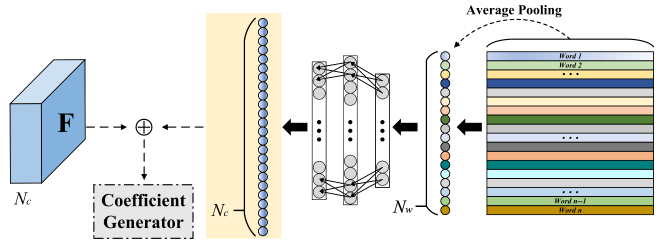


Fig. 5. Flowchart of the dictionary analysis module. First, each word vector within the dictionary is pooled averagely to one value. Second, all values are inputted into two FC layers to capture the relations between different words. Finally, the relation vector with the length of N_c is added to each feature pixel within $\mathbf{F}$ in channel dimension.

Furthermore, to accurately align the land-cover representation in different periods, the dictionary should contain sufficient and normalized information. Therefore, the deep dictionary is expected to be an orthogonal matrix (OM). On the one hand, the full-rank property of OM ensures that the information within the dictionary is sufficient. On the other hand, row vectors of OM are the unit vectors, which guarantee the dictionary is regularized. To this end, we develop the following loss function for $\mathbf{D}$:

$$\mathbf{O} = \mathbf{D}\mathbf{D}^T - \mathbf{E}$$
$$\mathcal{L}_{\text{om}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{O}_{i,j} \tag{5}$$

where $\mathbf{E}$ denotes the unit matrix and n represents the row/column number of the square matrix $\mathbf{O}$.

E. Change Map Generation

So far, we can get the sentences $S_A = \mathbf{D}\cdot\mathbf{C}_A$ and $S_B = \mathbf{D}\cdot\mathbf{C}_B$ for bitemporal HRRS images. Then, they are concatenated and inputted into a classifier to generate the change map. The classifier is composed of two 1×1 convolution layers and outputs 2-D prediction. To narrow the gaps between the prediction and ground truth, we select the cross-entropy loss function which is defined as

$$\mathcal{L}_c = -\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{p_i}\cdot\log\, p_i + \left(1-\widehat{p_i}\right)\cdot\log\,(1-p_i)\right) \tag{6}$$

where $\hat{p}$ and p denote the ground truth and prediction, respectively, and n is the number of image pixels. Combining $\mathcal{L}_c$ and $\mathcal{L}_{\text{om}}$ together, the final loss function for training our FDLdet is formulated as

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_{\text{om}}. \tag{7}$$

IV. Experiments

A. Dataset Introduction

Three public large-scale datasets are employed to evaluate the performance of our FDLdet, including season-varying change detection (SVCD) [55], Sun Yat-sen University change detection (SYSU-CD) [48], and LEVIR change detection (LEVIR-CD) [56]. The corresponding visual samples of three datasets are exhibited in Fig. 6.

1) SVCD: This dataset is obtained from Google Earth (GE). There are 11 season-varying HRRS pairs in SVCD, including seven real HRRS pairs with the size of 4725×2700 and four synthetic HRRS image pairs with the size of 1900×1000. Here, the synthetic pairs generate land-cover changes by incorporating additional objects manually. The spatial resolution of HRRS images in SVCD varies from 0.03 to 1 m per pixel. For easy access, the original HRRS pairs within SVCD are randomly rotated and cropped into 16 000 HRRS image patches with the size of 256×256 first. Then, the obtained HRRS patches are further divided into the training, validation, and testing sets, whose volumes are 10 000, 3000, and 3000.

2) SYSU-CD: This archive incorporates 20 000 aerial image patch pairs with a spatial resolution of 0.5 m and an image patch size of 256×256, which were constructed according to 800 original HRRS images which cover the Hong Kong area from 2007 and 2014. The SYSU-CD dataset contains many high-rise buildings, which are easily affected by illumination, shadows, and other environmental factors. Thus, the difficulties of change detection using this dataset are increased. According to the literature [48], we randomly select 10 000/4000/4000 image pairs for training/validation/testing sets.

3) LEVIR-CD: It comprises 637 HRRS images, each with a resolution of 0.5 m per pixel and an image size of 1024×1024 pixels. These images are sourced from GE and specifically capture changes in diverse building types across 20 distinct regions over the period from 2002 to 2018. In this article, we crop each image from 1024×1024 to $16 \times 256 \times 256$. According to the original division, the training, validation, and testing sets contain 7120, 1024, and 2048 image pairs.

B. Experimental Settings

The proposed FDLdet is implemented on the Pytorch [57] platform, and all experiments are conducted on the high-performance computer with GeForce RTX 3090 of 24-GB memory and Inter Xenon Silver 4214R. The backbone of our feature extractor (i.e., ResNet) is initialized with the pre-trained parameters (obtained by the ImageNet dataset [58]). The rest components of FDLdet are initialized randomly. We select the Adam optimizer to train FDLdet, and the batch size and epoch are set to 32 and 200, respectively. The initial learning rate is set to $1e^{-3}$, and it is multiplied by $1e^{-1}$ after every 60 epochs. In the training stage, some data augmentation schemes are used to improve the robustness of FDLdet. First, the bitemporal HRRS images are rotated with the same angle, which is randomly valued in a range of $[-180^\circ, 180^\circ]$. Second, the random vertical and horizontal flips are applied to them orderly. Third, two image patches are randomly cropped from two processed HRRS images, and the cropping scale is in a range of $[0.7, 1.0]$. Finally, the augmented images are normalized and input into FDLdet for training. In the inference stage, the bitemporal images are directly normalized and inputted in the trained FDLdet. The proposed FDLdet involves two hyperparameters, including the numbers of words within deep dictionary N_w and the volume of WRs N_r. Their empirical values and influence on FDLdet will be discussed in Section IV-C.

