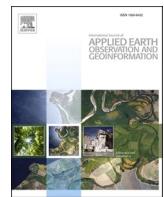




Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Deep learning change detection techniques for optical remote sensing imagery: Status, perspectives and challenges

Daifeng Peng ^{a,b,c,*}, Xuelian Liu ^a, Yongjun Zhang ^d, Haiyan Guan ^a, Yansheng Li ^d, Lorenzo Bruzzone ^e

^a School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

^b Technology Innovation Center for Integrated Applications in Remote Sensing and Navigation, Ministry of Natural Resources, Nanjing 210044, China

^c Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources, Wuhan 430079, China

^d School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

^e Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy



ARTICLE INFO

Keywords:

Change detection
Optical remote sensing image
Deep learning
Algorithms granularity
Review

ABSTRACT

Change detection (CD) aims to compare and analyze images of identical geographic areas but different dates, whereby revealing spatio-temporal change patterns of Earth's surface. With the implementation of the High-Resolution Earth Observation Project, an integrated sky-to-ground observation system has been continuously developed and improved. The accumulation of massive multi-modal, multi-angle, and multi-resolution remote sensing data have greatly enriched the CD data sources. Among them, high-resolution optical remote sensing images contain abundant spatial detail information, making it possible to interpret fine-grained scenes and greatly expand the application breadth and depth of CD. Generally, traditional optical remote sensing CD methods are cumbersome in steps and have a low level of automation. In contrast, artificial intelligence (AI) based CD methods possess powerful feature extraction and non-linear modeling capabilities, thereby gaining advantages that traditional methods cannot match. As a result, they have become the mainstream approaches in the field of CD. This review article systematically summarizes the datasets, theories, and methods of CD for optical remote sensing image. It provides a comprehensive analysis of AI-based CD algorithms based on deep learning paradigms from the perspectives of algorithm granularity. In-depth analysis of the performance of typical algorithms are further conducted. Finally, we summarize the challenges and trends of the CD algorithms in the AI era, aiming to provide important guidelines and insights for relevant researchers.

1. Introduction

Change detection (CD) techniques have been widely used in various research fields such as natural resource surveying, urban sprawl monitoring, environmental assessment, and rapid disaster response (Demir et al., 2013; Ertürk et al., 2017; De Alban et al., 2018). Due to the rapid advancements of sensor technology, the past decade has witnessed a massive accumulation of remote sensing imagery, providing huge opportunities for spatio-temporal change studies. Among these data sources, optical remote sensing (ORS) images are the most widely used, which can be easily accessible and have significant characteristics such as high spatial resolution, rich spectral information, and large image coverage. However, how to efficiently and rapidly extract useful features

and information from massive optical remote sensing data still poses great challenges, especially for accurate multi-temporal image CD task.

In general, CD workflow consists of five steps: *data input*, *pre-processing*, *change information extraction*, *change detection results*, and *accuracy evaluation* (as illustrated in Fig. 1), where change information extraction methods play a decisive role. Based on interpretation units, traditional CD algorithms mainly include pixel-level change detection (PLCD) and object-level change detection (OLCD). In PLCD, feature-level comparison of corresponding pixels are directly conducted to produce a difference image, and the final change map is obtained through thresholding or clustering analysis (Deng et al., 2008; Wu et al., 2017a). Nevertheless, contextual information is missed, which can be mitigated by using neighborhood windows (Celik, 2009) or probability

* Corresponding author at: School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.

E-mail address: daifeng@nuist.edu.cn (D. Peng).

graph model (Lv et al., 2018). Different from PLCD, OLCD methods are proposed by first segmenting the bi-temporal optical remote sensing images into independent and homogeneous image objects and then generating change maps using post-classification comparison (Su et al., 2011) or bi-temporal object comparison techniques (Zhang et al., 2017; Xu et al., 2019). In such case, spatial context information and rich object-level features can be utilized to improve CD performance. However, due to the multiscale effects of geographical objects and the complexity of their spatiotemporal distribution, determining a proper segmentation scale is often challenging.

