






Change Detection With Cross-Domain Remote Sensing Images: A Systematic Review

Jie Chen , *Member, IEEE*, Dongyang Hou , Changxian He, Yaoting Liu , Ya Guo ,
and Bin Yang , *Member, IEEE*

Abstract—Change detection (CD) is one of the most important research areas in remote sensing. With the fast development of imaging techniques, CD using cross-domain remote sensing images (CDCD) has attracted extensive attention from the community. However, CDCD is quite challenging because of the different properties of the cross-domain remote sensing images (collected from different sensors or under different imaging conditions), which triggers difference in physical quantity, noise effect, illumination, seasonal conditions, geometric, and visual appearance. In the past decades, although many attempts have been devoted to the above challenges, a review of CDCD is still lacking. To bridge this gap, this article provides a systematic review on CDCD, with emphasis on image preprocessing (i.e., geometric registration et al.), feature representation (i.e., conventional methods and deep learning (DL)-based methods), and change detectors (i.e., similarity-based detector and joint feature-loss detector). Moreover, extensive experiments have also been conducted to compare the performance of 17 widely utilized CDCD methods. Based on this comparison, directions for future developments of CDCD, which include large-scale CDCD datasets, foundation models, and specialized models, are also discussed.

Index Terms—Change detection (CD), cross-domain remote sensing images, heterogeneous images.

I. INTRODUCTION

THE surface of the Earth is dynamic and always undergoing changes. Change detection (CD) is the process to identify such differences in remote sensing images collected at different times [1]. It helps to know what and which land surfaces have changed, and consequently plays an important role in various applications such as disaster assessment, environmental monitoring, agriculture management, and government decision-making [2], [3], [4], [5]. Research on CD has become a hot topic in the remote sensing field.

Manuscript received 18 March 2024; revised 16 May 2024; accepted 10 June 2024. Date of publication 18 June 2024; date of current version 1 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42071427, Grant 42371393, and Grant 42171457; in part by the Natural Science Foundation of Hunan Province, China under Grant 2023JJ20018 and Grant 2023JJ30655; and in part by the Major scientific and technological projects of Yunnan Province under Grant 202202AD080010. (Corresponding author: Bin Yang.)

Jie Chen, Dongyang Hou, Changxian He, and Ya Guo are with the School of Geosciences and Info-Physics, Central South University, Changsha 410083, China (e-mail: cj2011@csu.edu.cn).

Yaoting Liu and Bin Yang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: binyang@hnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3416183

With the fast development of imaging techniques, a vast amount of remote sensing images are now collected, enriching the data sources for CD and triggering the development of CD methods. Generally, CD methods can be classified into different categories based on whether labels are needed [6], or types of networks [7], or input images [8], etc. This article focuses on the perspective of input images, and thus the CD methods can be classified into two categories, i.e., CD using images from the same domain (CSDS) and CD using cross-domain images (CDCD). The former employs input images acquired by the same sensor at similar time and imaging conditions, whereas the latter employs input images that are heterogeneous or collected under different imaging conditions. CSDS is straightforward and easy to be utilized, however, its performance depends heavily on the quality of the input images [9], i.e., the domain consistency. Given that such requirements are not easy to be satisfied in real applications, the performance of CSDS is quite limited. Compared to CSDS, CDCD is more flexible and thus preferable in real applications. For example, optical remote sensing images, due to their susceptibility to weather and environmental conditions, are often affected by cloud and mist, resulting in limited information about ground objects. On the contrary, synthetic aperture radar (SAR) images can be collected continuously in all weather conditions, during the day or the night [10]. The combination of optical and SAR images can be used for CD in emergency situations [11], [12]. However, CDCD is more challenging because it requires more powerful feature representation methods so that features extracted from different domains can be compared for identification of changed information.

CDCD methods can be broadly classified into two categories: 1) methods with heterogeneous images, i.e., pre- and postchange images collected by different sensors [8], and 2) methods based on images with different imaging conditions, such as images collected at different seasons or under different illumination conditions [13], as illustrated in Fig. 1. The two categories of CDCD methods can be further illustrated by four typical cases: let us assume that the prechange image is optical, then the postchange image in case 1 is SAR; in case 2, it is optical but acquired by another sensor; in case 3, it is optical but collected in a different season; and in case 4, it is optical but collected under different sunlight conditions, as shown in Fig. 2. The input images used for CDCD exhibit significant visual differences, which correspond to numerous difficulties in CDCD, such as differences in the physical conditions, noise effects, illumination, seasonal period,

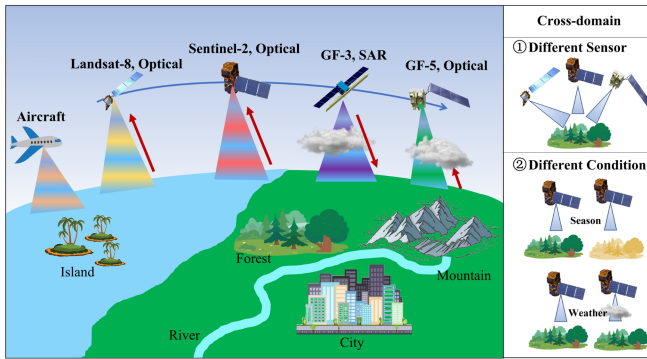


Fig. 1. Cross-domain remote sensing images collected by different sensors or under different imaging conditions.

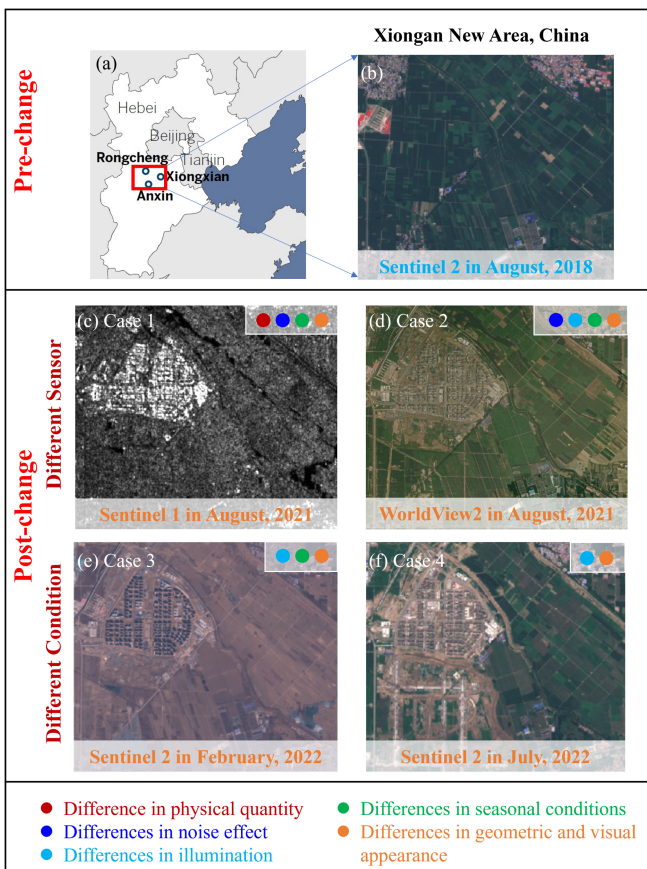


Fig. 2. Four typical cases of CDCD. (a)–(b) prechange images; (c)–(d) CDCD with heterogeneous images collected from different sensors; (e)–(f) CDCD with different imaging conditions. The color circles in the upper-right corner of (c)–(f) indicate relevant challenges.

geometric and visual factors, and differences in geometric and visual appearance. We can further describe these challenges as follows.

- 1) *Differences in Physical Quantity*: Optical images represent the ability of objects to reflect solar radiation, while SAR images represent the ability of objects to scatter actively emitted electromagnetic waves [14]. Under this circumstance, input images represent different physical

observations of the same object, and cannot be directly compared [8].

- 2) *Differences in Noise Effect*: Noise affects the quality of optical and SAR images. Optical images are mainly affected by radiometric noise [15], while SAR images are primarily affected by salt-and-pepper noise [16] and building occlusion [17]. Different optical or SAR sensors can also be impacted by different noise types and levels.
 - 3) *Differences in Illumination*: As the sensors traverse over the ground, differences in the intensity and angle of the sunlight received by the land surface may occur. This phenomenon can be caused by various factors, such as different time and weather [18]. Differences in illumination increase the variations in images, leading to unchanged areas misidentified as changed areas, thereby reducing the accuracy of CD.
 - 4) *Differences in Seasonal Conditions*: Seasonal variations can cause significant visual differences for the same object, such as deciduous forests or annual vegetation in summer and winter, snow cover and melting, among others [19], [20]. Changes caused by seasonal variations have high similarity with real changes; however, they do not belong to the category of real changes [21], [22].
 - 5) *Differences in Geometric and Visual Appearance*: It is difficult to collect images with the same observation angle due to differences in sensor orbits [23]. For example, optical sensors typically obtain images by downward viewing, while SAR sensors acquire images by side viewing [24], leading to geometric differences of the same object. Furthermore, the colors of ground objects may vary with time, or alterations in the color of building roofs can cause visual discrepancies.
- In the past decades, various CDCD methods have been proposed to address above difficulties. However, according to our best knowledge, a review of CDCD considering both heterogeneous images and images with different imaging conditions is lacking. To this end, this article provides a systematic review on CDCD methods, in which the standard processes, i.e., image preprocessing, feature representation, and change detector, are discussed. Experiments and future developments are also provided so that readers can understand the state-of-the-art in CDCD, design new methods, and extend applications.

II. MAIN CDCD TECHNIQUES

The standard processes of CDCD include three areas, i.e., image preprocessing, feature representation, and change detectors, as shown in Fig. 3. In this section, these three blocks will be introduced, with an emphasis on feature representation, which represents the core of CDCD.

A. Image Preprocessing

The aim of preprocessing remote sensing images is reducing the effect of interference factors. Preprocessing is an image-to-image operation, which only improves the quality of the image, but does not change the physical nature of the image, and thus provides better support for subsequent feature extraction and

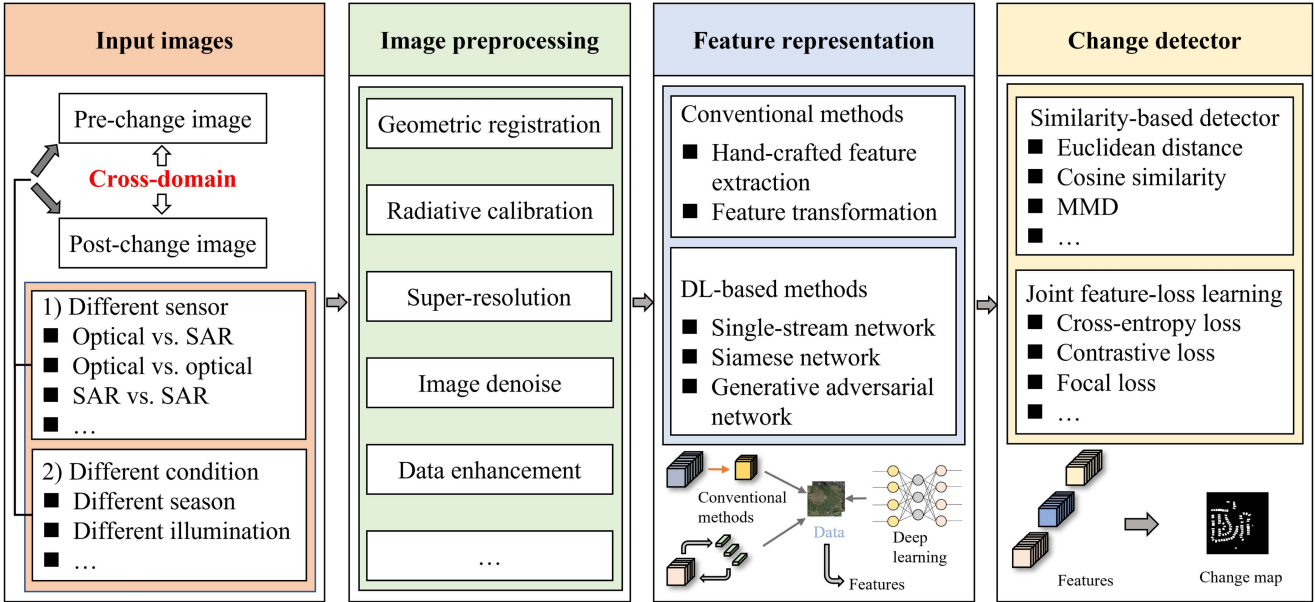


Fig. 3. General flowchart of CDCD. It includes three blocks, i.e., image preprocessing, feature representation, and change detectors.

TABLE I
TAXONOMY OF DATA PREPROCESSING METHODS FOR CDCD

Preprocessing methods	Definitions	Techniques	References
Geometric Registration	Geometric registration, which aims to correspond to the geometric positions of the image before and after the change, is a prerequisite for implementing change detection.	Conventional geometric registration methods include feature point extraction, feature point matching, transformation model establishment, image transformation, and image interpolation. Currently, various deep learning methods have been developed, as well as correction methods for registration errors.	[10], [37], [38], [39], [40], [41]
Radiative Calibration	Radiative calibration aims to eliminate the effects of different solar incidence angles, radiance levels, and atmospheric conditions on data acquisition.	Conventional radiative calibration methods include absolute calibration based on radiative transfer models and relative calibration using radiometric normalization. In remote sensing CD, radiometric normalization is a commonly used method.	[42], [43], [44], [45]
Image Super-resolution	For remote sensing images with inconsistent spatial resolution, image super-resolution can restore the remote sensing image with a lower spatial resolution to a higher spatial resolution.	The conventional super-resolution methods include interpolation, sub-pixel methods, etc. In recent years, deep learning methods, including generative adversarial networks and deep convolutional neural networks, have become increasingly popular.	[35], [46], [47], [48]
Image Denoising	Image denoising aims to remove the influence of noise on remote sensing images, such as noise in optical remote sensing images or speckle noise in SAR images. This process is crucial.	The conventional image denoising methods include spatiotemporal speckle filtering, Lee sigma filter, nonlinear diffusion filtering, etc.	[30], [31], [32], [49]
Data Enhancement	Data enhancement aims to increase the quantity and diversity of samples, which is essential.	Conventional data enhancement methods include flipping, color transformation, cropping, rotation, and other operations. In recent years, image generation-based networks, including GAN networks and instance augmentation networks, have gained popularity.	[32], [50]

CD. In CDCD, in addition to real changes in ground objects, other pseudochanged factors should be eliminated as much as possible, such as the unchanged area being misidentified as changed area, and the changed area being misidentified as unchanged area. Typical image preprocessing approaches include, but are not limited to, image registration, radiative calibration, image super-resolution, and image denoising. Moreover, it is often expensive and time-consuming to obtain images

with high-quality labels. Many supervised or semisupervised CDCD methods also utilize data enhancement as one of the image preprocessing operations. The definitions and the detailed techniques of each kind of images preprocessing method are summarized in Table I.