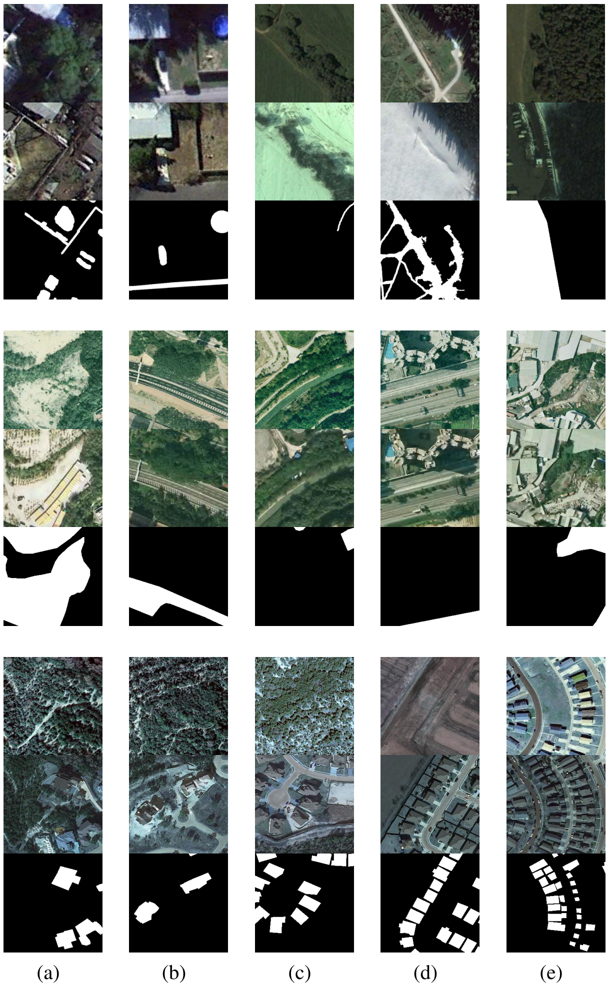


Fig. 6. Ten examples of the SVCD (first three rows), SYSU-CD (second three rows), and LEVIR-CD (last three rows) datasets. The first/fourth/seventh and second/fifth/eighth rows of (a)–(e) are aerial images of the former and later periods, respectively. Their ground truths are listed in the third/sixth/ninth rows of (a)–(e).

TABLE I
ANALYSIS OF TWO HYPERPARAMETERS (%)

Data Sets	Parameters		P	R	F	IoU	OA
SVCD	N_w	16	95.08	93.87	94.47	89.53	98.64
		32	**96.04**	94.29	**95.16**	**90.76**	**98.81**
		64	95.12	**94.96**	95.04	90.55	98.77
		128	95.36	94.52	94.94	90.37	98.75
	N_r	100	95.39	94.56	94.97	90.43	98.82
		200	**96.04**	94.29	**95.16**	**90.76**	98.81
		300	95.43	**94.64**	95.03	90.53	**98.83**
		400	95.89	94.06	94.96	90.41	98.82
SYSU-CD	N_w	16	81.84	74.15	77.81	63.68	90.03
		32	83.04	74.62	**78.61**	**64.75**	**90.42**
		64	**83.54**	73.28	78.07	64.03	90.29
		128	79.23	**76.70**	77.94	63.86	89.76
	N_r	100	78.47	**77.76**	78.11	64.09	89.72
		200	**83.04**	74.62	**78.61**	**64.75**	**90.42**
		300	79.58	76.96	78.25	64.27	89.91
		400	80.21	76.04	78.06	64.03	89.93
LEVIR-CD	N_w	16	90.58	**90.18**	90.38	82.45	99.02
		32	**91.88**	89.74	**90.80**	**83.15**	**99.07**
		64	91.06	90.09	90.57	82.77	99.04
		128	91.32	89.12	90.21	82.16	99.01
	N_r	100	90.74	90.52	90.63	82.86	99.05
		200	**91.88**	89.74	**90.80**	**83.15**	**99.07**
		300	90.08	91.35	90.71	83.00	99.05
		400	89.77	**91.40**	90.58	82.78	99.03

To quantitatively evaluate the performance of FDLdet, five assessment criteria are adopted [16], i.e., precision (P), recall (R), F1-score (F), intersection over union (IoU), and overall accuracy (OA). To calculate these evaluation metrics, we first count the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) at the pixel level. Here, TP and FP denote the number of changed pixels that are detected correctly and incorrectly, while TN and FN represent the number of unchanged pixels that are predicted correctly and incorrectly. Then, the definitions of the five assessment criteria are

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \quad F = \frac{2 \cdot P \cdot R}{P+R}$$
$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}, \quad \mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}. \quad (8)$$

C. Parameter Analysis

As mentioned before, two hyperparameters should be set in advance, i.e., the word number N_w and the region number N_r. To analyze their influences on FDLdet, we change their values alternately. For N_w, its values vary in [16, 32, 64, 128]. For N_r, its values are changed from 100 to 400 with an interval of 100. Note that we fix $N_w = 32$ and $N_r = 200$ when the other releases the changes. The results counted on three datasets are listed in Table I.

For SVCD, it is easy to find that various FDLdets with different N_w and N_r values perform well, and their behavior is close to each other. This indicates that FDLdet is not sensitive to these two parameters on the SVCD dataset. The best performance of FDLdet can be achieved when the values of $[N_w, N_r]$ equal [32, 200] (P, F, and IoU), [64, 300] (R), and [32, 300] (OA), respectively. Considering the tradeoff between performance and computational costs, we suggest the optimal values of N_w and N_r are set to 32 and 200 for SVCD. Different from the observations found in SVCD, the influences of N_w and N_r on FDLdet counted by the SYSU-CD dataset are noticeable. This is mainly because of the complexity of SYSU-CD, which encloses more diverse land covers. For F, IoU, and OA, the strongest FDLdet can be obtained when the values of $[N_w, N_r]$ are equal to [32, 200]. Nevertheless, For P and R, FDLdet can reach the highest performance when the values of $[N_w, N_r]$ are [64, 200] and [128, 100], respectively. Taking various aspects into account, we suggest the proper values of N_w and N_r are still 32 and 200 for SYSU-CD. For the LEVIR-CD dataset, altering hyperparameters does not result in significant performance changes, as the F1-score consistently exceeds 90 across all cases. Furthermore, it is evident that the optimal model performance is attained when the values

TABLE II
PERFORMANCE OF DIFFERENT METHODS COUNTED ON SVCD (%)

Methods	P	R	F	IoU	OA
FC-EF [21]	89.87	66.49	76.43	61.85	94.74
FC-Siam-conc [21]	89.50	73.67	80.82	67.81	95.52
FC-Siam-diff [21]	88.62	74.78	81.11	68.23	95.53
CDNet [59]	73.40	95.31	82.93	70.84	95.37
STANet [56]	87.35	95.77	91.36	84.10	97.86
BIT [51]	95.31	87.31	91.13	83.71	98.00
CLNet [60]	93.30	89.80	91.52	84.36	98.03
DSAMNet [48]	90.16	**98.42**	94.10	88.87	98.41
SNUNet [61]	94.82	92.45	93.62	88.00	98.51
ESCNet [2]	92.86	96.06	94.43	89.45	98.60
IFN [62]	**97.15**	91.70	94.34	89.29	98.70
ISNet [49]	95.18	94.43	94.80	90.12	98.78
FDLdet (our)	96.04	94.29	**95.16**	**90.76**	**98.81**