It is worth noting that the aforementioned CD methods are characterized by cumbersome processing steps and poor robustness against interference from image backgrounds. Recently, deep learning (DL) has emerged as a revolutionary tool to open the door of artificial intelligence (AI). Due to excellent capabilities of feature representation and nonlinear modeling, DL technologies have made tremendous achievements in the both computer vision and remote sensing fields, such as object detection (Tao et al., 2019), image classification (Hua et al., 2019), and image super-resolution reconstruction (Lanaras et al., 2018). In such context, deep learning based CD (DLCD) algorithms have gained increasing attention in remote sensing community. Experimental studies demonstrated that DLCD algorithms, which possess stronger feature representation and higher levels of automation, have achieved tremendous advantages over traditional CD methods. As a result, they have become the mainstream in CD area.

In recent years, DLCD methods have been summarized in several review articles from the perspectives of AI development paradigm (Khelifi and Mignotte, 2020; Shi et al., 2020; Jiang et al., 2024), information representation dimensions (Bai et al., 2023), change modes (Zhu et al., 2022), supervised modes (Shafique et al., 2022; Cheng et al., 2023; Wang et al., 2024a) and network modules (Wu et al., 2024). However, in-depth analysis of existing DLCD methods for ORS images from the perspectives of algorithm granularity is scarce, which is easier to understand and is significant to grasp the cores of development paradigms and trends of DLCD. In addition, new opportunities and challenges for DLCD continue to emerge with the fast development of AI (such as self-supervised learning, diffusion models, large foundation models, etc.), whose influence on CD needs to be systematically explored. To this end, we conduct a comprehensive analysis of DLCD algorithms from the perspective of algorithm granularity, accuracy and efficiency comparisons are further made to provide significant guidelines for CD algorithm design in various scenarios. Furthermore, ORS change detection (ORSCD) datasets in the past decade have been systematically summarized to provide valuable guidance for related researchers. Finally, in combination with the latest technological trends such as semi-

supervised learning, self-supervised learning, knowledge graphs, and foundation models, this article provides a systematic summary and analysis of the challenges and future trends of CD in the AI era.

The organization structure of this paper is illustrated in Fig. 2. In Section 1, we discuss the research background, objectives, and significance. Section 2 presents a detailed illustration and analysis of DLCD methods from the perspectives of feature-based, patch-based, and image-based approaches. We also summarize their general frameworks, commonly used network architectures, feature enhancement modules, and label-efficient learning strategies. In Section 3, we present a review of benchmark datasets for ORSCD from the perspectives of binary change detection (BCD) and semantic change detection (SCD). In Section 4, we compare and analyze typical DLCD algorithms based on quantitative performance and model complexity. Section 5 summarizes the challenges and future trends of existing DLCD methods systematically. Finally, in Section 6, we draw the conclusions of this article.

2. Algorithm analysis of DLCD methods

To conduct an in-depth analysis of AI-based CD algorithms based on deep learning paradigm, we have selected related journal articles ranging from January 1, 2015, to August 24, 2024. The English database used was “Web of Science,” and the search strategy was “TS = (“change detection”) AND TS = (“remote sensing” OR “remotely sensed”) AND TS = (“deep learning” OR “network”). The document type was set to “academic journals”. After a thorough analysis, we identified a total of 1127 relevant articles. For the benefit of visual analysis, we present the year-by-year growth trends of for the publication quantity on DLCD methods in Fig. 3(a). One can observe that the number of DLCD publications exhibits an obvious upward trend year-by-year. This indicates that with the boom of AI techniques, more related methods and strategies have been introduced into CD areas. Consequently, DLCD approaches are becoming the mainstream and are drawing increasing attention from RS community. In addition, the publication number with regards to classical RS journals is statistically analyzed in Fig. 3(b), where the most popular academic journals for DLCD are from TGRS, RS, JSTARS and GRSL, respectively.

Based on comprehensive survey of the aforementioned literature, DLCD methods can be roughly categorized into three main types from the perspective of algorithm granularity: Feature-based DLCD (FB-DLCD), Patch-based DLCD (PB-DLCD), and Image-based DLCD (IB-DLCD). The basic principles, advantages and limitations are presented in Fig. 4.