Image registration is the process of aligning the geographical coordinates and geometric positions of images using matching relationships of point and line features in multitemporal remote

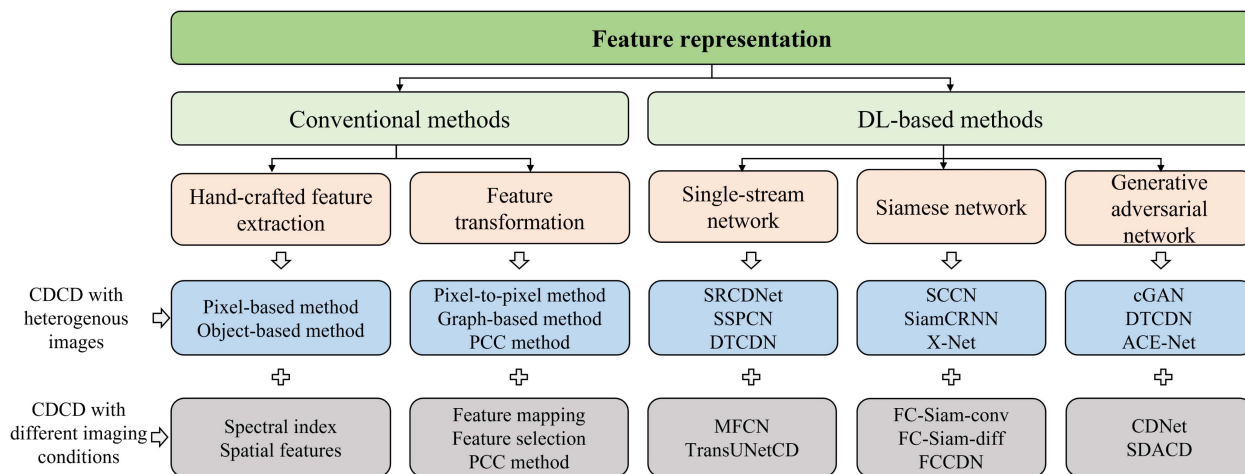


Fig. 4. Structure of the feature representation section, which includes conventional methods and DL-based methods for heterogenous images and images collected with different imaging conditions.

sensing images [25], [26]. Radiative calibration eliminates radiative differences in images captured under different imaging conditions (e.g., solar angles, sensor characteristics, and atmospheric conditions) using a series of methods such as pseudo-invariant features and histogram matching [27], [28]. Image super-resolution aims to recover high-resolution images from low-resolution images and is commonly used in CD using images with different spatial resolutions. This technique includes interpolation, DL-based methods, etc. Image denoising uses various denoising algorithms, including total variation denoising algorithm [29], Lee Sigma filter [30], and Rudin–Osher–Fatemi (ROF) denoising method [31], to suppress multiple types and levels of noise, such as speckle noise, Gaussian noise, impulse noise, and periodic noise, to obtain images with higher quality. Data enhancement techniques aim to increase the amount and variety of data [32], prevent model overfitting, and include geometric transformations, color transformations, and DL-based methods [33]. It is also notable that recent research works also take the preprocessing process as one part of the CD methods so that individual preprocessing process is not necessary in such special cases [34], [35], [36].

B. Feature Representation

It is necessary to represent consistent features between remote sensing images captured at different times for CDCD. Feature representation acts as the core of CDCD. In this section, we provide a comprehensive overview of the two categories involved in feature representation, which are conventional methods and DL-based methods. Conventional methods include hand-crafted feature design and feature transformation, which rely on expert-designed techniques and prior knowledge. DL-based methods primarily utilize the capabilities of DL to bridge the gap between cross-domain images and learn discriminative features using various network architectures, i.e., single-stream network, Siamese network and generative adversarial network (GAN). Fig. 4 illustrates the structure of the feature representation

section, and Fig. 5 summarizes the available literature for feature representation.

1) Conventional Methods:

(1) *Hand-Crafted Feature Design*: Hand-crafted feature extraction utilizes image processing techniques to extract representative features from remote sensing images. These techniques vary with the cross-domain categories, i.e., whether the input images are collected by different sensors or under different imaging conditions. For clarity, this section introduces these techniques separately.

(a) *Extraction from Heterogeneous Images*: Due to the differences in imaging sensors, spatial resolutions, and spectral bands among heterogeneous images, it is common to extract features either from images with different imaging modes or resolutions, or from different spectral bands of remote sensing images, to effectively detect changes. In this case, conventional pixel-level feature extraction methods, which rely on statistical counting and similarity measurements, are widely adopted. Stow et al. [51] found that performing ratio calculations on multisensor, multitemporal satellite image bands can improve the accuracy of CDCD. Prendes et al. [52] took into account the physical characteristics of the sensors, particularly the related measurement noise model and local joint distribution, and proposed a manifold estimation algorithm to measure the similarity of heterogeneous images. These methods based on pixel-level feature extraction are simple, intuitive, and easy to implement. However, these methods are vulnerable to adverse factors, including image noise and resolution, and they frequently fail to consider higher-level semantic information of land surface.

With the widespread use of high spatial resolution images, the spectral variability within individual images and the multiscale nature of ground objects limits the performance of traditional pixel-level feature extraction methods [53]. Object-based feature extraction using image segmentation algorithms can relieve this limitation because these methods can help to convert pixel-level features into object-level features, allowing for the extraction of richer semantic information, such as textures,

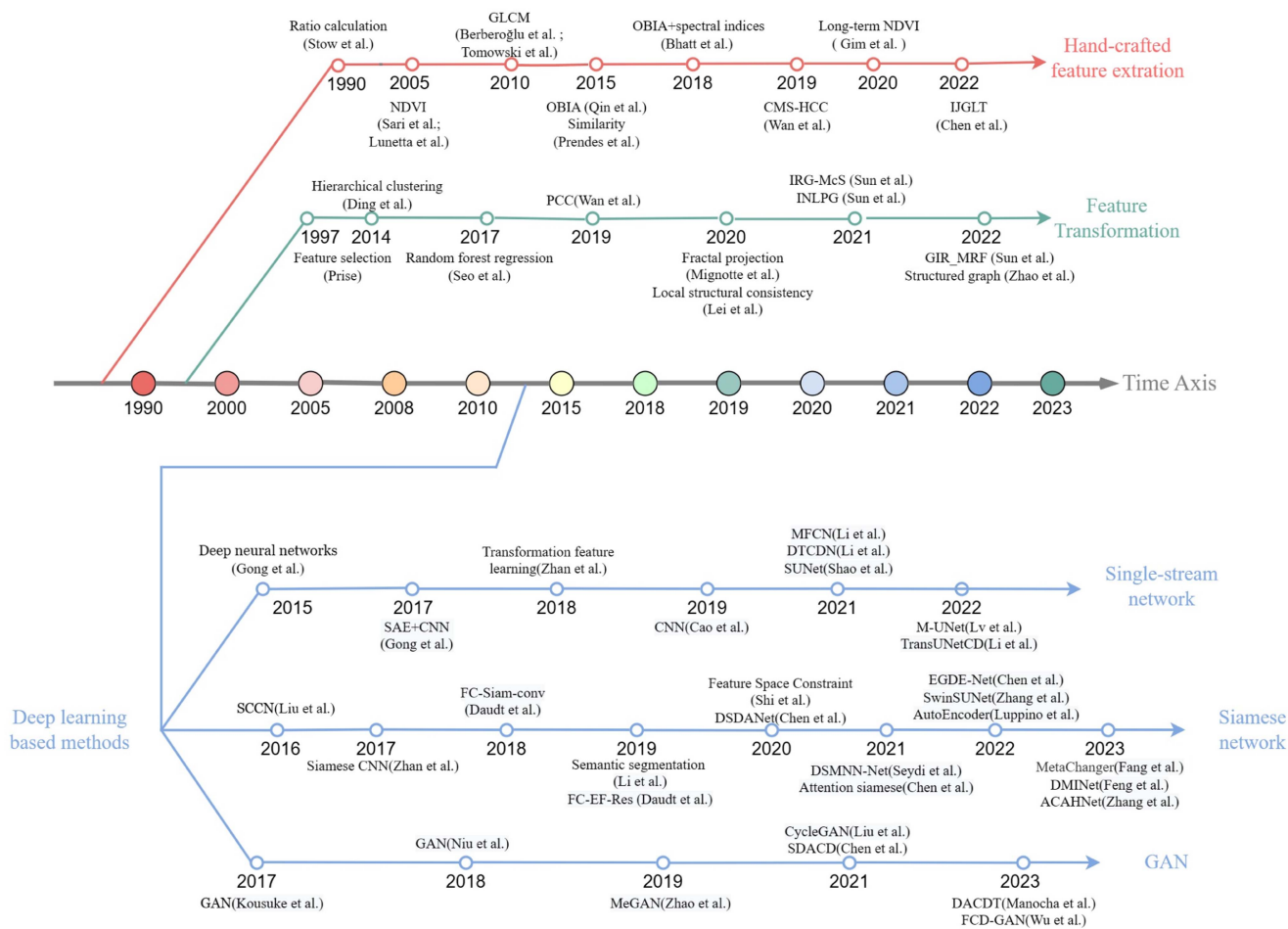


Fig. 5. Summary of available literature for CDCD methods.

shapes, spatial, and spatial context [53], [54]. For example, Wan et al. [55] used multirate segmentation to generate spectrally, spatially, and temporally homogeneous image objects and then performed CD on optical and SAR images based on the image objects. Bhatt et al. [56] utilized spectral indices (SIs) to characterize urban built-up areas and combined them with image segmentation techniques to reduce spectral variability and incorporate spatial context information. This approach has been proved to be effective in handling cross-domain image with spatial and spectral heterogeneity. Moreover, Wan et al. [57] took into account radiative and geometric differences between heterogeneous images and proposed a collaborative multitemporal segmentation method, termed as CMS-HCC, for CDCD in SAR and optical images. Chen et al. [58] segment two heterogeneous images into superpixels with the same boundaries and obtained a change map (CM) that represents the areas of change by iterative joint global-local translation (IJGLT). Compared to pixel-based methods, object-based methods can better capture the spatial information of objects, thereby improving feature representation and CDCD accuracy. However, the accuracy of object-based feature extraction can be influenced by the predefined segmentation parameters.

(b) *Extraction From Images With Different Imaging Conditions:* Different land covers have different spectral responses

under different imaging conditions, making it difficult to accurately identify changes solely based on spectral features. For example, variations in solar angles and season, and differences within ground objects can result in false changes [59]. In addition, the aging of buildings can cause spectral changes, leading to erroneous detections [60]. Therefore, selecting appropriate feature extraction methods is essential to enhance the accuracy and reliability of CDCD with different imaging conditions.

SIs can be helpful to some extent in distinguishing phenological changes from nonphenological changes. SIs are derived from algebraic calculations of spectral reflectance values from different bands of remote sensing images. They can characterize the spectral responses of ground objects under different seasons. Sari et al. [61] conducted CDCD in South Kalimantan province using an unsupervised classification method called RGB-normalized difference vegetation index (NDVI) on multitemporal SPOT images. Lunetta et al. [62] utilized discrete Fourier transform to estimate missing MODIS NDVI data, providing high-quality continuous data supporting CDCD in multitemporal sequences. Liu et al. [63] detected changes based on multiple morphological parameters by combining spectral correlation operator with morphological features of the NDVI time curve. Gim et al. [64] utilized AVHRR-estimated NDVI to detect the growing season of agricultural fields. These methods

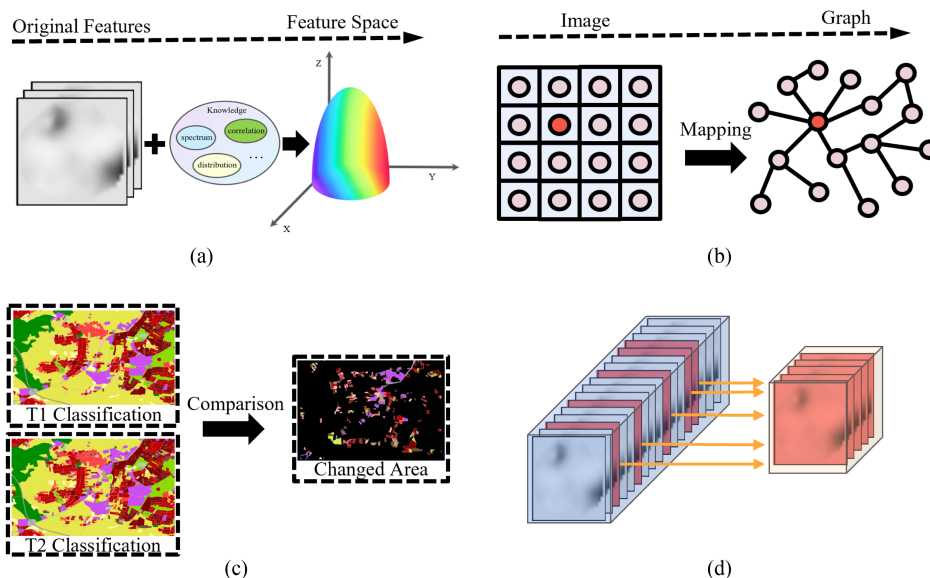


Fig. 6. Widely adopted feature transformation methods. (a) Prior-knowledge based method. (b) Graph-based method. (c) PCC. (d) Feature selection.

have high applicability and generality, making them suitable for feature representation in large-scale areas. However, in general, these methods do not take spatial information into account.