TABLE IV
PERFORMANCE OF DIFFERENT METHODS COUNTED ON LEVIR-CD (%)

Methods	P	R	F	IoU	OA
FC-EF [21]	87.37	52.46	65.56	48.76	97.19
FC-Siam-conc [21]	90.76	58.95	71.47	55.61	97.60
FC-Siam-diff [21]	91.97	57.84	71.02	55.06	97.60
CDNet [59]	82.26	82.95	82.60	70.36	98.22
STANet [56]	89.48	**89.89**	89.68	81.30	97.09
BIT [51]	89.24	89.37	89.31	80.68	98.92
CLNet [60]	90.07	85.70	87.83	78.30	98.79
DSAMNet [48]	80.61	88.98	84.59	73.29	98.35
SNUNet [61]	91.80	88.53	90.14	82.04	99.01
ESCNet [2]	87.09	86.27	86.68	76.49	98.65
IFN [62]	86.95	75.24	80.67	67.61	98.16
ISNet [49]	**92.46**	88.27	90.32	82.35	99.04
FDLdet (our)	91.88	89.74	**90.80**	**83.15**	**99.07**

TABLE III
PERFORMANCE OF DIFFERENT METHODS COUNTED ON SYSU-CD (%)

Methods	P	R	F	IoU	OA
FC-EF [21]	**84.77**	68.63	75.85	61.09	89.84
FC-Siam-conc [21]	83.03	71.60	76.90	62.46	89.90
FC-Siam-diff [21]	79.78	74.38	76.99	62.58	89.83
CDNet [59]	77.64	77.90	77.77	63.63	89.50
STANet [56]	70.26	**86.18**	77.41	63.15	88.14
BIT [51]	79.18	77.01	78.08	64.04	89.80
CLNet [60]	79.62	74.97	77.22	62.90	89.57
DSAMNet [48]	77.11	79.90	78.48	64.59	89.67
SNUNet [61]	78.26	74.42	76.30	61.68	89.10
ESCNet [2]	82.38	73.05	77.44	63.18	89.96
IFN [62]	82.39	73.57	77.73	63.57	90.06
ISNet [49]	80.27	76.41	78.29	64.44	90.01
FDLdet (our)	83.04	74.62	**78.61**	**64.75**	**90.42**

of N_w and N_r are set to $[32, 200]$, yielding the highest P, F, IoU, and OA. Therefore, if the readers want to apply the proposed model to other datasets, these two parameters' values are recommended for tuning around 32 and 200. In addition, we also recommend adjusting the number of superpixels to be less/more than 200 when the image size is smaller/larger than 256×256 to ensure computational efficiency and detail capture.

D. Compare With State-of-the-Art Methods

We adopt 12 methods to testify our FDLdet, including FC-EF [21], FC-Siam-conc [21], FC-Siam-diff [21], change detection network (CDNet) [59], spatial–temporal attention neural network (STANet) [56], BIT [51], cross layer convolutional neural network (CLNet) [60], DSAMNet [48], siamese network and nested UNet (SNUNet) [61], end-to-end superpixel-enhanced change detection network (ESCNet) [2], image fusion network (IFN) [62], and improved separability network (ISNet) [49]. The results of different methods are listed in Table II (SVCD), Table III (SYSU-CD), and Table IV (LEVIR-CD). Furthermore, we show the visualized results of all methods in Figs. 7–9.

For the SVCD dataset, the proposed FDLdet achieves the highest F, IoU, and OA values among all methods, which demonstrates its effectiveness. Taking OA values as examples, compared with the other 12 methods, the improvements of FDLdet are 4.07% (over FC-EF), 3.29% (over FC-Siam-conc), 3.28% (over FC-Siam-diff), 3.44% (over CDNet), 0.95% (over STANet), 0.81% (over BIT), 0.78% (over CLNet), 0.40% (over DSAMNet), 0.30% (over SNUNet), 0.21% (over ESCNet), 0.11% (over IFN), and 0.03% (over ISNet), respectively. The identical investigation can also be found in the results counted on the SYSU-CD dataset, i.e., our FDLdet performs the best under the indicators of F, IoU, and OA. For instance, the enhancements in F values achieved by FDLdet are 2.76% (over FC-EF), 1.71% (over FC-Siam-conc), 1.62% (over FC-Siam-diff), 0.84% (over CDNet), 1.20% (over STANet), 0.53% (over BIT), 1.39% (over CLNet), 0.13% (over DSAMNet), 2.31% (over SNUNet), 1.17% (over ESCNet), 0.88% (over IFN), and 0.31% (over ISNet), respectively. Similar to previous datasets, the model also achieves the best performance (F, IoU, and OA) on the LEVIR-CD dataset. Specifically, FDLdet surpasses the performance of the second-best method, ISNet, by 0.48% in terms of F and 0.8% in terms of IoU. These positive results are owing to the following points. First, the proposed feature extractor can learn the multiscale features from HRRS images, which are able to describe the diverse and complex land covers accordingly. Second, the pixel- and region-level knowledge is further explored based on those deep features, which is conducive to analyzing the essential information hidden in HRRS images. By integrating the above two parts and embedding them into a change detection-oriented DL framework, the change maps can be produced. Along with the "orthogonalization" constraint, both the completeness of the learned deep dictionary and the accuracies of change maps can be ensured.

However, our model's P and R values are not optimal compared to other methods. For SVCD, the best P and R values