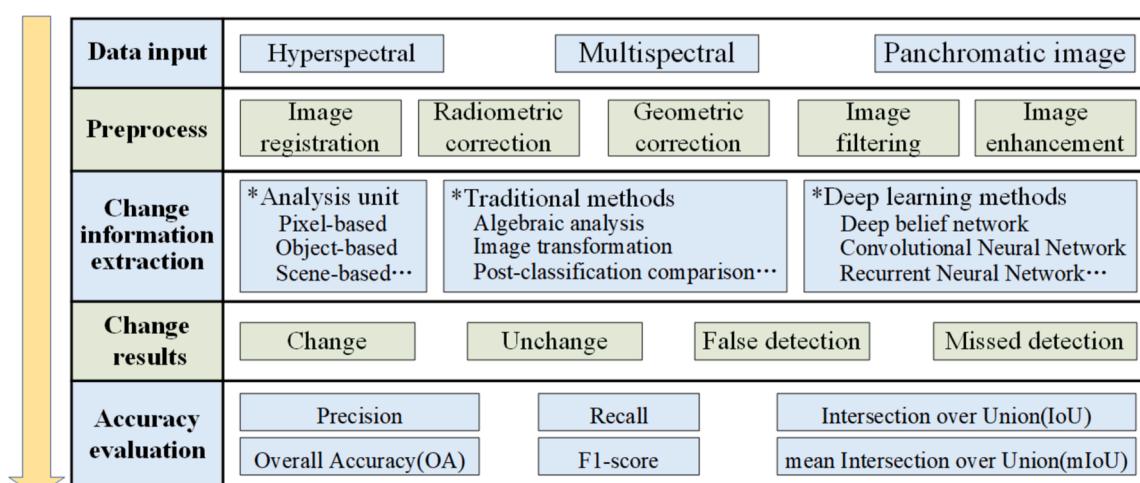


Fig. 1. Illustration of CD workflow.

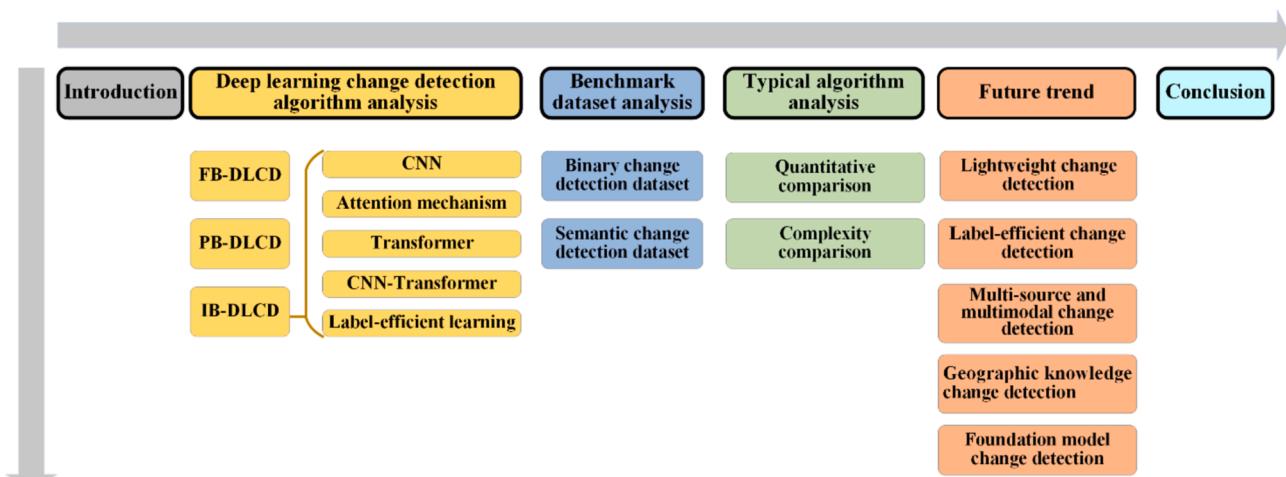


Fig. 2. The organization chart of this article.