Spatial information plays a crucial role in compensating of feature representation [65]. Spatial features tend to remain relatively stable compared to noise and illumination changes [66]. From this perspective, spatial features contribute to represent cross-domain image features under different imaging conditions. Yang et al. [67] utilized the homogeneous texture descriptor, which is based on the standardized MPEG-7 Gabor filters, to extract texture feature vectors for CDCD. Berberoğlu et al. [68] extracted spatial features of ground objects based on co-occurrence matrices and variation functions. Tomowski et al. [69] proposed a CDCD method based on color and gray-level co-occurrence matrix (GLCM) features, capable of detecting different types of changes. Li et al. [70] defined a local-similarity-based texture difference measure using GLCM. Spatial features can complement features when spectral features are not significant. However, spatial features do not possess rotation and scale invariance. Calculations usually need to consider multiple directions and scales, which can be computationally complex, especially for CDCD with high-resolution remote sensing images.

(2) *Feature Transformation*: Feature transformation not only automatically extract image features from remote sensing images, but also employ nonlinear mappings to project the original features into a feature space, in which better representation of the differences and changes can be achieved. Literately, the most widely adopted methods for feature transformation are shown in Fig. 6.

(a) *Transformation for Heterogeneous Images*: Heterogeneous image features can be transformed and represented based on prior knowledge or experience, enabling the application of these methods in CDCD. A variety of feature transformation methods have been proposed. For example, Liu et al. [71]

introduced a CDCD method (SAR and optical images) based on homogeneous pixel transformation. It transforms images from their original feature space (e.g., grayscale space) to a pixel-level mapping space (e.g., spectral space) so that the features before and after the change can be represented in a common embedding space. Touati et al. [72] [73] designed a framework for CDCD based on imaging modalities invariant operators. These operators detect the differences in high-frequency patterns of each structural region existing in two heterogeneous satellite images at different scales. Touati et al. [74] employed a circular invariant model and proposed a CD method, termed as CICMM, so that the prechange image can be transformed to the domain of the postchange image. Mignotte projected heterogeneous images into a feature embedding space using modified geometric fractal decomposition and shrinking mapping methods for CDCD [75]. These feature transformation methods mentioned above exhibit repeatability, stability, and fast computation speed, making them suitable for processing large-scale heterogeneous images. However, these methods primarily focus on the pixel-to-pixel transformation techniques, without directly considering the structural information in images. In many practical applications, researchers often wish to utilize the structural information in order to obtain more discriminative and representative features.

To leverage the structural information in heterogeneous images, graph based feature transformation can be used. In [76], a robust nearest neighbor graph was constructed to represent the structure of heterogeneous images, and the graphs within the same image domain through graph mapping were compared to calculate the forward and backward difference images for CDCD. In order to address the fact that graph-based methods often overlook prior knowledge about scene structures, Jimenez-Sierra et al. [77] proposed a CDCD framework that combines graph fusion and is driven by graph signal smoothness representation for heterogeneous images. Sun et al. [78] proposed an unsupervised image regression method

for CDCD based on the inherent structural consistency between heterogeneous images. This method utilized the self-expression property to preserve the global structure of features and employed an adaptive neighborhood approach to capture the local structure of features. Lei et al. [79] proposed a CDCD method based on adaptive local structure consistency. It is capable of constructing the local structure for each image and then measuring the consistency between the local structures of two images. Zhao et al. [80] utilized an auto-weighted structured graph to achieve domain transformation of the images. Specifically, it utilizes high-order neighbor information hidden in the graph. Graph based feature transformation methods focus on the structural features rather than pixel-level features [81], [82], [83]. They can capture both local and global structural information and exhibit strong robustness.

Postclassification comparison (PCC) is a special type of feature transformation method that identifies land cover change information by directly comparing the land cover maps of multitemporal images. PCC does not require strict prerequisites of radiometric consistency between images. Therefore, it is suitable for CDCD with images acquired from different imaging sensors or conditions [60]. For example, Mubea et al. [84] utilized the PCC method to detect changes in Nakuru, Kenya by combining optical and SAR imagery. Han et al. [85] proposed an improved hierarchical extreme learning machine method for CD after heterogeneous image classification. However, it also faces the issue of cumulative classification errors. Composite classification can be helpful to mitigate the cumulative errors by adopting time dependence. Wan et al. [86] proposed an optical and SAR image CDCD method based on multitemporal segmentation and composite classification. In addition, the adoption of multitemporal segmentation reduces the inevitable salt-and-pepper effect in pixel-based methods and mitigates misdetections caused by region transitions and object misalignment in object-based methods.

(b) *Transformation for Images With Different Imaging Conditions*: Feature mapping is the method that represents the features of images captured under different imaging conditions using mapping functions. This approach involves constructing a mapping model to describe the linear or nonlinear feature relationships between images. For example, Ding et al. [87] proposed a hierarchical clustering method with sparse representation suitable for CDCD. This method utilizes a tree-structured dictionary to represent features and uses sparse reconstruction error to determine the changes. Seo et al. [88] introduced a regression method normalizing the radiometric and phenological conditions of the original features. Feature mapping methods take into account the semantic information of images by representing features in the feature space. These approaches help to preserve the discriminative characteristics of the images, thus aiding in maintaining semantic consistency. However, feature mapping methods may lead to the loss or deformation of certain features and change the physical characteristics of the image.

Mapping images to a low-dimensional space using feature selection is a special type of feature transformation method. Feature selection refers to the process of selecting discriminative

features from existing features while preserving their physical quantity [89]. Popular feature selection methods often measure distances and similarities among multidimensional features to select the optimal subset of features. Price et al. [90] partitioned the bands of hyperspectral images into different spectral intervals and performed band selection to minimize the overall variability of the image. Bajcsy et al. [91] compared a wide range of feature selection methods and found that first-spectral derivatives and uniform spectral spacing can achieve effective feature selection. Ifarraguerri et al. [92] used multivariate Jeffries–Matusita distance measures to select representative features during the search optimization process. Chen et al. [93] proposed a feature selection algorithm based on genetic particle swarm optimization to address the optimization problem of feature selection in object-based CDCD. Cai et al. [94] aimed to address issues such as the inability to detect subtle changes, segmentation effects, and high false detection rates in object-based CDCD methods. They extracted spectral and texture features from different time periods and then used the maximum relevance minimum redundancy algorithm for feature selection [94]. Feature selection helps eliminate task-irrelevant redundant features, reduces noise, and interference caused by different imaging conditions, and improves the generalization of features. However, the effectiveness of feature selection is influenced by the choice of the feature selection algorithm. Inappropriate feature selection algorithms may result in feature subsets that are not representative or lack discriminative ability.

2) *DL-Based Methods*: The increasing availability of cross-domain images has provided a foundation for DL-based feature representation methods [95], [96]. With their exceptional feature abstraction and autonomous learning capabilities, DL can learn discriminative features from cross-domain images [97], [98], [99]. The DL-based methods leveraging learned features and further improve the model generalization. Notably, machine learning methods can also be utilized for feature representation, as studied in [12], [100], [101], [102]. However, they are generally believed to be less powerful than DL-based methods. We thus put more attention to the DL-based methods. According to the literature, the most widely used DL-based models are single-stream network, siamese network, and GANs, as shown in Fig. 7.

(1) *Single-Stream Network*: The single-stream deep feature learning method takes the original images of the multitemporal phase as inputs or performs subtraction or concatenation on the features obtained through simple feature extraction. Regarding CDCD with heterogeneous images or different imaging conditions, the strategies involved in feature representation using single-stream network are different.

(a) *Single-Stream Network for Heterogeneous Images*: As for CDCD with heterogeneous images, Zhan et al. [103] applied a logarithmic transformation to SAR images to obtain statistical distribution properties similar to optical images. They learned high-dimensional feature representations from the transformed feature pairs. Li et al. [104] generated pseudolabels using spatial fuzzy clustering and trained a semantic segmentation network based on the generated pseudolabels. Ma et al. [105] transformed two heterogeneous images in the pixel-level feature space and

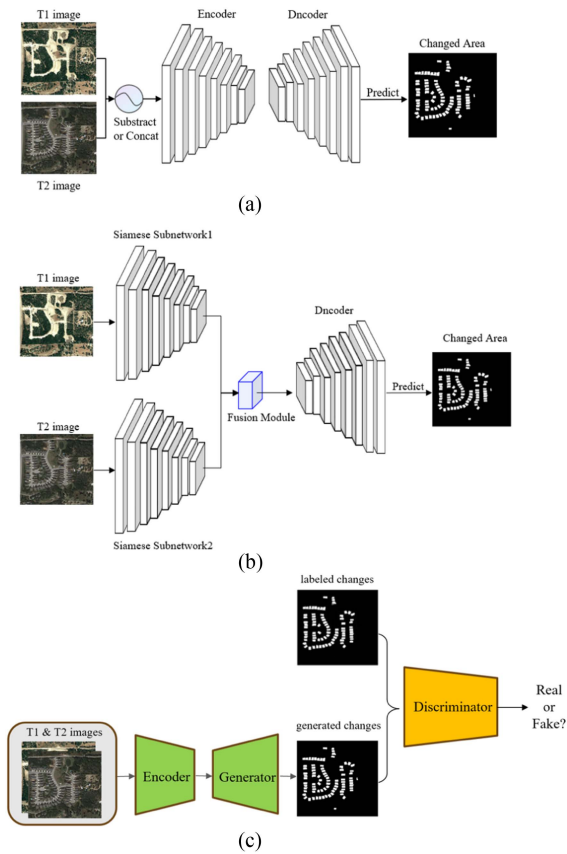


Fig. 7. Widely adopted DL based models in CDCD. (a) Single-stream network. (b) Siamese network. (c) GAN.

compared and classified them in their respective feature spaces. They used the postclassification pseudolabels to train a capsule neural network for CDCD. Shao et al. [106] used a neural network to map heterogeneous satellite and unmanned aerial vehicle images to a high-dimensional feature space for CD, mitigating the differences between heterogeneous images, such as color, resolution, disparity, and image distortion. Liu et al. [35] proposed a super-resolution-based change detection network with stacked attention modules, which can reduce intraclass heterogeneity and inter-class similarity in images with different spatial resolutions. Li et al. [107] constructed a spatial self-paced convolutional network for unsupervised CDCD. It integrated self-paced learning into the convolutional network to dynamically select reliable samples for learning the representation of the relationship between heterogeneous images. Li et al. [108] proposed a deep translation-based change detection network (DTCDN) for optical and SAR images. The deep translation mapped images from one domain (e.g., optical) to another domain (e.g., SAR) using a cyclic structure, enabling feature representation in the same feature space. Lv et al. [109] connected a pair of heterogeneous image patches and used the UNet neural network to learn shared abstract features for CDCD. They embedded multiscale convolution modules in the UNet to cover ground objects of various sizes and shapes in the image scene.

(b) *Single-Stream Network for Images With Different Imaging Conditions:* Regarding CDCD with different imaging conditions, Gong et al. [110] proposed a DL-based CDCD algorithm for multitemporal SAR image analysis. They first performed joint classification on two original images and then built a single-stream deep neural network. This method reduces the influence of speckle noise. Gong et al. [111] used sparse autoencoders to transform logarithmic ratio difference features into an appropriate feature space, enabling CDCD of SAR images at different imaging conditions. Jaturapitpornchai et al. [112] used the UNet architecture to identify the locations of newly built structures in SAR images. This method not only learns changes in structures and other regions based on intensity variations, but also incorporates visual features of structures and nonbuilding objects, distinguishing changes related to newly built structures from other types of changes. Cao et al. [113] combined a deep denoising model and CNN to construct a CDCD algorithm for SAR images. They used deep denoising to preserve useful features and suppress noise in SAR images, and then input the denoised difference map into the CNN for feature learning to obtain the CM. Li et al. [114] proposed a multiscale fully convolutional neural network (MFCN) to relief the influence of unbalanced samples so that deep features could be effectively extracted. Li et al. [115] introduced an end-to-end CDCD network named TransUNetCD, which amalgamates the merits of both the transformer and UNet architectures. This innovative network adeptly represents local-global semantic attributes of images acquired under different imaging conditions, consequently mitigating the influence of spurious variations.

The network structure of single-stream deep feature learning is relatively simple, as it contains only one primary data flow, thereby reducing the complexity of the model and the difficulty of training. However, in the single-stream structure, the network typically performs simple subtraction or concatenation operations on multitemporal images at the feature input stage, without fully utilizing the interrelatedness and complementarity of the image data. As a result, it may not fully exploit the discriminative features required for CDCD.

(2) *Siamese Network:* A Siamese network consists of two identical or similar subnetworks. Each image is fed into one branch of the Siamese network's subnetwork for feature extraction. The features from both branches are then fused using a feature fusion module for feature representation, and the features are decoded to generate the CDCD results.