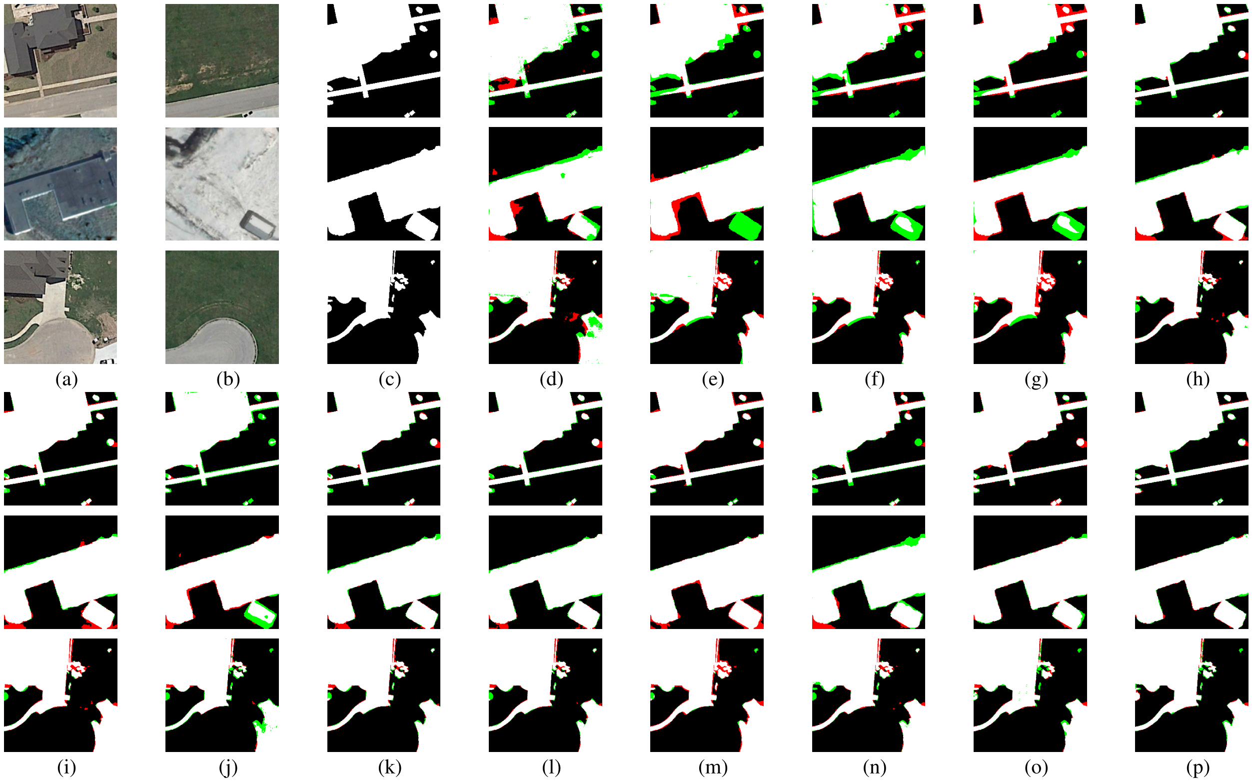


Fig. 7. Visual results of six examples for SVCD dataset. (a) Image T_1. (b) Image T_2. (c) Ground truth. (d) FC-EF. (e) FC-SC. (f) FC-SD. (g) CDNet. (h) STANet. (i) BIT. (j) CLNet. (k) DSAMNet. (l) SNUNet. (m) ESCNet. (n) IFN. (o) ISNet. (p) FDLdet (our). Here, white, black, red, and green denote TP, TN, FP, and FN, respectively.

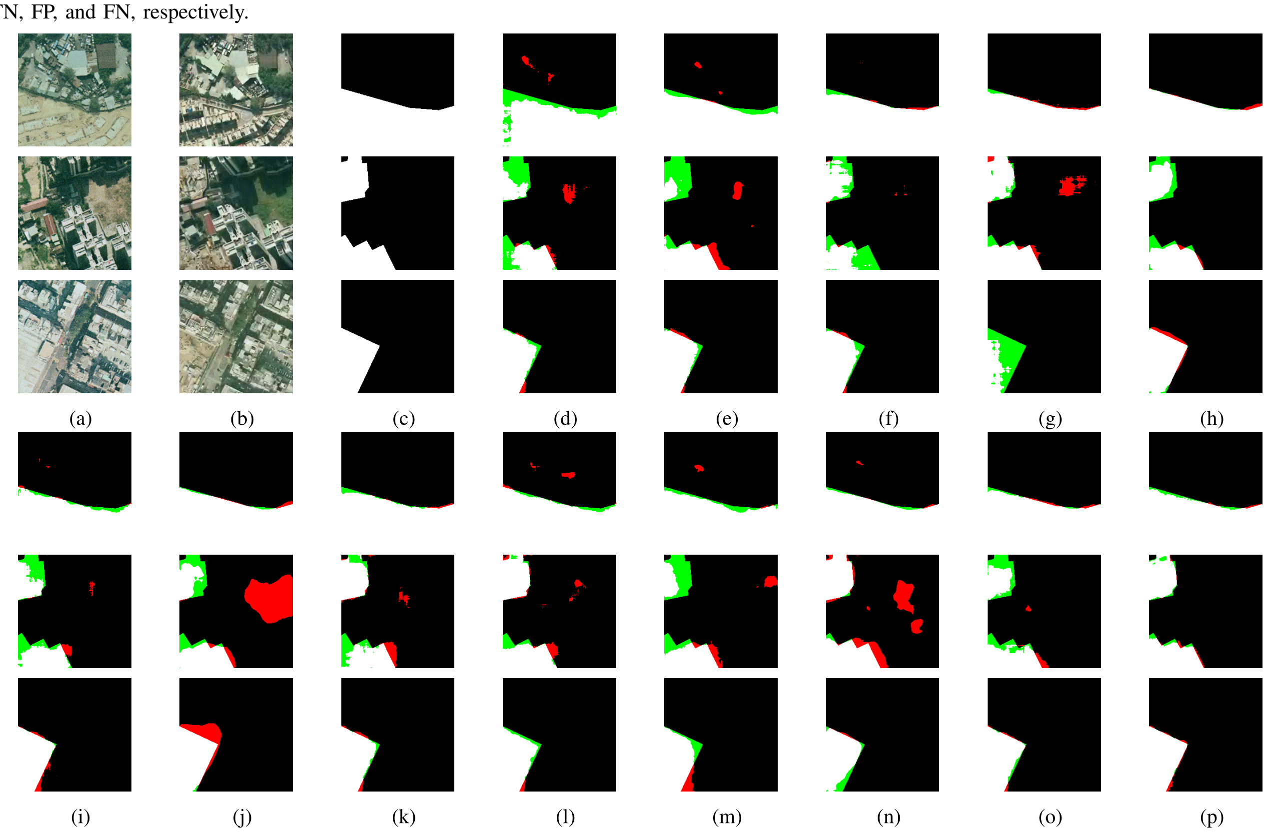


Fig. 8. Visual results of six examples for SYSU-CD dataset. (a) Image T_1. (b) Image T_2. (c) Ground truth. (d) FC-EF. (e) FC-SC. (f) FC-SD. (g) CDNet. (h) STANet. (i) BIT. (j) CLNet. (k) DSAMNet. (l) SNUNet. (m) ESCNet. (n) IFN. (o) ISNet. (p) FDLdet (our). Here, white, black, red, and green denote TP, TN, FP, and FN, respectively.

are achieved by IFN [62] and DSAMNet [48], separately. For SYSU-CD, FC-EF [21] and STANet [56] perform best in P and R, respectively. For LEVIR-CD, STANet [56] and ISNet [49] achieve the peak values in P and R. The main