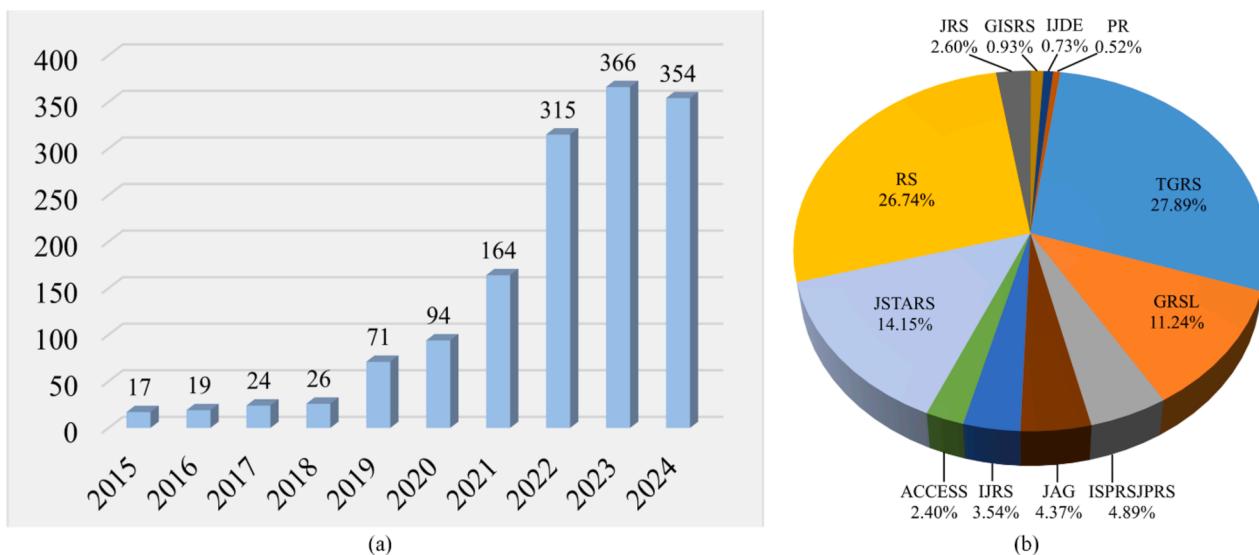


Fig. 3. Illustration of publication trends on DLCD. (a) Year-by-year growth trends of the number of DLCD publications. (b) Statistics of classical RS journals for DLCD.

2.1. Feature-based deep learning change detection

Traditional handcrafted features typically require large amounts of expert knowledge. As a result, they tend to lack generality and may not perform well in different scenarios. Differently, deep features are learned through pre-trained networks based on large-scale datasets, which are free from expert knowledge and of strong robustness. To exploit the potential of deep features for robust CD, FB-DLCD methods generally employ pre-trained networks to extract deep features for difference maps generation (Liu et al., 2019; Peng and Guan, 2019; Zhang and Shi, 2020). Then, difference image analysis techniques such as thresholding or clustering are used to produce final result map. The technical framework of these approaches is illustrated in Fig. 5.

In general, pre-trained models are capable of extracting deep features quickly and effectively. However, there is often a significant domain gap between the pre-trained dataset and the actual testing CD dataset, making it difficult to directly use pre-trained deep features. Two possible approaches can be considered to address this issue. One approach is to fine-tune the network using a real RS dataset such as MillionAID (Long et al., 2021). Another approach is to train a specific model by using a specific dataset to learn more discriminative features. Note that the input for FB-DLCD can be either 1D vector or 2D tensor. In

the former case, images are flattened to 1D vectors to train a simple neural network (Zhang et al., 2016), which is only capable of capturing spectral information while neglecting spatial structural information. On the contrary, spatial context information can be well preserved in the latter case, where a Siamese Convolutional Neural Network (Siamese CNN) architecture is usually constructed to obtain distinctive feature representations between different pixels (Zhan et al., 2017; Zhang et al., 2019). For instance, to produce deep features that are invariant to color and seasonal variations, Bandara and Patel (2023) proposed a novel unsupervised deep metric learning based CD by introducing a deep change probability generator (D-CPG) and a deep feature extractor (D-FE), which are iteratively optimized by similarity-dissimilarity loss and context consistency loss. Noh et al. (2024) proposed to learn robust deep features based on image reconstruction loss, Segment Anything Model (SAM) was further employed to refine coarse change map. Thanks to the flexibility of user-defined networks, more spatio-temporal features can be learned to include spatial hierarchical relationship (Xu et al., 2020). Additionally, segmented feature maps or difference maps can be combined with bi-temporal images to facilitate multi-scale information fusion, thereby detailed change information is captured (Hou et al., 2021; Tang et al., 2021). More recently, with the development of large foundation models, more discriminative and robust deep features can be