(a) *Siamese Network for Heterogeneous Images:* Regarding CDCD with heterogeneous images, Zhao et al. [116] constructed a neural network with coupling layers on both sides, containing an equal number of layers, to transform heterogeneous images into the same feature space. Liu et al. [117] proposed a symmetric convolutional coupling network (SCCN) for heterogeneous images. It transforms the input images into a feature space and applies a thresholding algorithm to learn deep features. Chen et al. [10] presented a universal deep Siamese convolutional multiple-layer recurrent neural network (SiamCRNN) applicable to heterogeneous images. Shi et al. [118] proposed an end-to-end CDCD method based on feature space constraint. It adopts a Siamese encoder to extract image features and then

computes the gram matrix of the feature correlations to project the image representation into different feature spaces. Seydi et al. [119] introduced a deep Siamese morphological neural network (DSMNN-Net), combining multiscale convolutional layers and morphological layers to generate deep features for CDCD. Chen et al. [120] proposed an attention-guided Siamese fusion network for CDCD, where feature fusion is achieved by using an attention information fusion module. Wu et al. [121] presented an unsupervised CDCD method that includes a convolutional autoencoder (CAE) for feature extraction. The CAE eliminates most redundancies in heterogeneous images, and the commonality autoencoder aligns the feature representations of the transformed images. Luppino et al. [122] utilized two fully convolutional networks to map the heterogeneous images from one domain to another and proposed the X-Net architecture. Jiang et al. [123] proposed a semi-supervised Siamese network (S3N) based on transfer learning. In S3N, the layer weights of deep-level features are trained using the transfer learning strategy, reducing the computational cost. Moreover, some researchers have used domain-specific affinity matrices and enforced the feature representations of heterogeneous images in the embedding space using these relationships [122], [124].

(b) *Siamese Network for Images With Different Imaging Conditions*: As for CDCD with different imaging conditions, Zhan et al. [125] addressed the problem by first training a Siamese convolutional network using a weighted contrastive loss (CL). This network directly extracts features from image pairs, and the distance between the feature vectors is used to detect changes between image pairs with significant temporal differences. Daudt et al. [126] [127] proposed three widely used end-to-end CD frameworks, termed as FC-Siam-conc, FC-Siam-diff, and FC-EF-Res, by adopting fully convolution network without pretraining or transfer learning. Liu et al. [128] proposed a novel network architecture that combines CNNs with bidirectional long short-term memory (BiLSTM) networks to extract joint spectral-spatial-temporal features from dual-temporal multispectral images. Chen et al. [129] proposed a deep Siamese domain adaptation convolutional neural network (DSDANet) architecture for CDCD. First, the model extracts spatiotemporal spectral features from multitemporal images. Then, the learned features are embedded into a reproducing kernel Hilbert space to achieve explicit distribution matching between the two domains. Xu et al. [130] proposed a multidirectional fusion perception network that combines different types of features through multidirectional fusion paths (MFP) and an adaptive weighted fusion (AWF) strategy. Zhang et al. [131] proposed a Siamese CD network called SMD-Net, which is based on the incremental enhancement of change region information using multiscale difference maps. Sun et al. [132] constructed a Siamese nested UNet (SANet) with a graph attention mechanism. By training the model on the distorted images, this method effectively improved the CDCD accuracy with different imaging conditions. Chen et al. [133] designed a feature constraint CD network (FCCDN) to extract and fuse multiscale features. They also proposed a self-supervised learning-based strategy to constrain feature learning and reduce the impact of varying imaging conditions. Chen et al. [134] designed a feature

difference enhancement module to learn more discriminative features, aiming to reduce intraclass variance and mitigate the impact of imaging conditions. Hu et al. [135] proposed a novel self-supervised hyperspectral spatiotemporal spectral learning network called HyperNet, which aims to enhance the robustness of building CD in diverse imaging conditions. Zhang et al. [136] introduced a Siamese U-shaped architecture within the framework of the transformer network, termed as SwinSUNet, capable of concurrently processing images with different imaging conditions while effectively represent dual-temporal features. Fang et al. [137] proposed a generalized CD architecture, MetaChanger, which uses a flow-based double alignment fusion (FDAF) module to interactively fuse the two-branch features of the twin network to overcome the problem of domain differences in different temporal images. Feng et al. [138] proposed a dual-branch multilevel intertemporal network (DMINet) to efficiently derive change representations and suppress change-independent interference by unifying self-attention and cross-attention to obtain inter-period joint-attention. Zhang et al. [139] proposed the asymmetric cross-attention hierarchical network (ACAHNet), which combines CNN and transformer in a series-parallel manner to eliminate pseudovariations caused by different imaging conditions based on their complementary advantages.

The Siamese network allows for separate feature extraction from cross-domain images, and the features from both branches are compared and fused in the feature fusion module. This enables Siamese network to effectively integrate information from multiple data sources and leverage the characteristics of heterogeneous or differently imaged conditions to enhance the overall performance of CD. However, the Siamese network also requires learning additional parameters to explore the relationship between the subnetworks, making the training and optimization process more complex compared to the single-stream neural networks.

(3) *Generative Adversarial Network (GAN)*: Currently, most DL-based CDCD methods attempt to design complex neural networks with powerful feature learning capabilities, overlooking the common occurrence of domain shift phenomena, which results in limited generalization of the models. To address this problem, researchers have been exploring the development of more generalized feature representation models. Among them, generative cross-domain feature learning has attracted extensive attention. This method focuses on representing the features of different domains, and the knowledge learned depends on the two input domains.

(a) *GAN for Heterogeneous Images*: As for CDCD with heterogeneous images, Niu et al. [140] proposed an unsupervised translation network based on conditional GAN (cGAN) to convert optical images into SAR images in a shared embedding space. Gong et al. [141] constructed a coupled GAN associated with the coupled variational autoencoder to perform isomorphic transformations on the heterogeneous images. Li et al. [108] utilized a GAN to construct a deep translation network that achieves domain representation of features extracted from heterogeneous image and proposed a DTCDN for CDCD. Luppino et al. [122] utilized two autoencoders and proposed the ACE-Net framework

by aligning code spaces with adversarial training. Jia et al. [142] proposed novel binary adversarial autoencoders with structural self-similarity (BASNet) for detecting land cover changes in heterogeneous images. The structural consistency loss is defined by the cross-pattern distance in a novel affinity space, which enforces the network to transform the heterogeneous images into a common domain for feature representation. Based on CycleGAN, Liu et al. [143] proposed an unsupervised CD method to learn the mapping relationship between heterogeneous images. Specifically, it transforms an image (e.g., SAR) from its original feature space to another feature space (e.g., optical). Manocha et al. [144] proposed an image translation process-oriented Deep Adaptation-based Change Detection Technique (DACDT), in which GAN is used to encode and transform optical and SAR images to the same domain.

(b) *GAN for Images With Different Imaging Conditions*: As for CDCD with different imaging conditions, Kuo et al. [145] proposed a framework for handling CDCD at the semantic level. This framework connects the source and target domains through semantic consistency. Kousuke et al. [146] utilized GANs to address domain shift issues. Specifically, they employed a GAN-based transformer to map the features from the target domain to the source domain, treating the CDCD as an intradomain detection task. Jian et al. [147] developed a GANs-based one-class classification technique for CDCD. The method utilizes unaltered image sequences to train a GAN that generates changed data. Yang et al. [148] proposed a selective adversarial adaptation-based framework, which utilizes two domain discriminators to establish domain adaptation relationships between multiple source domains and a target domain. Chen et al. [50] utilized generative adversarial training to generate dual-temporal images that reflect various types of changes in buildings. They then employed CDNet to extract the changed regions in the buildings. Liu et al. [149] proposed a supervised domain adaptation framework method, which utilizes generative adversarial learning with cycle-consistency constraints to achieve cross-domain style transfer and learns domain-invariant features to represent different feature distributions in the feature space. Wu et al. [6] proposed a framework called fully convolutional change detection with generative adversarial networks (FCD-GAN), which unifies unsupervised, weakly supervised, semi-supervised, and fully supervised CD and models spectral and spatial variations between multitemporal images.

GANs achieve feature representation through cross-modal learning, enabling the learning of feature mapping relationships between different domains. By optimizing the balance between the generator and discriminator, GANs can obtain more robust feature representations. It means that GANs can handle images from different sensors and imaging conditions, thereby better adapting to cross-domain scenes. However, the training process of GANs is complex and prone to issues like mode collapse and mode oscillation, which can lead to training instability or convergence difficulties. In addition, generated images may suffer from missing details, blurriness, or lack of realism, which can impact the accuracy and reliability of feature representation.

C. Change Detectors

Change detectors refer to a collective term for mathematical methods and artificial models that transform acquired features and various prior information into CDCD results. As the final stage of the CDCD process, the efficiency of change detectors significantly impacts the result of CDCD. Based on whether the detection results are learned in conjunction with features, change detectors can be classified into two categories: feature similarity-based and joint feature-loss learning methods. Details of these change detectors are shown in Table II.

1) *Feature Similarity-Based Change Detectors*: Feature similarity-based change detectors rely on feature engineering and detect the change by measuring the degree of similarity between image features in two periods. The main quantitative metrics for feature similarity are Euclidean parity [150], cosine similarity [152], and maximum mean discrepancy (MMD) [154]. In addition to generating distance maps by performing pixel-by-pixel distance calculations on two-phase images [155], [157], features selected in layers can be concatenated to construct depth-varying super-vectors [158] with specific components having higher absolute values on the changed pixels than the unchanged ones. For MMD, single-core MMD features have limited capability to represent and studies often use multiple-kernel MMD (MK-MMD) [153], [154]. The MK-MMD-based change detectors can use training data containing only sparse annotations to accomplish knowledge migration between different domains. In addition, initial CM optimization strategies [140] are also widely used in CDCD, and this class of methods adjusts the detection results by performing algebraic operations [148] on the patches marked as changed versus unchanged in the initial CM. For example, Celik et al. [159] and Pual et al. [151] obtained the image difference feature maps by simple algebraic computation and depth feature maps, respectively, and then processed them by conventional PCA and clustering to obtain the CMs. Sun et al. [76] achieved CD with Markov Random Fields (MRFs) using local similarity as a constraint and change energy as a basis in heterogeneous images. In addition, Sun et al. [16] fused superpixel-based feature extraction with change vectors and achieved multimodal CDCD by segmentation optimization of conditional random fields. Considering the advantages of graph structure in establishing heterogeneous image associations, Sun et al. [81] constructed similarity maps for images to generate difference maps, and then achieved CD by threshold segmentation and superpixel-based MRF. Niu et al. [140] used cGAN to represent SAR images and optical remote sensing images in an approximately consistent form and then achieved CDCD by direct comparison for difference. Mignotte used spatial self-similarity at different scales of heterogeneous images for approximate encoding and MRF in iterative conditional mode for CDCD of the final difference map [75]. Liu et al. [149] used convolution and coupling layers to extract domain invariant features from heterogeneous images and then used threshold segmentation based on pixel distance to obtain the CM. Touati et al. [74] converted the multimodal CD problem to a unimodal CD problem using concentric invariant convolution model mapping to project pre-event images

TABLE II
SUMMARY OF AVAILABLE CHANGE DETECTORS

Change detectors	Methods	Formula	Explanation	References
	Euclidean distance	$d(X^1, X^2) = \sqrt{\sum_{i=1}^N (x_i^1 - x_i^2)^2}$	$d(X^1, X^2)$ is the distance between features; N is the feature dimension; x_i^1 and x_i^2 are eigenvalues for each dimension.	[150] [151]
Feature similarity	Cosine similarity	$d(X^1, X^2) = \frac{\sum_{i=1}^N x_i^1 x_i^2}{\sqrt{\sum_{i=1}^N (x_i^1)^2} \cdot \sqrt{\sum_{i=1}^N (x_i^2)^2}}$	$d(X^1, X^2)$ is the distance between features; N is the feature dimension; x_i^1 and x_i^2 are eigenvalues for each dimension.	[152]
	Maximum mean discrepancy	$d(X^1, X^2) = \left\ \frac{1}{N} \sum_{i=1}^N (f(x_i^1) - f(x_i^2)) \right\ _H$	$d(X^1, X^2)$ is the distance between features; N is the feature dimension; x_i^1 and x_i^2 are the respective dimensional eigenvalues; $f(\cdot)$ denotes a mapping from the original space to the regenerated Hilbert space; $\ \cdot\ _H$ denotes regenerative Hilbert space normalization.	[153] [154]
	Cross-entropy loss	$Loss = mean(-(1 - L_v) \log(y_d) - L_v \cdot \log(1 - y_d))$	$Loss$ denotes loss values; $mean(\cdot)$ denotes averaging; and L_v is the label value, which is taken as 0 when the pixel prediction result is consistent with the true label, and 1 when it is not; y_d is the predicted probability of the pixel category obtained by feature similarity determination.	[108]
Joint feature-loss learning	Contrastive loss	$Loss = \frac{1}{2} mean((1 - L_v) d^2 + L_v \cdot max(0, m - d)^2)$	$Loss$ denotes loss values; the $mean(\cdot)$ and $max(\cdot)$ denote the averaging and maximizing operations, respectively; d is the inter-pixel distance and m is the distance threshold, if $d > m$ then the pixel pair is non-similar; L_v is the label value, which is 0 when the pixel pair is judged to be similar and 1 if not.	[155] [156]
	Focal loss	$Loss = mean(-\alpha(1 - L_v)(1 - y_d)^\gamma \log(y_d) - L_v(1 - \alpha)y_d^\gamma \cdot \log(1 - y_d))$	$Loss$ denotes loss values; $mean(\cdot)$ denotes averaging; and L_v is the label value, which is taken as 0 when the pixel prediction result is consistent with the true label, and 1 otherwise; y_d is the predicted probability of the pixel category obtained by feature similarity determination; parameters α and γ are used to adjust the weights of positive and negative samples and hard cases, respectively.	[108] [156]

into postevent images for imaging and then used unsupervised Markov models to generate binary CM. The method based on feature similarity is highly interpretable but is difficult to handle weighted optimization and pays insufficient attention to important feature information. Usually, it can only be improved by methods such as secondary judgment of results by manually setting thresholds and lacks the ability to cope with CD problems in complex scenes.