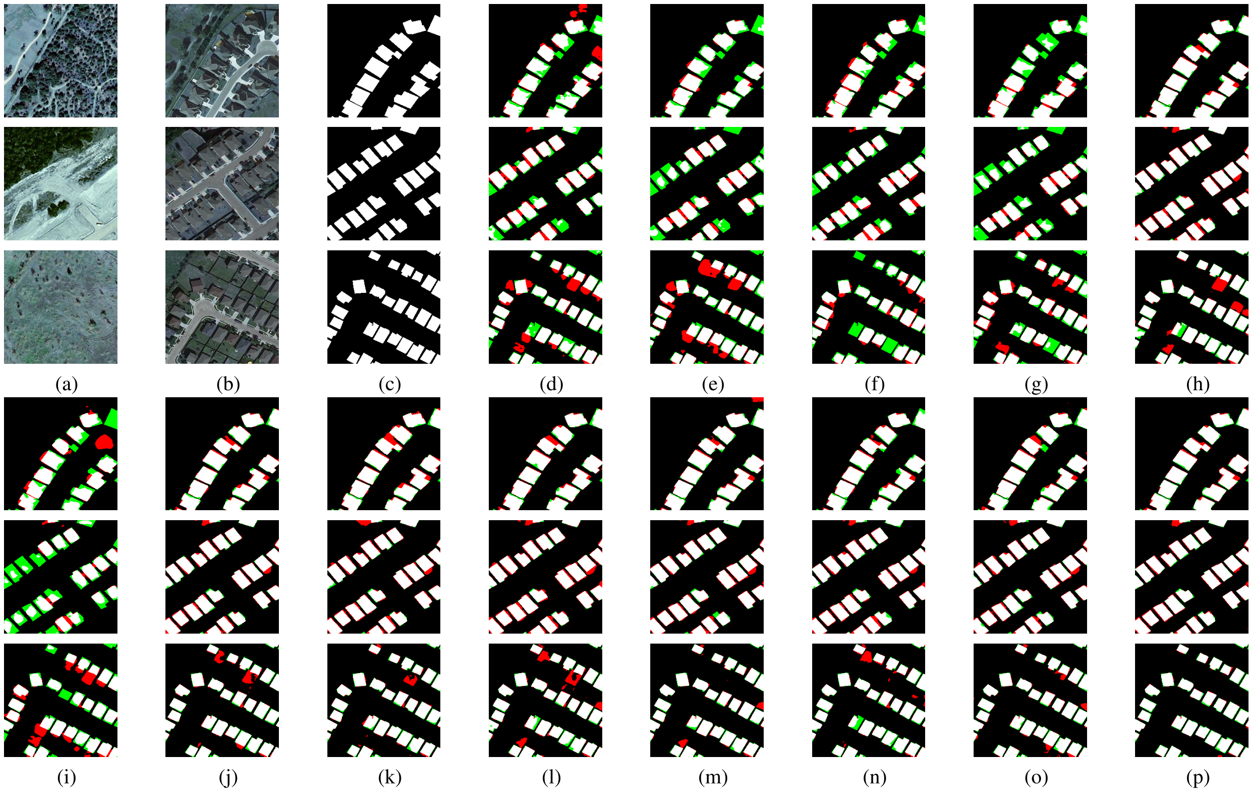


Fig. 9. Visual results of six examples for LEVIR-CD dataset. (a) Image T_1. (b) Image T_2. (c) Ground truth. (d) FC-EF. (e) FC-SC. (f) FC-SD. (g) CDNet. (h) STANet. (i) BIT. (j) CLNet. (k) DSAMNet. (l) SNUNet. (m) ESCNet. (n) IFN. (o) ISNet. (p) FDLdet (our). Here, white, black, red, and green denote TP, TN, FP, and FN, respectively.

reasons behind this can be summarized as follows. P and R are two mutually contradictory indexes since the number of FP and FN pixels cannot be small simultaneously. If models pay more attention to P values, their R values must be relatively low, and vice versa. Taking IFN in SVCD as an example, its P value is as high as 97.15%. Nevertheless, its R value is only 91.70%, which is uncompetitive among various methods. Therefore, an expected change detection model should keep a balance between them. Looking back to our FDLdet, although its P and R values are not the best, their differences are slight. This implies that FDLdet is a robust model. Furthermore, the performance gaps (measured by P and R) between FDLdet and the optimal methods are small, and FDLdet's P and R values are still at the top ranks. This demonstrates that the proposed model is helpful for HRRS change detection.

To further visually show the advantages of FDLdet, we randomly select six sample pairs from three datasets and visually report their change maps in Figs. 7–9, where the pixels of TP, TN, FP, and FN within change maps are marked in white, black, red, and green for convenience. By observing Figs. 7–9, it is evident that FDLdet has fewer FP (red color) and FN (green color) pixels than the other methods. For FC-EF, STANet, BIT, and IFN, red and green colors take up a large portion of the prediction map, meaning they have high error rates in predicting changed and unchanged pixels. For DSAMNet, SNUNet, ESCNet, and ISNet, the areas in red and green colors are smaller than in previous methods. However, some similar regions are mispredicted. In comparison, the quality of the change maps produced by FDLdet is higher, which are closer to the ground truths. These satisfying visual results illustrate the usefulness of our FDLdet again.

E. Ablation Study

In this section, we study the contributions of FDLdet's different components. It can be regarded as an enhanced DL framework embedded with a basic feature extractor, and the enhanced DL framework can be deemed as a forward DL with a regional clue supplement and a dictionary "orthogonalization" constraint. To study their significance to FDLdet, we construct the following four networks.

1) Net_1: Feature extractor.
2) Net_2: Feature extractor + forward DL.
3) Net_3: Feature extractor + forward DL + regional clue.
4) Net_4: Feature extractor + forward DL + regional clue + orthogonalization.

Net_1 means that only the deep features extracted by the feature extractor are used to predict change maps directly. Net_2 represents that the forward DL is added after the feature extractor to estimate the change areas. Net_3 denotes that the regional information of HRRS images is complemented to the forward DL. Net_4 is our FDLdet that encloses the feature extractor and the enhanced DL.