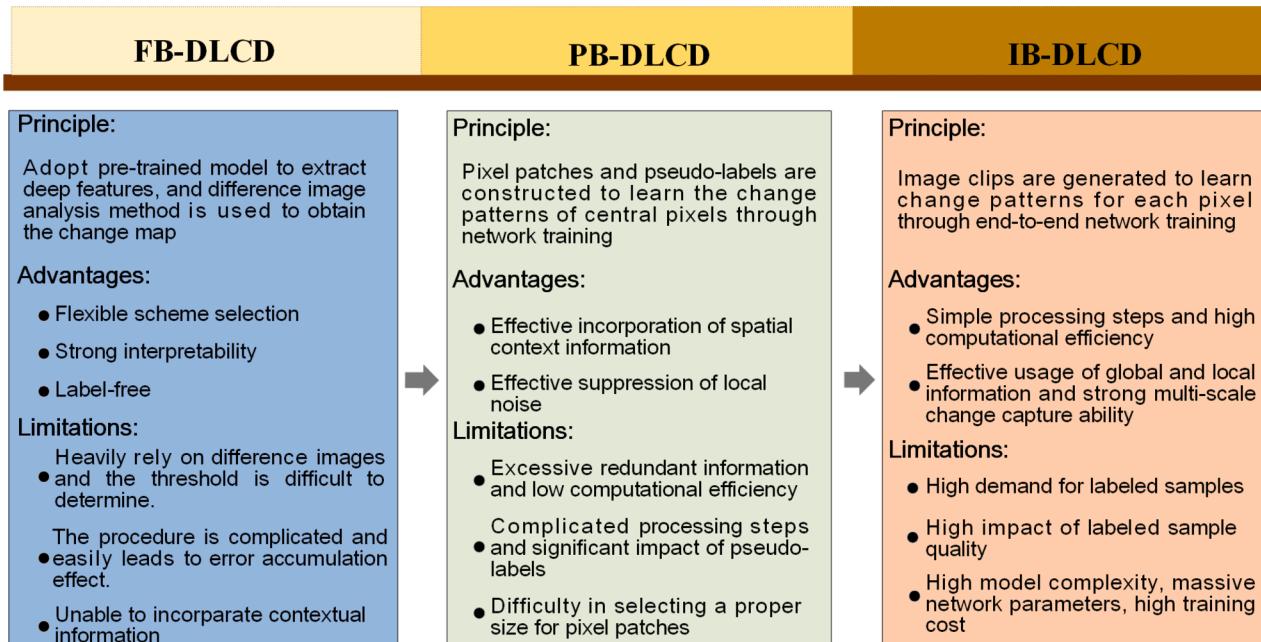


Fig. 4. Comparison and summary of three DLCD methods (FB-DLCD, PB-DLCD, IB-DLCD).

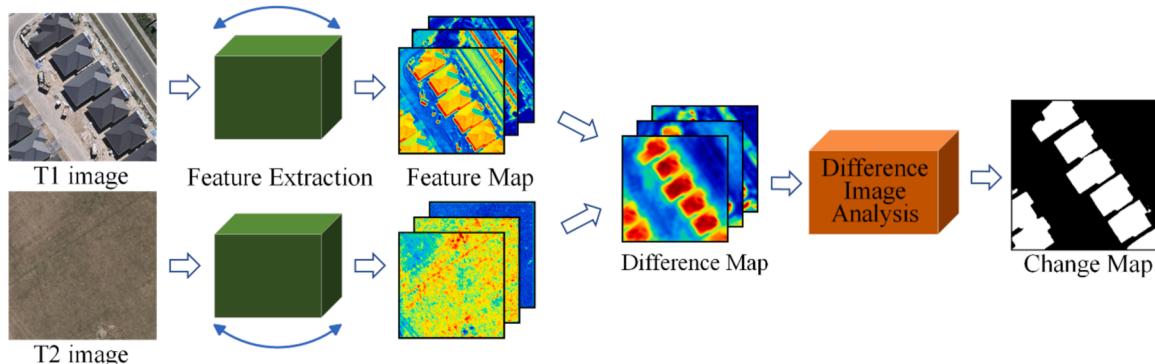


Fig. 5. Framework of FB-DLCD.

generated to facilitate the improvement of FB-DLCD. For example, Tan et al. (2024) proposed a novel unsupervised CD method by combining SAM and contrastive language-image pre-training (CLIP), where difference images were produced by using multi-scale deep features from FastSAM. A piecewise semantic attention scheme was then proposed to refine the difference map by injecting semantic information from CLIP. In addition, similar to unsupervised change detection, zero-shot change detection is emerging by exploiting the strong zero-shot prediction and generalization capabilities of foundation models via prompt engineering (Zheng et al., 2024a).