2) *Joint Feature-Loss Learning Based Change Detectors:* The feature-loss joint learning based approach uses feature similarity measures to obtain the independent variables of the loss function in the feature learning process, and then constructs the loss function by combining the change information of the ground truth, and jointly learns the final change with the criterion of minimizing the value of the loss function. For example, Chen et al. [154] and Wang et al. [159] used the image pair metric distance and frequency of occurrence as the basis for parameter settings, respectively, and used CL to penalize the pixel pairs whose sample pair similarity and variation were inconsistent. Li et al. [108] used deep image translation based on the cyclic structure to obtain similar features of heterogeneous images as the basis for constructing the loss function and then combined the information of each scale by focal loss (FL) to generate the

CM. Luppino et al. [122] used the L2 distance measure between image blocks to participate in the reweighting of change pixel pairs by image translation loss, and then generated CM with random field iterative optimization filtering and Otsu thresholding method. Daudt et al. [127] combined land cover category prediction tasks, prompting CD losses to take into account land cover information for CM optimization. Li et al. [114] optimized the loss function to reweight the binary cross-entropy loss using the proportion of background (negative samples) in the image to measure the degree of sample imbalance. However, due to the distribution of change objects and the interference of background objects in the real world, it is difficult for a single loss function to address issues such as sample imbalance and false changes. Typically, a combination of different types of loss functions is required to improve the robustness of change detectors in complex backgrounds. By merging CL and FL, Wang et al. [160] made the joint loss function more focused on low-confidence nonbackground samples, while ensuring that the positive and negative instance distance measures are reasonable. Cao et al. [161] started from the inverse volume of the regional area and performed weighted processing on the Generalized Dice Loss (GDL). They then combined this with class frequency weighted cross-entropy loss to improve change detector performance in

TABLE III
DETAILED INFORMATION ON THE SELECTED CD METHODS

No.	Method	Type	Publication	Source
1	NPSG [16]	Conventional methods	PR, 2021	https://github.com/yulisun/NPSG
2	IRG-McS [76]		TIP, 2021	https://github.com/yulisun/IRG-McS
3	INLPG [81]		TGRS, 2022	https://github.com/yulisun/INLPG
4	GIR_MRF [78]		ISPRS, 2022	https://github.com/yulisun/GIR-MRF
5	ISTCRF [82]		PR, 2022	https://github.com/yulisun/IST-CRF
6	SCASC [83]		TGRS, 2022	https://github.com/yulisun/SCASC
7	CICMM [74]		ASTESJ, 2020	https://www-labs.iro.umontreal.ca/~mignotte/ResearchMaterial/index.html
8	FPMABA [75]		TGRS, 2020	https://www-labs.iro.umontreal.ca/~mignotte/ResearchMaterial/index.html
9	cGAN [140]	DL-based methods	GRSL, 2019	https://github.com/llu025/Heterogeneous_CD/tree/master/Code-Aligned_Autoencoders
10	SCCN [117]		TNNLS, 2018	https://github.com/llu025/Heterogeneous_CD/tree/master/Code-Aligned_Autoencoders
11	X-Net [122]		TGRS, 2022	https://github.com/llu025/Heterogeneous_CD/tree/bbae43f8ac5a08d6f79a5253377ebe866e8e2634/legacy/Deep_Image_Translation
12	ACE-Net [122]		TGRS, 2022	https://github.com/llu025/Heterogeneous_CD/tree/bbae43f8ac5a08d6f79a5253377ebe866e8e2634/legacy/Deep_Image_Translation
13	FC-Siam-conv [126]		ICIP, 2018	https://github.com/rcdaudt/fully_convolutional_change_detection
14	FC-Siam-diff [126]		ICIP, 2018	https://github.com/rcdaudt/fully_convolutional_change_detection
15	MFCN [114]		GRSL, 2021	https://github.com/lixinghua5540/MFCN
16	FC-EF-Res [127]		CVIU, 2019	https://github.com/rcdaudt/fully_convolutional_change_detection
17	DTCDN [108]		ISPRS, 2021	https://github.com/lixinghua5540/DTCDN

scenarios with small targets and sample imbalances [161]. Zheng et al. [162] combined the cross-entropy loss with the cube loss, which effectively improved the CD performance of the detector under edge errors and sample imbalance. To endow the loss function with both spatial constraint and regional perceptual properties, Peng et al. [163] focused on the hard samples with FL and then combined it with the dice loss, which is robust to sample imbalance, thereby enhancing the performance of the semantic CD detector. In addition, the adversarial loss function is often combined with GAN so as to reduce the feature domain shift between two periods of images due to illumination, seasonal changes, etc. For example, Zhao et al. [164] used the maximum mean deviation MMD to construct a loss function for generative adversarial adaptation networks that can adjust the mapping relations between predicted and real changes. The loss function is easy to apply and can better handle problems such as sample imbalance while improving the generalization performance and noise immunity of change detectors, but is highly dependent on hyperparameter settings when performing weighted combination or assignment optimization.

III. EXPERIMENTS

In order to verify the performance of different CDCD methods, 17 methods are selected to conduct the experiment. These 17 methods can be divided into two categories: conventional methods and DL-based methods. The specific information for the selected 17 methods are summarized in Table III, including method type, publication, and code link. In addition, this review paper uses eight datasets corresponding to CDCD in two different situations, including heterogeneous images and images with different imaging conditions. Each situation includes four specific scenes, as shown in Figs. 8 and 9. These datasets were widely utilized in the CD tasks. To quantitatively evaluate the accuracy of these models, this article selected 10 evaluation metrics, including false alarm (FA), missing alarm (MA), total error (TE), overall accuracy (OA), Kappa coefficient (Kappa), average accuracy (AA), completeness (CP), correctness (CR),

quality (QA), and F1-score (F1). These metrics are widely used for evaluating CD accuracy and described in [8], [165].

Fig. 10, Tables IV and V present the results for the evaluation. For clarity, Fig. 10 and Table IV show the qualitative and quantitative performances of the 17 methods for the California dataset, respectively, whereas Table V presents the results of the FreshNet model for the eight datasets. A summary of the main observations after conducting the experimental results follows.

- 1) The experiment demonstrates the superiority of DL-based methods over conventional methods in CDCD. This can be attributed to the remarkable learning capabilities of diverse network architectures employed in DL-based methods. These methods exhibit enhanced proficiency in detecting both pseudochanges and real changes across various cross-domain scenes. Consequently, they contribute significantly to the advancement of CDCD.
- 2) The performance of a specific approach, such as FC-EF-Res (see Table V), is different when applied to different datasets. Due to the different interfering factors for the detection of real change areas in different cross-domain scenes, and the uniqueness of each application case. Therefore, improving the generalization of detecting approaches is still required in the future study of CDCD.

IV. FUTURE DEVELOPMENTS OF CDCD METHODS

With the rapid development of computer technology, CDCD methods have also been developed to be more intelligent, accurate, and efficient. Datasets and models are two of the most important factors that determine the capability of CDCD methods. In the future, CDCD methods will be developed in the following three directions: 1) Large-scale cross-domain datasets. Currently, such datasets are not available and constructing such datasets can provide data foundation for the development of CDCD. 2) Foundation models. These models can be adapted to various cross-domain scenes, which have strong generality and robustness but may not achieve the highest accuracy in specific scenes. 3) Specialized models. These models are only adapted to specific cross-domain scenes, which are of high accuracy but

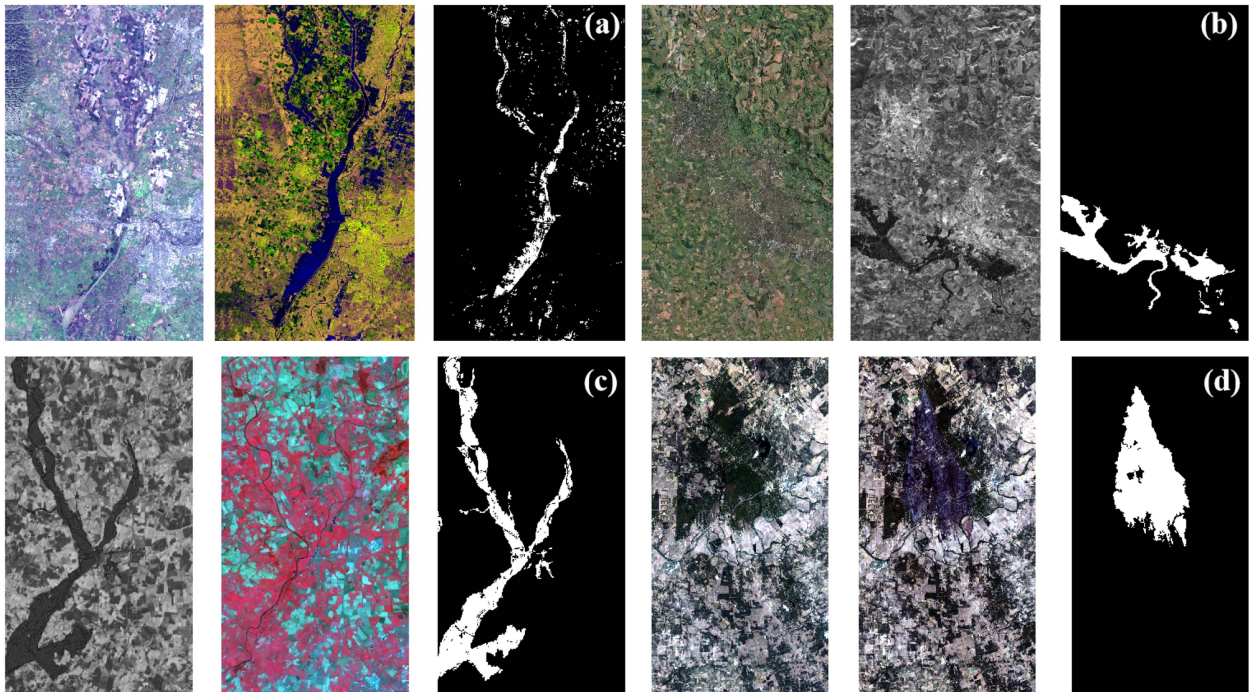


Fig. 8. Heterogeneous remote sensing images for evaluation of CDCD methods (CDCD 1), (a) California dataset, (b) U.K.-1, (c) U.K.-2, (d) Texas dataset.

TABLE IV
QUANTITATIVE COMPARISON USING TEN METRICS AMONG THE SELECTED METHODS FOR THE CALIFORNIA DATASET

No.	Methods	FA	MA	TE	OA	Kappa	AA	CP	CR	QA	F1
1	NPSG	0.0243	0.5612	0.0495	0.9505	0.4283	0.7216	0.4388	0.4707	0.2938	0.4542
2	IRG-McS	0.0232	0.4658	0.0410	0.9590	0.4903	0.7356	0.5342	0.4908	0.3437	0.5116
3	INLPG	0.0145	0.6774	0.0774	0.9226	0.4059	0.8159	0.3226	0.6989	0.2832	0.4414
4	GIR_MRF	0.0237	0.4421	0.0394	0.9606	0.4953	0.7311	0.5579	0.4795	0.3475	0.5157
5	ISTCRF	0.0193	0.4793	0.0417	0.9583	0.5274	0.7783	0.5207	0.5810	0.3785	0.5492
6	SCASC	0.0263	0.4644	0.0413	0.9587	0.4490	0.7012	0.5356	0.4190	0.3073	0.4702
7	CICMM	0.0389	0.8947	0.1009	0.8991	0.0811	0.5533	0.1053	0.1743	0.0702	0.1312
8	FPMABA	0.0307	0.5747	0.0487	0.9513	0.3409	0.6505	0.4253	0.3208	0.2238	0.3657
9	cGAN	0.0258	0.6970	0.0692	0.9308	0.3265	0.7006	0.3030	0.4484	0.2207	0.3616
10	SCCN	0.0069	0.7243	0.1052	0.8948	0.3766	0.8800	0.2757	0.8638	0.2642	0.4180
11	X-Net	0.0188	0.7388	0.0930	0.9070	0.3253	0.7679	0.2612	0.6154	0.2245	0.3667
12	ACE-Net	0.0189	0.7352	0.0912	0.9088	0.3285	0.7668	0.2648	0.6112	0.2266	0.3695
13	FC-Siam-conv	0.0132	0.1613	0.0187	0.9813	0.7593	0.8518	0.8387	0.7099	0.6246	0.7689
14	FC-Siam-diff	0.0187	0.2540	0.0268	0.9732	0.6436	0.7892	0.7460	0.5875	0.4895	0.6573
15	MFCN	0.0098	0.1598	0.0159	0.9841	0.8035	0.8892	0.8402	0.7853	0.6832	0.8118
16	FC-EF-Res	0.0088	0.1099	0.0128	0.9872	0.8400	0.9013	0.8901	0.8072	0.7341	0.8466
17	DTCDN	0.0147	0.4375	0.0371	0.9629	0.5973	0.8290	0.5625	0.6823	0.4457	0.6166

TABLE V
PERFORMANCE COMPARISON OF FC-EF-RES ON DIFFERENT CROSS DOMAIN SCENES INCLUDING CDCD 1 AND CDCD 2

Scenes	Dataset	FA	MA	TE	OA	Kappa	AA	CP	CR	QA	F1
CDCD 1	California	0.0088	0.1099	0.0128	0.9872	0.8400	0.9013	0.8901	0.8072	0.7341	0.8466
	UK-1	0.0097	0.3319	0.0353	0.9647	0.7318	0.9139	0.6681	0.8559	0.6006	0.7504
	UK-2	0.0072	0.0504	0.0123	0.9877	0.9413	0.9701	0.9496	0.9469	0.9016	0.9483
	Texas	0.0027	0.0364	0.0063	0.9937	0.9669	0.9865	0.9636	0.9773	0.9426	0.9704
CDCD 2	CDD-1	0.0263	0.2376	0.0396	0.9604	0.6871	0.8226	0.7624	0.6613	0.5483	0.7082
	CDD-2	0.0168	0.3183	0.0307	0.9693	0.7939	0.9825	0.6817	0.9977	0.6806	0.8100
	LEVIR-1	0.0069	0.0599	0.0127	0.9873	0.9346	0.9680	0.9401	0.9433	0.8898	0.9417
	LEVIR-2	0.0120	0.1008	0.0199	0.9801	0.8778	0.9343	0.8992	0.8783	0.7996	0.8886

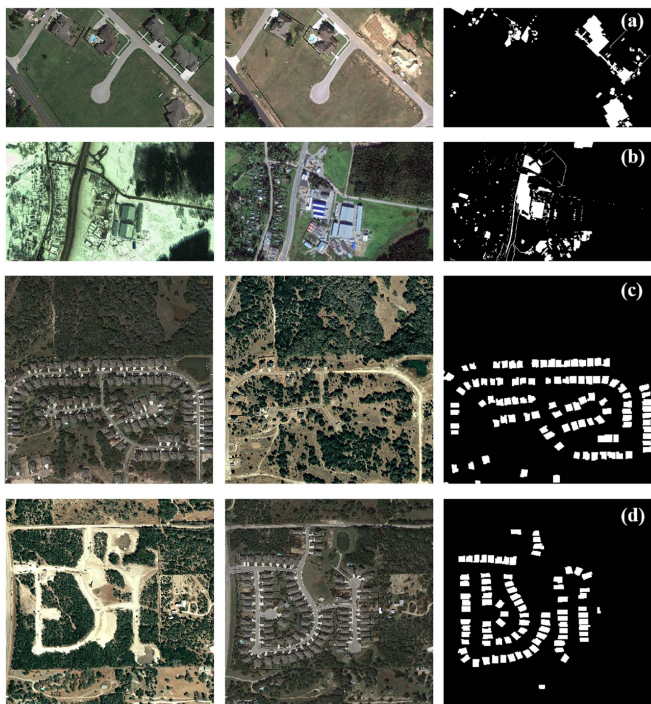


Fig. 9. Images acquired with different conditions for evaluation of CDCD methods (CDCD 2), (a-b) CDD dataset, (c-d) LEVIR dataset.

may not strong generality when the predict domain are different from the specific scene.