The performance of different nets counted on SVCD, SYSU-CD, and LEVIR-CD is exhibited in Table V. For the SVCD dataset, we can find that the performance of four nets is incremental in all cases, which illustrates that each component plays a positive role. Taking IoU as examples, compared with Net_i, $i = 1, 2, 3$, the enhancements achieved by Net_{i+1} are 0.63%, 0.68%, and 0.95%, respectively. The reasons behind this are threefold. First, the developed forward DL is good at maintaining the consistency of HRRS image features, which helps the net find the changed/unchanged information (see

TABLE V
ABLATION EXPERIMENTAL RESULTS OF FOUR NETWORKS (%)

Data Set	Network	P	R	F	IoU	OA
SVCD	*Net*$_1$	94.37	93.43	93.90	88.50	98.49
	Net$_2$	94.81	93.71	94.26	89.13	98.58
	Net$_3$	95.34	93.93	94.63	89.81	98.68
	Net$_4$	**96.04**	**94.29**	**95.16**	**90.76**	**98.81**
SYSU-CD	*Net*$_1$	75.96	**79.43**	77.65	63.47	89.22
	Net$_2$	80.44	75.88	78.10	64.06	89.96
	Net$_3$	80.72	76.22	78.40	64.48	90.10
	Net$_4$	**83.04**	74.62	**78.61**	**64.75**	**90.42**
LEVIR-CD	*Net*$_1$	88.85	89.63	89.24	80.57	98.90
	Net$_2$	89.64	**90.62**	90.13	82.03	98.99
	Net$_3$	90.64	89.81	90.22	82.18	99.01
	Net$_4$	**91.88**	89.74	**90.80**	**83.15**	**99.07**

TABLE VI
PERFORMANCE OF FDLDETS WITH DIFFERENT COEFFICIENTS (%)

	P	R	F	IoU	OA
Sparse	95.16	93.30	94.22	89.07	98.58
Grid	95.34	94.15	94.74	90.01	98.70
Slic	**96.04**	**94.29**	**95.16**	**90.76**	**98.81**

Net_1 and Net_2). Second, the mined regional clues assist the net in discovering the essential information of HRRS images so that the accuracies of the changed areas can be improved (see Net_2 and Net_3). Third, the "orthogonalization" constraint ensures that the learned dictionary could provide sufficient and normalized messages for change detection (see Net_3 and Net_4). Besides, an encouraging observation is that there is a distinct performance gap between Net_4 and Net_1 (1.67% of P, 0.86% of R, 1.26% of F, 2.26% of IoU, and 0.32% of OA). This further confirms the usefulness of the forward DL and two supplements. For the SYSU-CD and LEVIR-CD dataset, we can find similar conclusions in most cases, i.e., the behavior of Net_{i+1} is stronger than that of Net_{i+1}, where $i = 1, 2, 3$. Fortunately, the performance of Net_4 is the best among all nets in other assessment criteria.

F. In-Depth Components Study

This section further investigates the enhanced DL framework of FDLdet. As mentioned in the ablation study, the enhanced DL framework is a forward DL with the regional clues (for coefficients) and the orthogonal constraint (for dictionary). Thus, we will study the enhanced DL framework from the following aspects. First, the effectiveness of the regional clues will be discussed. Second, the necessity of dictionary orthogonalization. Third, the function of the forward DL. Finally, the superiorities of the enhanced DL framework will be studied visually.

1) Effectiveness of Regional Clues: As mentioned in Section III-C, the regional clues (obtained by SLIC) of HRRS images are taken into account to generate the word coefficients. We design the following experiments to further study how those clues help our FDLdet. First, we remove the regional clues during the coefficient generation. In other words, the coefficients are generated by only the learned feature maps. Note that the sparse constraint ($L1$ norm) would be added to the coefficients in this case. Second, instead of using the SLIC algorithm, we use a simple grid scheme to divide an HRRS image into regular regions. Then, the information of these regular regions would be added to produce coefficients. Third, we leverage the SLIC algorithm to mine the irregular regions and merge them into coefficients. We name the three mentioned scenarios "Sparse," "Grid," and "Slic" for convenience.

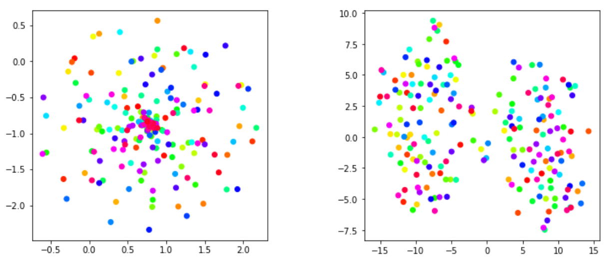
The performance of FDLdets with different coefficients is shown in Table VI. From the observation of the results, we can find that the behavior of FDLdet with "Sparse" is the weakest. This is mainly because: 1) the "Sparse" coefficients only incorporate the pixel-level information which cannot reflect the spatial relations within HRRS images and 2) the sparse constraint puts most word coefficients toward zero, which leads to the information loss. When the regular regions are added (i.e., the "Grid" coefficients), the performance of FDLdet is enhanced distinctly. For instance, the IoU values are increased from 89.07% to 90.01%. This indicates that the regional information can improve the coefficients for the final change detection tasks. Once we incorporate the irregular regions obtained by the SLIC algorithm, the obtained word coefficients promote the performance of FDLdet significantly. For example, the F-values rise from 94.74% to 95.16%. The reason is that the irregular regions generated by "Slic" contain rich homogeneous messages, which is conducive to identifying the changed/unchanged information.

To further show the superiority of "Slic" over "Grid" visually, we randomly select a pair of bitemporal HRRS images from the SVCD dataset and generate corresponding regional pooling results. The chosen HRRS images and their change detection ground truth are shown in Fig. 10(a) and (b). First, we use the SLIC algorithm to over-segment two HRRS images. The results are displayed in Fig. 10(c), whose numbers of irregular regions are 1731 and 1670, respectively. After using average pooling in the channel dimension, the "Slic" regional pooling maps can be produced [see Fig. 10(d)]. Then, the simple grid scheme is applied to bitemporal HRRS images to partition them into regular regions. For the sake of fairness, the grid size is set to 42×42 so that the number of regular regions is 1764. The "Grid" regional pooling maps are also obtained by the channel-wise average pooling, and they are exhibited in Fig. 10(e). Finally, the "Slic" and "Grid" regional pooling maps are masked by the ground truth, and the masked maps are shown in Fig. 10(f) and (g) separately. By observing these two kinds of masked maps, we can find that the information involved in the "Slic" regional pooling maps is more complete than the "Grid" maps. For example, the shape and texture of land covers are apparent, and their edges are clear and smooth. These profitable materials could support our FDLdet to achieve satisfactory change detection results.