Notably, FB-DLCD methods can be regarded as the unsupervised deep learning CD methods, which largely adhere to traditional CD approaches and suffer from several challenges such as tedious processing steps, strong reliance on difference images, difficulties in determining thresholds, and susceptibility to error accumulation effect. As a result, the overall performance of these methods is poor, leading to limited practical application values.

2.2. Patch-based deep learning change detection

In contrast to FB-DLCD method, which heavily relies on difference images, pixel-patches or super-pixels are generated in PB-DLCD from either original images or difference maps to serve as network inputs. Subsequently, the change patterns of the central pixels (or super-pixels)

are learned directly by utilizing training samples. Here, the construction of training samples is largely influenced by pseudo-labels, which are usually generated through traditional PLCD or OLCD methods. The overall technical framework of PB-DLCD is illustrated in Fig. 6.

In particular, based on ground truth labeled samples, Daudt et al. (2018b) for the first time proposed two simple yet effective PB-DLCD frameworks, achieving promising CD results. To reduce the uncertainty and noise impact at the local pixel level, Gong et al. (2017) proposed an ensemble strategy based on super-pixels, which utilizes a voting rule to integrate the differential representation results at super-pixel level. In addition, the Spectral-Spatial Joint Learning Network (SSJLN) and the pre-event super-pixel constraint (PreSC) strategy are subsequently introduced to further improve CD performance (Zhang and Lu, 2019; Qing et al., 2022). Lyu et al. (2016) were the first to use Recurrent Neural Networks (RNNs) to simulate spatiotemporal relationships in PB-DLCD. Building upon this work, Mou et al. (2018) further combined CNN with RNN to jointly learn spatiotemporal feature representations.

However, it is challenging to determine the patch extent, and there exists severe redundant information between neighboring patches, which inevitably leads to huge waste of computational resources. Consequently, the scalability and large-scale usage of such methods are limited.