A. Large-Scale CDCD Datasets

Constrained by the costs of manual annotation and the challenges of visual interpretation, the current collection of open source remote sensing CD datasets predominantly comprises same-domain images, with relatively fewer cross-domain images. Moreover, in terms of change annotation types, both same-domain and cross-domain images only focus on a single category or a few specific types of change objects, such as buildings [166], vegetation [167], rivers [167], [168], etc. Generally, the same-domain multicategory datasets containing change information for building-related surface cover objects and vegetation-related surface cover objects [168], [169], [170], and the cross-domain multicategory CD datasets consisting of instances of change in different categories such as buildings and roads [171]. This also leads to high accuracy on same-domain/cross-domain CD datasets containing only a single/few types, but weak robustness and generalization against multitype CDCD tasks when using current data-driven DL models. With the types and numbers of Earth observation satellites increasing and the diversity of imaging conditions improving, the detection of changes in large-scale multitype cross-domain remote sensing images has gradually become a regular application scenario, which poses great challenges to the datasets in terms of sample size, coverage, semantic label richness, and heterogeneity of the preceding and following temporal phases. At present, a number of large-scale remote sensing datasets for multiple applications, represented

by the Wuhan Multi-Application VHR Scene classification dataset (WH-MAVS) [172], attempt to bridge the gap between algorithmic research and field applications by improving the completeness of the datasets; and the artificial intelligence large models, led by ChatGPT, have shown great potential in remote sensing research with their excellent performance in research such as geographic quiz. Therefore, building large-scale multitype cross-domain datasets, with large model-oriented, wide range, multitemporal, and multiapplication as the main objective in designing datasets, is one of the important research directions for CDCD in the future.

B. Foundation Models

Foundation models have made significant progress in natural image-based visual tasks due to their good scalability and representation ability. These models are usually pretrained using large-scale and diverse datasets and can be adapted to a range of downstream tasks by fine-tuning model parameters, such as CLIP [173], ALIGN [174], Florence [175], and NÜWA [176] models. Foundation model has also gradually extended to the field of remote sensing image perception. Wang et al. [177] proposed the remote sensing foundation model, which uses masked image modeling to obtain initial weights through unsupervised pretraining on a large-scale remote sensing dataset. Sun et al. [178] proposed a foundation model, which is called as RingMo, with masked image modeling for remote sensing feature representation. RingMo consists of two steps: the first step for generating a large-scale dataset, whereas the second step for proposing a training method for downstream tasks. In these works, fine-tuning experiments on multiple remote sensing tasks show that the foundation model has advantages in accuracy, efficiency, and interpretability. Incorporating the foundation vision-language model, Dong et al. [179] established ChangeCLIP, which is a new paradigm for CDCD. It incorporated a vision-language-driven decoder for combination of image-text encoding and visual feature decoding. In order to enhance the mining capabilities from spectral data, Hong et al. [180] proposed a universal remote sensing foundation model, named SpectralGPT, by using a novel 3-D generative pretrained transformer, which advances downstream CD tasks. Li et al. [181] built a universal foundation model-based CD method to make full use of knowledge of foundation models for CD, in which the frozen foundation model (e.g., CLIP) was utilized. Aiming at enhance the ability of foundation models in dealing with cross-domain images, Guo et al. [182] introduced SkySense, which incorporated multimodal spatiotemporal encoders and included over billions of parameters. With the development of remote sensing technology and computer science, proposing a CD foundation model or a high-level remote sensing data foundation model containing CD will be the direction of future development. Due to the large number of parameters in the foundation model and the large training dataset used, how to extract discriminative information that adapts to various downstream tasks to learn better ground object representations while significantly reducing computational cost and memory footprint will be a challenge for this research.

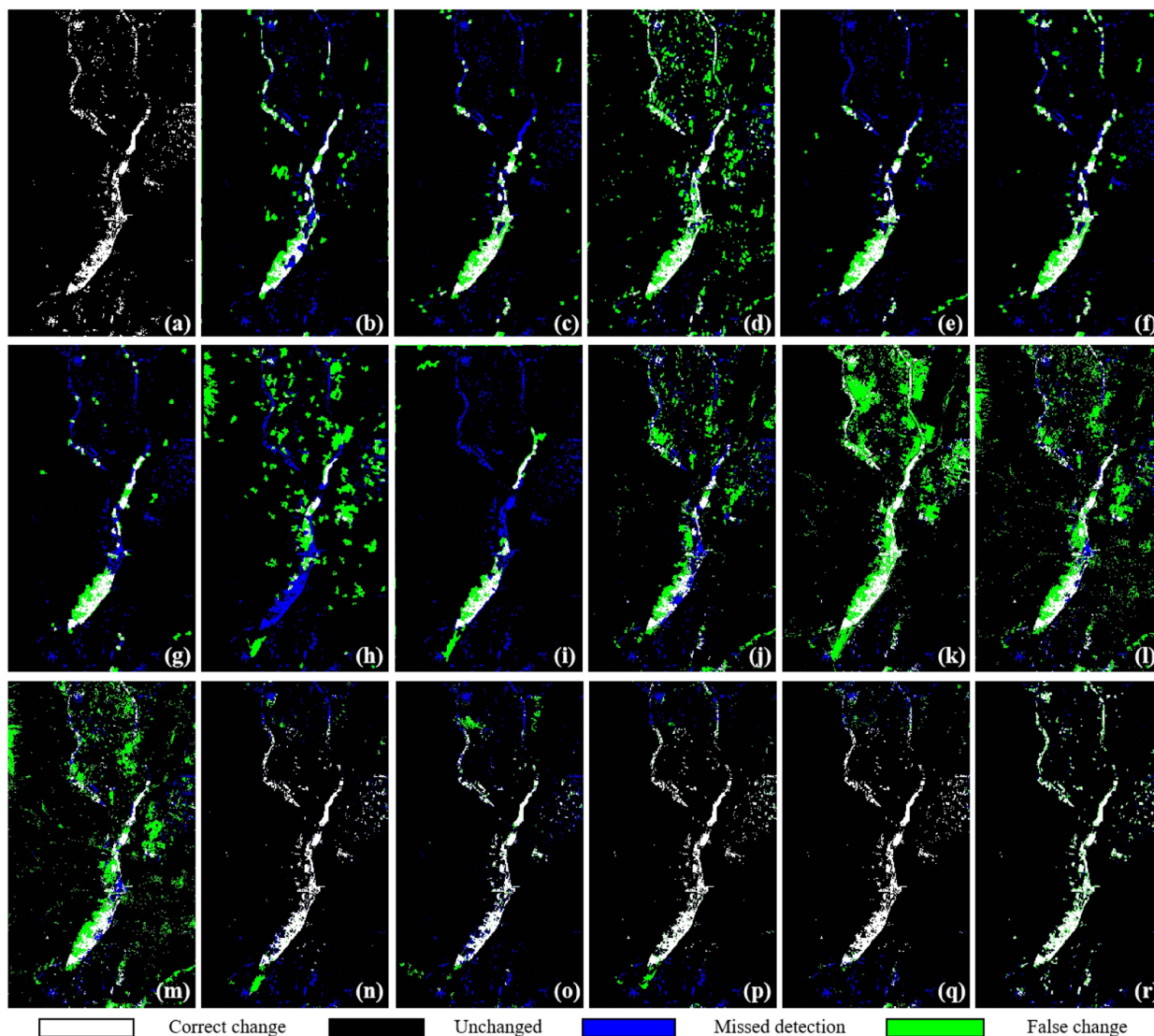


Fig. 10. CDCD results with different methods for the California dataset: (a) Reference image, (b) NPSG [16], (c) IRG-McS [76], (d) INLPG [81], (e) GIR_MRF [78], (f) ISTCRF [82], (g) SCASC [83], (h) CICMM [74], (i) FPMABA [75], (j) cGAN [140], (k) SCCN [117], (l) X-Net [122], (m) ACE-Net [122], (n) FC-Siam-conv [126], (o) FC-Siam-diff [126], (p) MFCN [114], (q) FC-EF-Res [127], (r) DTCDN [108].

C. Specialized Models

In contrast to foundation models, specialized models effectively solve one or more bottlenecks in CDCD by specializing in the design of some modules, thereby improving the accuracy of CD in specific scenes. Sun et al. [16] used nonlocal similarity to detect changes between optical and SAR images. They established a connection between heterogeneous data based on nonlocal patch similarity and then measured the change level based on whether the graph structures of the two images were consistent, overcoming the problem that the optical and SAR images cannot be directly compared due to differences in the physical quantity. Yang et al. [20] proposed an unsupervised CD method based on a time-distance-guided convolutional recurrent network, which uses a novel time-distance-guided long short-term memory unit to suppress the influence of pseudochanges caused by seasonal variations. They added time-distance-guided

gate modeling to the input gate and forget gate of the traditional long short-term memory (LSTM) network to adapt irregular time distances, thereby improving the feature extraction ability of time series images with irregular time distances and thus solving the problem caused by seasonal variations. To effectively utilize remote sensing images with different spatial resolutions, Liu et al. [35] proposed a CD method based on image super-resolution and attention mechanism, where super-resolution can solve the influence of different distributions, and attention mechanism can enhance the useful information in multiscale features, thus achieving high-precision CDCD with different spatial resolutions. Generally speaking, one type of technology can effectively solve a specific problem from CDCD. Some technologies, such as generative adversarial models, can also be used for multiple CDCD scenes, such as CD using optical and SAR images [108] and CD of optical images in different imaging conditions [13]. Nevertheless it can be concluded that,

to improve the accuracy of CD in specific cross-domain scenes, specialized models need relatively complex network structures or relatively strict constraints, resulting in numerous model parameters and weak generalization performance. How to use CDCD models with simpler structures or looser constraints for a specific scene will be future challenge.

V. CONCLUSION

With the rapid development of remote sensing imaging technology, a significant amount of cross-domain images are now collected by different sensors or under different imaging conditions. CDCD methods aim to address the problems caused by these cross-domain images, such as different physical observations, noise effects or illumination, and thus play critical role in emergencies and high-precision demanded CD tasks. This article summarizes the recent advances in image preprocessing, feature representation, and change detectors. Emphasis was paid to the feature representation part because it represents the core of CDCD methods. We classified the feature representation technology into two key categories, i.e., conventional methods and DL-based methods. The former one relies on expert-designed techniques, whereas the latter one utilizes the capabilities of DL-based methods. Experiments have also been conducted to compare 17 widely adopted methods over different cross-domain scenes and reveal the advantages of DL-based methods, compared to conventional methods. Finally, with the availability of cross-domain images and development of DL technology, future developments of CDCD are presented, including large-scale CDCD datasets, foundation models, and specialized models. We anticipate that the performance of CDCD methods will be greatly improved in terms of efficiency, accuracy and intelligence, and consequently benefit downstream applications in the future.