2) Necessity of Dictionary Orthogonalization: To ensure the learned dictionary covers sufficient information for change detection, we add an orthogonal constraint to it [see (5)]. Here, we visually study the positive effect of dictionary

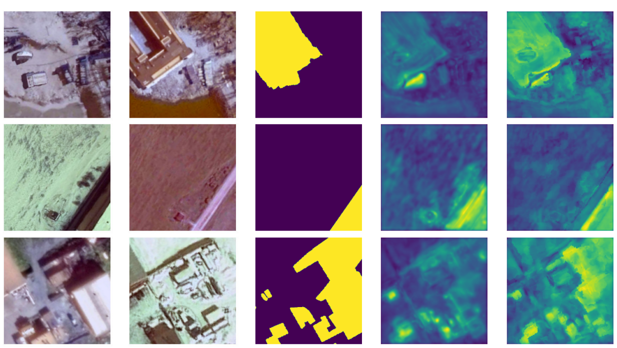
(a) (b) (c) (d) (e) (f) (g)

Fig. 10. Visual exhibition of “Slic” and “Grid” regional pooling results. (a) Ground truth. (b) Bitemporal HRRS images. (c) Irregular regions obtained by SLIC. (d) “Slic” regional pooling maps. (e) “Grid” regional pooling maps. (f) “Slic” masked maps. (g) “Grid” masked maps.



(a) (b)

Fig. 11. 2-D scatterplots of the words from two dictionaries. (a) Orthogonized dictionary. (b) Non-orthogonized dictionary.



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

Fig. 12. Feature and sentence DIs of three temporal HRRS image pairs. (a) HRRS images of the former period. (b) HRRS images of the later period. (c) Ground truths. (d) Feature DIs. (e) Sentence DIs.

orthogonalization. First, two dictionaries are trained with/without the orthogonal constraint. Then, the t-distributed stochastic neighbor embedding (T-SNE) algorithm [63] is adopted to reduce the dictionaries’ dimension. Finally, the 2-D words of the two dictionaries are displayed in Fig. 11. It is easy to find that the words within the orthogonal dictionary follow an approximate uniform distribution, while the distribution of words from the ordinary dictionary is disordered and biased. This incomplete feature distribution of the ordinary dictionary may cause the model to struggle to distinguish similar land covers. Therefore, the orthogonal dictionary takes more advantages for accurately distinguishing the land-cover changes.

3) Superiorities of Enhanced DL Framework: To visually exhibit the enhanced DL framework’s advantages, three pairs of bitemporal HRRS images are randomly picked up from SVCD dataset. Then, we put them into a trained FDLdet and collect the image features $\{\mathbf{F}_A^i, \mathbf{F}_B^i, i = 1, 2, 3\}$ (before DL) and the sentences $\{\mathbf{S}_A^i, \mathbf{S}_B^i, i = 1, 2, 3\}$ (after DL). Next, the Euclidean distance is used to generate the feature DIs $\{\mathbf{DI}_F^i, i = 1, 2, 3\}$ and sentence DIs $\{\mathbf{DI}_S^i, i = 1, 2, 3\}$. The bitemporal HRRS images, their ground truths, and the different DIs are shown in Fig. 12. From observing them, we can find that compared with feature DIs, the sentence DIs are closer to the ground truth and involve more detailed information. These findings indicate that: 1) the enhanced DL framework could improve the consistency of temporal image features and 2) besides the visual contents, the sentences represent essential information hidden in HRRS images. The encouraging investigations demonstrate that the enhanced DL framework positively impacts change detection tasks.

V. Conclusion

In this article, a change detector FDLdet based on forward DL is proposed to capture the potential land-cover changes between bitemporal HRRS images. FDLdet consists of three main components, including a feature extractor, a coefficient generator, and a deep dictionary. Feature extractor aims to extract the image features from HRRS images. The coefficient generator is employed to transform image features into regional word coefficients. The words within the dictionary are weighted and combined by word coefficients to generate dictionary features for describing HRRS images. Finally, the land-cover changes can be accurately captured by simply analyzing differences between dictionary features. Compared with the existing DL, the proposed forward DL explores the essential feature of HRRS images, removing the complicated reverse output and sparse procedure. The promising experimental results on three public datasets demonstrate the effectiveness of FDLdet.

References

[1] D. Wen et al., “Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions,” *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.

[2] H. Zhang, M. Lin, G. Yang, and L. Zhang, “ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 28–42, Jan. 2023.

[3] Z. Zhu, S. Qiu, and S. Ye, “Remote sensing of land change: A multifaceted perspective,” *Remote Sens. Environ.*, vol. 282, Dec. 2022, Art. no. 113266.

[4] Y. Li, B. Dang, Y. Zhang, and Z. Du, “Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives,” *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 306–327, May 2022.