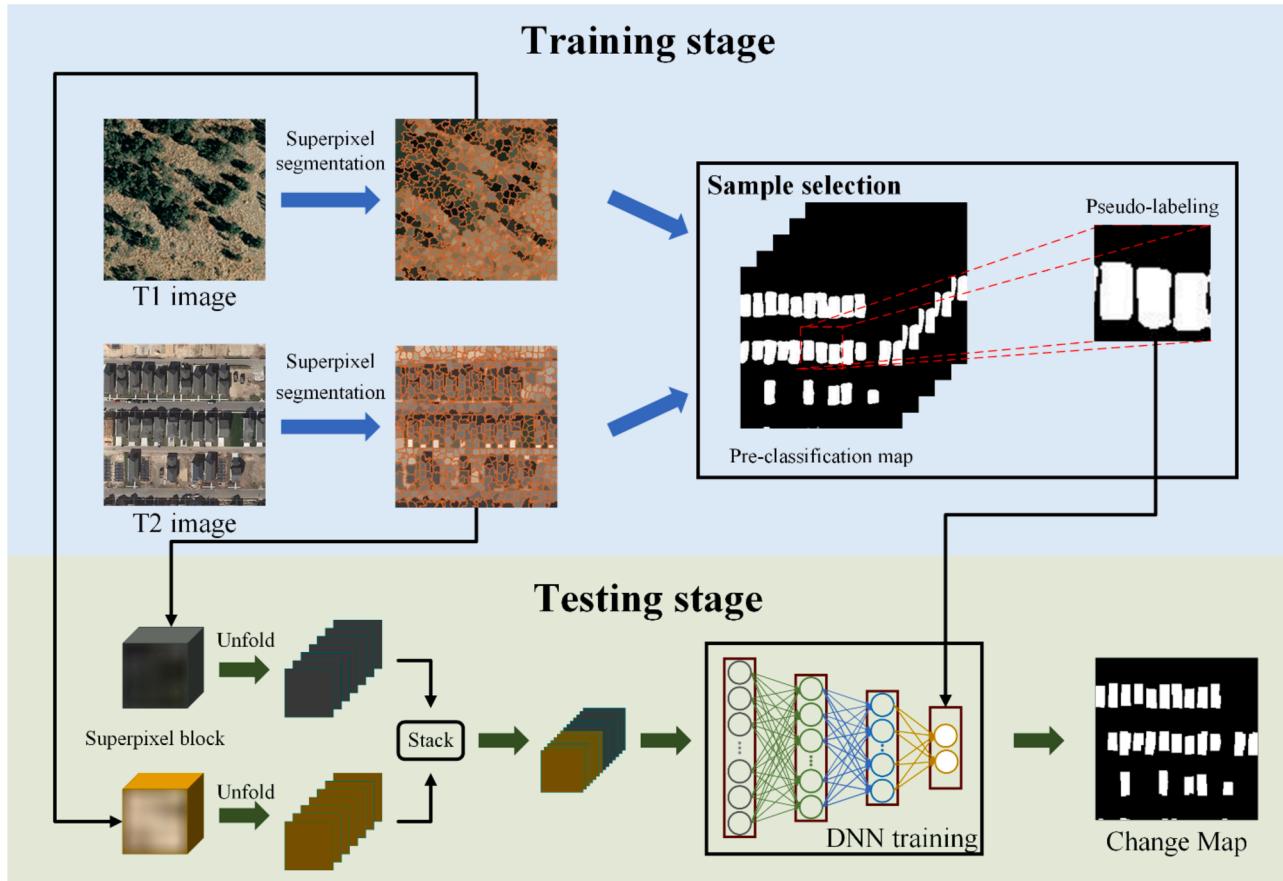


Fig. 6. Framework of PB-DLCD.

2.3. Image-based deep learning change detection

This approach employs bi-temporal image clips as the network input, then change categories of each pixel are learned directly through end-to-end training. Consequently, the accuracy and efficiency of CD are significantly improved, making IB-DLCD the mainstream DLCD

methods. The technical workflow of IB-DLCD is illustrated in Fig. 7.

To be specific, IB-DLCD methods can be categorized into four sub-categories based on the differences in network architecture and feature enhancement modules, namely: 1) CNN-based IB-DLCD methods, 2) Attention Mechanism-based IB-DLCD methods, 3) Transformer-based IB-DLCD methods, and 4) CNN-Transformer Hybrid-based IB-DLCD

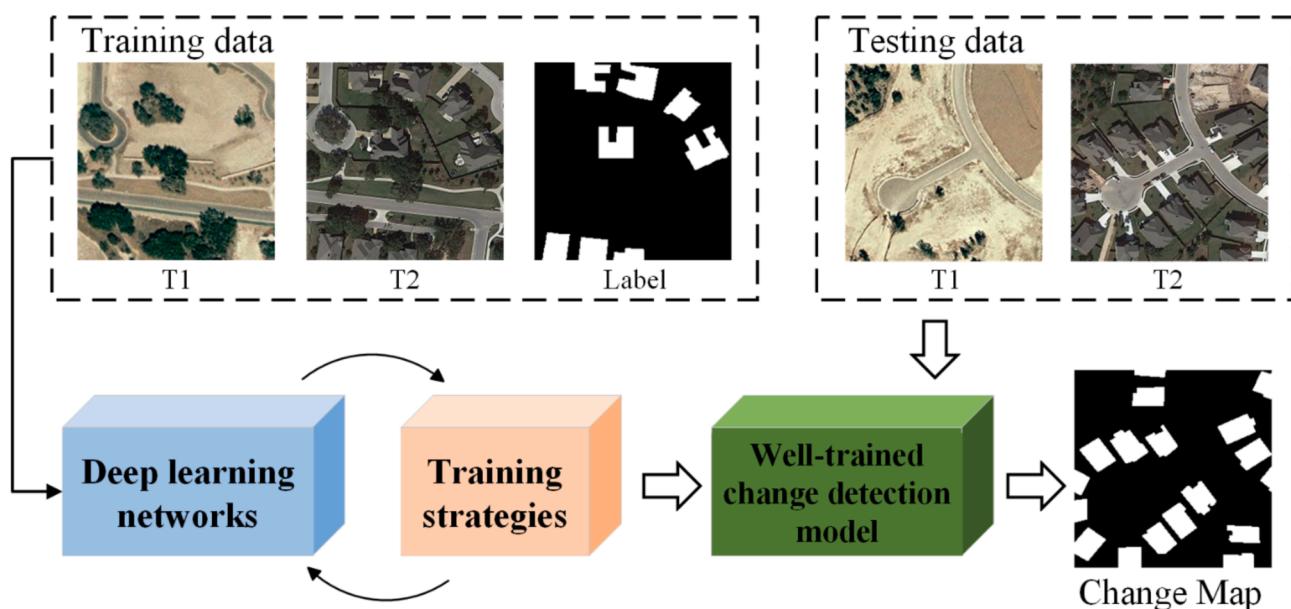


Fig. 7. Framework of IB-DLCD.