REFERENCES

- [1] H. Mahmoudzadeh, "Digital change detection using remotely sensed data for monitoring green space destruction in Tabriz," *Int. J. Environ. Res.*, vol. 1, no. 1, pp. 35–41, 2007.
- [2] Z. Zhu, C. E. Woodcock, C. Holden, and Z. Yang, "Generating synthetic landsat images based on all available landsat data: Predicting landsat surface reflectance at any given time," *Remote Sens. Environ.*, vol. 162, pp. 67–83, 2015.
- [3] Y. Qing et al., "Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, 2022, Art. no. 102899.
- [4] X. Huang, Y. Cao, and J. Li, "An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images," *Remote Sens. Environ.*, vol. 244, 2020, Art. no. 111802.
- [5] B. Yang, Y. Mao, L. Liu, X. Liu, Y. Ma, and J. Li, "From trained to untrained: A novel change detection framework using randomly initialized models with spatial-channel augmentation for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Mar. 2023, Art. no. 4402214.
- [6] C. Wu, B. Du, and L. Zhang, "Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9774–9788, Aug. 2023.
- [7] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1688.
- [8] Z. Lv et al., "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proc. IEEE*, vol. 110, no. 12, pp. 1976–1991, Dec. 2022.
- [9] L. Ru, B. Du, and C. Wu, "Multi-temporal scene classification and scene change detection with correlation based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 1382–1394, Nov. 2021.
- [10] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.
- [11] Y. Sun, L. Lei, D. Guan, J. Wu, G. Kuang, and L. Liu, "Image regression with structure cycle consistency for heterogeneous change detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1613–1627, Feb. 2024.
- [12] 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.
- [13] W. Zhao, L. Mou, J. Chen, Y. Bo, and W. J. Emery, "Incorporating metric learning and adversarial network for seasonal invariant change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2720–2731, Apr. 2020.
- [14] C. Zhang et al., "A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 109, 2022, Art. no. 102769.
- [15] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.
- [16] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognit.*, vol. 109, 2021, Art. no. 107598.
- [17] 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.
- [18] L. Bergamasco, S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised change detection using convolutional-autoencoder multiresolution features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4408119.
- [19] B. Yang, L. Qin, J. Liu, and X. Liu, "IRCNN: An irregular-time-distanced recurrent convolutional neural network for change detection in satellite time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2503905.
- [20] B. Yang, L. Qin, J. Liu, and X. Liu, "UTRNet: An unsupervised time-distance-guided convolutional recurrent network for change detection in irregularly collected images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410516.
- [21] Z. Zhu, "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 370–384, 2017.
- [22] Z. Zhu, S. Qiu, and S. Ye, "Remote sensing of land change: A multifaceted perspective," *Remote Sens. Environ.*, vol. 282, 2022, Art. no. 113266.
- [23] F. Pacifici, N. Longbotham, and W. J. Emery, "The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6241–6256, Oct. 2014.
- [24] N. Longbotham, C. Chaapel, L. Bleiler, C. Padwick, W. J. Emery, and F. Pacifici, "Very high resolution multiangle urban classification analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1155–1170, Apr. 2012.
- [25] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.
- [26] W. Shi and A. Shaker, "The line-based transformation model (LBTM) for image-to-image registration of high-resolution satellite image data," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3001–3012, 2006.
- [27] P. Xiao, X. Zhang, D. Wang, M. Yuan, X. Feng, and M. Kelly, "Change detection of built-up land: A framework of combining pixel-based detection and object-based recognition," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 402–414, 2016.
- [28] X. Wan, J. Liu, S. Li, J. Dawson, and H. Yan, "An illumination-invariant change detection method based on disparity saliency map for multitemporal optical remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1311–1324, Mar. 2019.
- [29] X. Wang, Z. Jia, J. Yang, and N. Kasabov, "Change detection in SAR images based on the logarithmic transformation and total variation denoising method," *Remote Sens. Lett.*, vol. 8, no. 3, pp. 214–223, 2017.
- [30] S. Iino, R. Ito, K. Doi, T. Imaizumi, and S. Hikosaka, "CNN-based generation of high-accuracy urban distribution maps utilising sar satellite imagery for short-term change monitoring," *Int. J. Image Data Fusion*, vol. 9, no. 4, pp. 302–318, 2018.

- [31] X. Lou, Z. Jia, J. Yang, and N. Kasabov, "Change detection in SAR images based on the ROF model semi-implicit denoising method," *Sensors*, vol. 19, no. 5, 2019, Art. no. 1179.
- [32] C. Chen et al., "Remote sensing image augmentation based on text description for waterside change detection," *Remote Sens.*, vol. 13, no. 10, 2021, Art. no. 1894.
- [33] B. Peng, Z. Meng, Q. Huang, and C. Wang, "Patch similarity convolutional neural network for urban flood extent mapping using Bi-temporal satellite multispectral imagery," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2492.
- [34] R. Zhou et al., "A unified deep learning network for remote sensing image registration and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5101216.
- [35] 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 Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4403718.
- [36] X. Ou, L. Liu, S. Tan, G. Zhang, W. Li, and B. Tu, "A hyperspectral image change detection framework with self-supervised contrastive learning pretrained model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7724–7740, Sep. 2022.
- [37] Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9059–9070, Nov. 2019.
- [38] Y. Ye, T. Tang, B. Zhu, C. Yang, B. Li, and S. Hao, "A multiscale framework with unsupervised learning for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5622215.
- [39] W. Ma et al., "Remote sensing image registration with modified sift and enhanced feature matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 3–7, Jan. 2017.
- [40] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Inf. Fusion*, vol. 73, pp. 22–71, 2021.
- [41] L. Yan, Z. Wang, Y. Liu, and Z. Ye, "Generic and automatic Markov random field-based registration for multimodal remote sensing image using grayscale and gradient information," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1228.
- [42] S. M. Vicente-Serrano, F. Pérez-Cabello, and T. Lasanta, "Assessment of radiometric correction techniques in analyzing vegetation variability and change using time series of landsat images," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3916–3934, 2008.
- [43] X. Chen, L. Vierling, and D. Deering, "A simple and effective radiometric correction method to improve landscape change detection across sensors and across time," *Remote Sens. Environ.*, vol. 98, no. 1, pp. 63–79, 2005.
- [44] C. Song, C. E. Woodcock, K. C. Seto, M. P. Lenney, and S. A. Macomber, "Classification and change detection using landsat tm data: When and how to correct atmospheric effects?," *Remote Sens. Environ.*, vol. 75, no. 2, pp. 230–244, 2001.
- [45] T. A. Schroeder, W. B. Cohen, C. Song, M. J. Canty, and Z. Yang, "Radiometric correction of multi-temporal landsat data for characterization of early successional forest patterns in western oregon," *Remote Sens. Environ.*, vol. 103, no. 1, pp. 16–26, 2006.
- [46] Z. Liu, R. Feng, L. Wang, W. Han, and T. Zeng, "Dual learning-based graph neural network for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5628614.
- [47] X. Dong, X. Sun, X. Jia, Z. Xi, L. Gao, and B. Zhang, "Remote sensing image super-resolution using novel dense-sampling networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1618–1633, Feb. 2021.
- [48] L. Chen, H. Liu, M. Yang, Y. Qian, Z. Xiao, and X. Zhong, "Remote sensing image super-resolution via residual aggregation and split attentional fusion network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9546–9556, Sep. 2021.
- [49] Y. Afaq and A. Manocha, "Analysis on change detection techniques for remote sensing applications: A review," *Ecological Inform.*, vol. 63, 2021, Art. no. 101310.
- [50] H. Chen, W. Li, and Z. Shi, "Adversarial instance augmentation for building change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5603216.
- [51] D. A. Stow, D. Collins, and D. McKinsey, "Land use change detection based on multi-date imagery from different satellite sensor systems," *Geocarto Int.*, vol. 5, no. 3, pp. 3–12, 1990.
- [52] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tournet, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.
- [53] K. Johansen, L. A. Arroyo, S. Phinn, and C. Witte, "Comparison of GEO-object based and pixel-based change detection of riparian environments using high spatial resolution multi-spectral imagery," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 2, pp. 123–136, 2010.
- [54] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogrammetry Remote Sens.*, vol. 80, pp. 91–106, 2013.
- [55] L. Wan, T. Zhang, and H. You, "Object-based method for optical and SAR images change detection," *J. Eng.*, vol. 2019, no. 21, pp. 7410–7414, 2019.
- [56] A. Bhatt, S. K. Ghosh, and A. Kumar, "Spectral indices based object oriented classification for change detection using satellite data," *Int. J. Syst. Assurance Eng. Manage.*, vol. 9, pp. 33–42, 2018.
- [57] L. Wan, Y. Xiang, and H. You, "An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019.
- [58] H. Chen, F. He, and J. Liu, "Heterogeneous images change detection based on iterative joint global-local translation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9680–9698, Jul. 2022.
- [59] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.
- [60] J. Chen, X. Chen, X. Cui, and J. Chen, "Change vector analysis in posterior probability space: A new method for land cover change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 317–321, Mar. 2011.
- [61] D. K. Sari, E. T. Hermawan, and G. Hudman, "Study on vegetation cover change in the province of South Kalimantan using RGB-NDVI unsupervised classification method," in *Proc. Map Asia Conf.*, 2005, pp. 1–9.
- [62] R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy, "Land-cover change detection using multi-temporal modis NDVI data," *Remote Sens. Environ.*, vol. 105, no. 2, pp. 142–154, 2006.
- [63] B. Liu, J. Chen, J. Chen, and W. Zhang, "Land cover change detection using multiple shape parameters of spectral and NDVI curves," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1251.
- [64] H.-J. Gim, C.-H. Ho, S. Jeong, J. Kim, S. Feng, and M. J. Hayes, "Improved mapping and change detection of the start of the crop growing season in the us corn belt from long-term AVHRR NDVI," *Agricultural forest Meteorol.*, vol. 294, 2020, Art. no. 108143.
- [65] P. Gong, D. J. Marceau, and P. J. Howarth, "A comparison of spatial feature extraction algorithms for land-use classification with spot HRV data," *Remote Sens. Environ.*, vol. 40, no. 2, pp. 137–151, 1992.
- [66] L. Li and M. K. Leung, "Integrating intensity and texture differences for robust change detection," *IEEE Trans. image Process.*, vol. 11, no. 2, pp. 105–112, Feb. 2002.
- [67] F. Yang and R. Lishman, "Land cover change detection using Gabor filter texture," in *Proc. 3rd Int. Workshop Texture Anal. Synth.*, 2003, vol. 17, pp. 78–83.
- [68] S. Berberoğlu, A. Akin, P. Atkinson, and P. Curran, "Utilizing image texture to detect land-cover change in mediterranean coastal wetlands," *Int. J. Remote Sens.*, vol. 31, no. 11, pp. 2793–2815, 2010.
- [69] D. Tomowski, M. Ehlers, and S. Klonus, "Colour and texture based change detection for urban disaster analysis," in *Proc. Joint Urban Remote Sens. Event*, 2011, pp. 329–332.
- [70] Z. Li, W. Shi, M. Hao, and H. Zhang, "Unsupervised change detection using spectral features and a texture difference measure for VHR remote-sensing images," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 7302–7315, 2017.
- [71] 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.
- [72] R. Touati, M. Mignotte, and M. Dahmane, "A reliable mixed-norm-based multiresolution change detector in heterogeneous remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3588–3601, Sep. 2019.
- [73] 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, Aug. 2020.
- [74] R. Touati, M. Mignotte, and M. Dahmane, "A circular invariant convolution model-based mapping for multimodal change detection," *Adv. Sci., Tech. Eng. Syst. J.*, vol. 5, pp. 1288–1298, 2020.

- [75] M. Mignotte, "A fractal projection and Markovian segmentation-based approach for multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8046–8058, Nov. 2020.
- [76] 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, Jul. 2021.
- [77] D. A. Jimenez-Sierra, D. A. Quintero-Olaya, J. C. Alvear-Munoz, H. D. Benitez-Restrepo, J. F. Florez-Ospina, and J. Chanussot, "Graph learning based on signal smoothness representation for homogeneous and heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410416.
- [78] Y. Sun, L. Lei, X. Tan, D. Guan, J. Wu, and G. Kuang, "Structured graph based image regression for unsupervised multimodal change detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 185, pp. 16–31, 2022.
- [79] L. Lei, Y. Sun, and G. Kuang, "Adaptive local structure consistency-based heterogeneous remote sensing change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Nov. 2022, Art. no. 8003905.
- [80] L. Zhao, Y. Sun, L. Lei, and S. Zhang, "Auto-weighted structured graph-based regression method for heterogeneous change detection," *Remote Sens.*, vol. 14, no. 18, 2022, Art. no. 4570.
- [81] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 4700221.
- [82] Y. Sun, L. Lei, D. Guan, J. Wu, and G. Kuang, "Iterative structure transformation and conditional random field based method for unsupervised multimodal change detection," *Pattern Recognit.*, vol. 131, 2022, Art. no. 108845.
- [83] Y. Sun, L. Lei, D. Guan, M. Li, and G. Kuang, "Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 4405814.
- [84] K. Mubea and G. Menz, "Monitoring land-use change in Nakuru (Kenya) using multi-sensor satellite data," *Adv. Remote Sens.*, vol. 1, no. 3, pp. 74–84, 2012.
- [85] T. Han, Y. Tang, X. Yang, Z. Lin, B. Zou, and H. Feng, "Change detection for heterogeneous remote sensing images with improved training of hierarchical extreme learning machine (HELM)," *Remote Sens.*, vol. 13, no. 23, 2021, Art. no. 4918.
- [86] L. Wan, Y. Xiang, and H. You, "A post-classification comparison method for SAR and optical images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.
- [87] 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.
- [88] D. K. Seo, Y. H. Kim, Y. D. Eo, W. Y. Park, and H. C. Park, "Generation of radiometric, phenological normalized image based on random forest regression for change detection," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1163.
- [89] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
- [90] J. C. Price, "Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1277–1285, Sep. 1997.
- [91] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogrammetric Eng. Remote Sens.*, vol. 70, no. 7, pp. 793–802, 2004.
- [92] A. Ifarraguerri and M. W. Prairie, "Visual method for spectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 2, pp. 101–106, Apr. 2004.
- [93] Q. Chen, Y. Chen, and W. Jiang, "Genetic particle swarm optimization-based feature selection for very-high-resolution remotely sensed imagery object change detection," *Sensors*, vol. 16, no. 8, 2016, Art. no. 1204.
- [94] F. Cai, Z. Xiang, H. Cai, and D. Shan, "CVA multi-scale remote sensing image change detection combined with feature selection," *Bull. Surveying Mapping*, no. 8, pp. 101–104, 2020.
- [95] J. Chen, J. Zhu, Y. Guo, G. Sun, Y. Zhang, and M. Deng, "Unsupervised domain adaptation for semantic segmentation of high-resolution remote sensing imagery driven by category-certainty attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5616915.
- [96] J. Zhu, Y. Guo, G. Sun, L. Yang, M. Deng, and J. Chen, "Unsupervised domain adaptation semantic segmentation of high-resolution remote sensing imagery with invariant domain-level prototype memory," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5603518.
- [97] S. Wang, D. Hou, and H. Xing, "A self-supervised-driven open-set unsupervised domain adaptation method for optical remote sensing image scene classification and retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5605515.
- [98] J. Chen et al., "Memory-contrastive unsupervised domain adaptation for building extraction of high-resolution unsupervised remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5605615.
- [99] D. Hou, S. Wang, X. Tian, and H. Xing, "PCLUDA: A pseudo-label consistency learning-based unsupervised domain adaptation method for cross-domain optical remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Dec. 2022, Art. no. 5600314.
- [100] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.
- [101] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogrammetry Remote Sens.*, vol. 107, pp. 50–63, 2015.
- [102] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," 2019, *arXiv:1909.05948*.
- [103] T. Zhan, M. Gong, X. Jiang, and S. Li, "Log-based transformation feature learning for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1352–1356, Sep. 2018.
- [104] Y. Li, C. Peng, Y. Chen, L. Jiao, L. Zhou, and R. Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5751–5763, Aug. 2019.
- [105] W. Ma, Y. Xiong, Y. Wu, H. Yang, X. Zhang, and L. Jiao, "Change detection in remote sensing images based on image mapping and a deep capsule network," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 626.
- [106] R. Shao, C. Du, H. Chen, and J. Li, "Sunet: Change detection for heterogeneous remote sensing images from satellite and UAV using a dual-channel fully convolution network," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3750.
- [107] H. Li, M. Gong, M. Zhang, and Y. Wu, "Spatially self-paced convolutional networks for change detection in heterogeneous images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4966–4979, May 2021.
- [108] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 179, pp. 14–34, 2021.
- [109] Z. Lv, H. Huang, L. Gao, J. A. Benediktsson, M. Zhao, and C. Shi, "Simple multiscale UNet for change detection with heterogeneous remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, May 2022, Art. no. 2504905.
- [110] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
- [111] M. Gong, H. Yang, and P. Zhang, "Feature learning and change feature classification based on deep learning for ternary change detection in sar images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 129, pp. 212–225, 2017.
- [112] R. Jaturapitornchai, M. Matsuoka, N. Kanemoto, S. Kuzuoka, R. Ito, and R. Nakamura, "Newly built construction detection in SAR images using deep learning," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1444.
- [113] X. Cao, Y. Ji, L. Wang, B. Ji, L. Jiao, and J. Han, "SAR image change detection based on deep denoising and CNN," *IET Image Process.*, vol. 13, no. 9, pp. 1509–1515, 2019.
- [114] X. Li, M. He, H. Li, and H. Shen, "A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jul. 2021, Art. no. 8017505.
- [115] 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, Apr. 2022, Art. no. 5622519.
- [116] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7066–7080, Dec. 2017.
- [117] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.