[5] N. Wang, W. Li, R. Tao, and Q. Du, "Graph-based block-level urban change detection using Sentinel-2 time series," *Remote Sens. Environ.*, vol. 274, Jun. 2022, Art. no. 112993.
[6] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, Apr. 2015.
[7] L. I. Kuncheva and W. J. Faithfull, "PCA feature extraction for change detection in multidimensional unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 69–80, Jan. 2014.
[8] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.
[9] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.
[10] M. Yang, L. Jiao, F. Liu, B. Hou, S. Yang, and M. Jian, "DPFL-nets: Deep pyramid feature learning networks for multiscale change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6402–6416, Nov. 2022.
[11] L. T. Luppino et al., "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 60–72, Jan. 2024.
[12] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.
[13] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016.
[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 7553.
[15] G. Novakovsky, N. Dexter, M. W. Libbrecht, W. W. Wasserman, and S. Mostafavi, "Obtaining genetics insights from deep learning via explainable artificial intelligence," *Nature Rev. Genet.*, vol. 24, no. 2, pp. 125–137, Feb. 2023.
[16] X. Tang et al., "An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
[17] J. Liu, M. Gong, A. K. Qin, and K. C. Tan, "Bipartite differential neural network for unsupervised image change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 876–890, Mar. 2020.
[18] Y. Wu, J. Li, Y. Yuan, A. K. Qin, Q.-G. Miao, and M.-G. Gong, "Commonality autoencoder: Learning common features for change detection from heterogeneous images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4257–4270, Sep. 2022.
[19] Z. Lv, T. Liu, J. A. Benediktsson, and N. Falco, "Land cover change detection techniques: Very-high-resolution optical images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 44–63, Mar. 2022.
[20] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 982–986, Jun. 2019.
[21] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 4063–4067.
[22] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, "FCCDN: Feature constraint network for VHR image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 101–119, May 2022.
[23] X. Tang et al., "Hyperspectral image classification based on 3-D octave convolution with spatial–spectral attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2430–2447, Mar. 2021.
[24] C. Liu, J. Ma, X. Tang, F. Liu, X. Zhang, and L. Jiao, "Deep hash learning for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3420–3443, Apr. 2021.
[25] Y. Yang et al., "AR^{2}Det: An accurate and real-time rotational one-stage ship detector in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
[26] S.-B. Chen, Q.-S. Wei, W.-Z. Wang, J. Tang, B. Luo, and Z.-Y. Wang, "Remote sensing scene classification via multi-branch local attention network," *IEEE Trans. Image Process.*, vol. 31, pp. 99–109, 2022.
[27] X. Tang et al., "Meta-hashing for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2021.
[28] X. Tang, M. Li, J. Ma, X. Zhang, F. Liu, and L. Jiao, "EMTCAL: Efficient multiscale transformer and cross-level attention learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5626915.
[29] J. Liu, W. Zhang, F. Liu, and L. Xiao, "A probabilistic model based on bipartite convolutional neural network for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3071347.
[30] S. Mahdizadehaghdam, A. Panahi, H. Krim, and L. Dai, "Deep dictionary learning: A parametric network approach," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 4790–4802, Oct. 2019.
[31] H. Tang, H. Liu, W. Xiao, and N. Sebe, "When dictionary learning meets deep learning: Deep dictionary learning and coding network for image recognition with limited data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2129–2141, May 2021.
[32] F. Gao, X. Deng, M. Xu, J. Xu, and P. L. Dragotti, "Multi-modal convolutional dictionary learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1325–1339, 2022.
[33] K. Ding, C. Huo, Y. Xu, Z. Zhong, and C. Pan, "Sparse hierarchical clustering for VHR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 577–581, Mar. 2015.
[34] Y. Li, M. Gong, L. Jiao, L. Li, and R. Stolkin, "Change-detection map learning using matching pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4712–4723, Aug. 2015.
[35] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Patch similarity graph matrix-based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4841–4861, Jun. 2021.
[36] W. Zhang, J. Li, F. Zhang, J. Sun, and K. Zhang, "Unsupervised change detection of multispectral images based on PCA and low-rank prior," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[37] Y. Sun, L. Lei, D. Guan, X. Li, and G. Kuang, "SAR image change detection based on nonlocal low-rank model and two-level clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 293–306, 2020.
[38] L. Li et al., "Deformable dictionary learning for SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4605–4617, Aug. 2018.
[39] Q. Yu, M. Zhang, L. Yu, R. Wang, and J. Xiao, "SAR image change detection based on joint dictionary learning with iterative adaptive threshold optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5234–5249, 2022.
[40] G. Yang, H.-C. Li, W.-Y. Wang, W. Yang, and W. J. Emery, "Unsupervised change detection based on a unified framework for weighted collaborative representation with RDDL and fuzzy clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8890–8903, Nov. 2019.
[41] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model," *IEEE Trans. Image Process.*, vol. 29, pp. 757–767, 2020.
[42] Y. Sun, L. Lei, D. Guan, and G. Kuang, "Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images," *IEEE Trans. Image Process.*, vol. 30, pp. 6277–6291, 2021.
[43] P. Akiva, M. Purri, and M. Leotta, "Self-supervised material and texture representation learning for remote sensing tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8203–8215.
[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.
[45] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[46] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, p. 1382, 2019.
[47] K. Han et al., "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.
[48] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.
[49] G. Cheng, G. Wang, and J. Han, "ISNet: Towards improving separability for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5623811.
[50] Z. Li, C. Yan, Y. Sun, and Q. Xin, "A densely attentive refinement network for change detection based on very-high-resolution bitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2022.3159544.

[51] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.

[52] Q. Li, R. Zhong, X. Du, and Y. Du, "TransUNetCD: A hybrid transformer network for change detection in optical remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2022.3169479.

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

[54] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[55] M. A. Lebedev, Y. V. Vizilter, O. V. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 565–571, May 2018.

[56] H. Chen and Z. Shi, "A spatial–temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, p. 1662, 2020.

[57] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.

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

[60] Z. Zheng, Y. Wan, Y. Zhang, S. Xiang, D. Peng, and B. Zhang, "CLNet: Cross-layer convolutional neural network for change detection in optical remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 247–267, May 2021.

[61] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

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

[63] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

Yuqun Yang (Student Member, IEEE) received the B.Sc. degree in information and computing science from Xi'an University of Technology, Xi'an, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence, Xidian University, Xi'an.

His research interests include machine learning, object detection, and image classification.

Xu Tang (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electronic circuit and system from Xidian University, Xi'an, China, in 2007, 2010, and 2017, respectively.

From 2015 to 2016, he was a Joint Ph.D. Student along with Prof. W. J. Emery at the University of Colorado at Boulder, Boulder, CO, USA. He is currently a Full Professor with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University. His research interests include remote sensing image content-based retrieval and reranking, hyperspectral image processing, remote sensing scene classification, and object detection. For more details, please refer to: https://web.xidian.edu.cn/tangxu/.

Xiangrong Zhang (Senior Member, IEEE) received the B.S. and M.S. degrees from the School of Computer Science, Xidian University, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2006.

From January 2015 to March 2016, she was a Visiting Scientist with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University. Her research interests include pattern recognition, machine learning, and remote sensing image analysis and understanding.

Jingjing Ma (Member, IEEE) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2012, respectively.

She is currently an Associate Professor with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University. Her research interests include computational intelligence and image understanding.

Yiu-Ming Cheung (Fellow, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2000.

He is currently a Chair Professor (artificial intelligence) with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, SAR, China. His research interests include machine learning, visual computing, data science, pattern recognition, multiobjective optimization, and information security.

Dr. Cheung is a fellow of American Association for the Advancement of Science (AAAS), Institution of Engineering and Technology (IET), British Computer Society (BCS), and Asia-Pacific Artificial Intelligence Association (AAIA). He is currently the Editor-in-Chief of IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE. He has also served as an Associate Editor for IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2014 to 2020, *Pattern Recognition*, *Pattern Recognition Letters* from 2022 to 2024, and *Neurocomputing*, to name a few. For details, please refer to: https://www.comp.hkbu.edu.hk/ ymc.

Licheng Jiao (Fellow, IEEE) received the B.S. degree in high voltage from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees in electronic engineering from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

From 1984 to 1986, he was an Assistant Professor with the Civil Aviation Institute of China, Tianjin, China. From 1990 to 1991, he was a Post-Doctoral Fellow with the Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China. He has authored or co-authored more than 200 scientific articles. His research interests include signal and image processing, nonlinear circuits and systems theory, wavelet theory, natural computation, and intelligent information processing.

Dr. Jiao is a member of the IEEE Xi'an Section Executive Committee and an Executive Committee Member of the Chinese Association of Artificial Intelligence. He is the Chairperson for the Awards and Recognition Committee.