- [118] N. Shi, K. Chen, G. Zhou, and X. Sun, "A feature space constraint-based method for change detection in heterogeneous images," *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 3057.
- [119] S. T. Seydi, M. Hasanlou, and J. Chanussot, "DSMNN-Net: A deep siamese morphological neural network model for burned area mapping using multispectral Sentinel-2 and hyperspectral PRISMA images," *Remote Sens.*, vol. 13, no. 24, 2021, Art. no. 5138.
- [120] P. Chen, L. Guo, X. Zhang, K. Qin, W. Ma, and L. Jiao, "Attention-guided siamese fusion network for change detection of remote sensing images," *Remote Sens.*, vol. 13, no. 22, 2021, Art. no. 4597.
- [121] Y. Wu, J. Li, Y. Yuan, A. 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.
- [122] L. T. Luppino et al., "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2021, Art. no. 4700422.
- [123] X. Jiang, G. Li, X.-P. Zhang, and Y. He, "A semisupervised Siamese network for efficient change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, Art. no. 4700718.
- [124] 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.
- [125] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.
- [126] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.
- [127] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understanding*, vol. 187, 2019, Art. no. 102783.
- [128] R. Liu, Z. Cheng, L. Zhang, and J. Li, "Remote sensing image change detection based on information transmission and attention mechanism," *IEEE Access*, vol. 7, pp. 156349–156359, 2019.
- [129] H. Chen, C. Wu, B. Du, and L. Zhang, "DSDANet: Deep siamese domain adaptation convolutional neural network for cross-domain change detection," 2020, *arXiv:2006.09225*.
- [130] J. Xu, C. Luo, X. Chen, S. Wei, and Y. Luo, "Remote sensing change detection based on multidirectional adaptive feature fusion and perceptual similarity," *Remote Sens.*, vol. 13, no. 15, p. 3053, 2021.
- [131] X. Zhang et al., "SMD-Net: Siamese multi-scale difference-enhancement network for change detection in remote sensing," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1580.
- [132] C. Sun, J. Wu, H. Chen, and C. Du, "SemiSAnet: A semi-supervised high-resolution remote sensing image change detection model using siamese networks with graph attention," *Remote Sens.*, vol. 14, no. 12, 2022, Art. no. 2801.
- [133] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, "FCCDN: Feature constraint network for VHR image change detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 187, pp. 101–119, 2022.
- [134] Z. Chen et al., "Egde-Net: A building change detection method for high-resolution remote sensing imagery based on edge guidance and differential enhancement," *ISPRS J. Photogrammetry Remote Sens.*, vol. 191, pp. 203–222, 2022.
- [135] M. Hu, C. Wu, and L. Zhang, "HyperNet: Self-supervised hyperspectral spatial-spectral feature understanding network for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2022, Art. no. 5543017.
- [136] C. Zhang, L. Wang, S. Cheng, and Y. Li, "SwinSUNet: Pure transformer network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5224713.
- [137] S. Fang, K. Li, and Z. Li, "Changer: Feature interaction is what you need for change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5610111.
- [138] Y. Feng, J. Jiang, H. Xu, and J. Zheng, "Change detection on remote sensing images using dual-branch multilevel intertemporal network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 4401015.
- [139] X. Zhang, S. Cheng, L. Wang, and H. Li, "Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 2000415.
- [140] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.
- [141] M. Gong, X. Niu, T. Zhan, and M. Zhang, "A coupling translation network for change detection in heterogeneous images," *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3647–3672, 2019.
- [142] M. Jia, C. Zhang, Z. Lv, Z. Zhao, and L. Wang, "Bipartite adversarial autoencoders with structural self-similarity for unsupervised heterogeneous remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Aug. 2022, Art. no. 6515705.
- [143] Z.-G. Liu, Z.-W. Zhang, Q. Pan, and L.-B. Ning, "Unsupervised change detection from heterogeneous data based on image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2021, Art. no. 4403413.
- [144] A. Manocha and Y. Afaq, "Optical and SAR images-based image translation for change detection using generative adversarial network (GAN)," *Multimedia Tools Appl.*, vol. 82, no. 17, pp. 26289–26315, 2023.
- [145] C.-L. Kuo and J.-H. Hong, "Interoperable cross-domain semantic and geospatial framework for automatic change detection," *Comput. Geosci.*, vol. 86, pp. 109–119, 2016.
- [146] Y. Kousuke, T. Kanji, and S. Takuma, "Use of generative adversarial network for cross-domain change detection," 2017, *arXiv:1712.08868*.
- [147] P. Jian, K. Chen, and W. Cheng, "GAN-based one-class classification for remote-sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jul. 2021, Art. no. 8009505.
- [148] M. Yang, L. Jiao, B. Hou, F. Liu, and S. Yang, "Selective adversarial adaptation-based cross-scene change detection framework in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2188–2203, Mar. 2021.
- [149] J. Liu, W. Xuan, Y. Gan, Y. Zhan, J. Liu, and B. Du, "An end-to-end supervised domain adaptation framework for cross-domain change detection," *Pattern Recognit.*, vol. 132, 2022, Art. no. 108960.
- [150] B. Fang, L. Pan, and R. Kou, "Dual learning-based Siamese framework for change detection using bi-temporal vhr optical remote sensing images," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1292.
- [151] J. Paul, B. U. Shankar, B. Bhattacharyya, and A. K. Datta, "Unsupervised change detection in remote sensing images using CNN based transfer learning," in *Proc. 5th Int. Conf. Adv. Comput. Data Sci.*, 2021, pp. 463–474.
- [152] X. Wang, T. E. Huang, T. Darrell, J. E. Gonzalez, and F. Yu, "Frustratingly simple few-shot object detection," 2020, *arXiv:2003.06957*.
- [153] A. Gretton et al., "Optimal kernel choice for large-scale two-sample tests," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2012, vol. 25, pp. 1205–1213.
- [154] 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, 2020, Art. no. 1662.
- [155] H. Chen, C. Wu, B. Du, and L. Zhang, "Deep siamese domain adaptation convolutional neural network for cross-domain change detection in multispectral images," 2020, *arXiv:2004.05745*.
- [156] 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 Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5510018.
- [157] R. Kou, B. Fang, G. Chen, and L. Wang, "Progressive domain adaptation for change detection using season-varying remote sensing images," *Remote Sens.*, vol. 12, no. 22, 2020, Art. no. 3815.
- [158] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised multiple-change detection in VHR multisensor images via deep-learning based adaptation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5033–5036.
- [159] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k -means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
- [160] Z. Wang, C. Peng, Y. Zhang, N. Wang, and L. Luo, "Fully convolutional siamese networks based change detection for optical aerial images with focal contrastive loss," *Neurocomputing*, vol. 457, pp. 155–167, 2021.
- [161] Z. Cao, M. Wu, R. Yan, F. Zhang, and X. Wan, "Detection of small changed regions in remote sensing imagery using convolutional neural network," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 502, no. 1, May 2020, Art. no. 012017.
- [162] D. Zheng, Z. Wei, Z. Wu, and J. Liu, "Learning pairwise potential CRFs in deep siamese network for change detection," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 841.
- [163] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, and P. He, "SCDNet: A novel convolutional network for semantic change detection in high resolution optical remote sensing imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 103, 2021, Art. no. 102465.

- [164] W. Zhao, X. Chen, X. Ge, and J. Chen, "Using adversarial network for multiple change detection in bitemporal remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Nov. 2020, Art. no. 8003605.
- [165] 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.
- [166] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.
- [167] P. J. S. Vega et al., "An unsupervised domain adaptation approach for change detection and its application to deforestation mapping in tropical biomes," *ISPRS J. Photogrammetry Remote Sens.*, vol. 181, pp. 113–128, 2021.
- [168] 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.
- [169] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, Oct. 2009.
- [170] N. Bourdis, D. Marraud, and H. Sahbi, "Constrained optical flow for aerial image change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 4176–4179.
- [171] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. A. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 565–571, 2018.
- [172] J. Yuan, L. Ru, S. Wang, and C. Wu, "WH-MAVS: A novel dataset and deep learning benchmark for multiple land use and land cover applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1575–1590, Jan. 2022.
- [173] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8748–8763.
- [174] C. Jia et al., "Scaling up visual and vision-language representation learning with noisy text supervision," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 4904–4916.
- [175] L. Yuan et al., "Florence: A new foundation model for computer vision," 2021, *arXiv:2111.11432*.
- [176] C. Wu et al., "NÜWA: Visual synthesis pre-training for neural visual world creation," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 720–736.
- [177] D. Wang et al., "Advancing plain vision transformer toward remote sensing foundation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2022, Art. no. 5607315.
- [178] X. Sun et al., "RingMo: A remote sensing foundation model with masked image modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jul. 2022, Art. no. 5612822.
- [179] S. Dong, L. Wang, B. Du, and X. Meng, "ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 208, pp. 53–69, Feb. 2024.
- [180] D. Hong et al., "SpectralGPT: Spectral remote sensing foundation model," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: [10.1109/TPAMI.2024.3362475](https://doi.org/10.1109/TPAMI.2024.3362475).
- [181] K. Li, X. Cao, and D. Meng, "A new learning paradigm for foundation model-based remote-sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5610112.
- [182] X. Guo et al., "SkySense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 27672–27683.



Jie Chen (Member, IEEE) received the M.S. degree in cartography and geographical information engineering from East China University of Technology, Nanchang, China, and the Ph.D. degree in cartography and geographical information engineering from Central South University, Changsha, China, in 2006 and 2010, respectively.

He is currently a Professor with the Department of Geoinformatics, Central South University, Changsha, China. He is a reviewer for several journals in the field of remote sensing. His research interests include

intelligent analysis and understanding of remote sensing image.



Dongyang Hou received the Ph.D. degree in geographic information system from the China University of Mining and Technology, Xuzhou, China, in 2016.

He is currently an Associate Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. His main research interests include spatial-temporal data crawling, mining, and remote sensing image retrieval.



Changxian He received the B.S. degree in geographic information science in 2020 from Central South University, Changsha, China, where he is currently working toward the M.S. degree in surveying and mapping science and technology.

His research interests include remote sensing image intelligent interpretation and road cross-domain extraction of high-resolution remote sensing images.



Yaoting Liu received the B.S. degree in automation from the University of South China, Hengyang, China, in 2020, and the M.S. degree in control science and engineering with the School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha, China, in 2023. He is currently working toward the Eng.D. degree in energy and power engineering with the College of Electrical and Information Engineering, Hunan University, Hunan, China.

His current research interests include remote sensing, data fusion, change detection, and deep learning.



Ya Guo received the M.S. degree in surveying and mapping engineering from China University of Geosciences, Beijing, China, in 2020. He is currently working toward the Ph.D. degree in cartography and geographical information engineering with Central South University in Changsha, China.

His research interests include remote sensing image intelligent interpretation and artificial intelligence.



Bin Yang (Member, IEEE) received the B.S. degree in science and technology of remote sensing from Beihang University, Beijing, China, in 2012, and the Ph.D. degree in photogrammetry and remote sensing from Peking University, Beijing, in 2017.

He is currently an Associate Professor with the College of Electrical and Information Engineering, Hunan University, Changsha, China, and with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Changsha, China. His research interests include remote sensing and artificial intelligence.