methods. In addition, label-efficient learning-based IB-DLCD methods have gained increasing attention due to the difficulty of obtaining sufficient pixel-level labeled samples. To visually demonstrate their influence, a pie chart of different methods is drawn with the publication amount as the index value, as shown in Fig. 8. The results indicate that the CNN-based methods (22 %) and attention mechanism-based methods (30 %) occupy a large proportion in the IB-DLCD methods due to their excellent CD performance. Due to the success of Transformer in vision task, the Transformer-based methods (6 %) and the CNN-Transformer hybrid-based methods (29 %) have experienced rapid development. Specifically, label-efficient learning-based CD methods (13 %) account for a significant percentage, indicating the increasing attention from researchers to deal with the challenge of limited labeled samples.

In addition, a word cloud map is generated according to the keywords from the literature, as shown in Fig. 9. The word cloud reveals that each category of methods includes algorithms with significant impact. Among them, the CNN-based methods represented by FC-EF, the attention mechanism-based methods represented by STANet, the CNN-Transformer hybrid-based methods represented by BIT, and the label-efficient learning-based methods represented by SemiCDNet exhibit higher influence within their respective IB-DLCD method categories.

2.3.1. CNN-based IB-DLCD methods

Inspired by semantic segmentation networks, the majority of existing CNN-based CD networks adopt encoder-decoder architectures similar to UNet (Ronneberger et al., 2015), as shown in Fig. 10. This network efficiently extracts features at different levels through a series of convolutional and pooling layers in the encoder. Subsequently, by employing transpose convolutional layers in the decoder, it performs up-sampling operations to progressively recover the feature map resolutions. Simultaneously, skip connections are included to compensate for the loss of detailed information in the decoder effectively.

Notably, according to the stages of bi-temporal information fusion, these CD network architectures can be divided into three types, namely early fusion architecture (EFA), mid fusion architecture (MFA), and late fusion architecture (LFA), as shown in Fig. 11. Here, the fusion procedure mainly consists of feature concatenation, feature subtraction, and feature summation.

Among them, the EFA (Fig. 11(a)) directly concatenates two input images at the network input to facilitate the interaction of multi-temporal information. In such case, the CD task can be easily transformed into a semantic segmentation task, where large amounts of segmentation strategies and tricks can be adopted to improve CD performance (Daudt et al., 2018a; Daudt et al., 2019; Chen et al., 2022a). For example, Peng et al. (2019) proposed an improved version of UNet++ network, where input images are stacked to feed into the network, multiple side output fusion strategy is further introduced to

capture multi-scale change information. Differently, MFA (Fig. 11(b)) is proposed by introducing a weight-shared Siamese encoder, where bi-temporal images can be processed independently to generate multi-scale deep features through Siamese branches. Subsequently, temporal information fusion can be performed at deep features level, which is more robust and stable against low-level features. In particular, large amounts of pre-trained networks from computer vision can be introduced to serve as the Siamese encoder, which facilitates to flexible extraction and fusion of multi-scale deep features. Consequently, the training efficiency, accuracy, and reliability of CD network is improved significantly, making MFA the mainstream fusion architecture (Fang et al., 2021; Yang et al., 2021b; Ding et al., 2022; Wei et al., 2022; Zhang et al., 2022d; Wang et al., 2023d; Zhu et al., 2023b; Li et al., 2024b). To enhance the reliability of fused features, Lin et al. (2022) treated CD as a video understanding problem and introduced a temporal aggregation module to enhance the discriminative capability of features from different periods. Different from MFA, LFA (Fig. 11(c)) simultaneously employs Siamese encoders and decoders with shared weight, and deep feature fusion is conducted based on two decoder outputs (Chen et al., 2022c; Liang et al., 2023; Zhao et al., 2023b).

Particularly, in EFA, it is flexible to extend the input images from two phases to three phases or even more phases without introducing new network parameters. However, EFA only performs fusion operations at the raw images level, which easily leads to the entanglement of bi-temporal information and failure to consider the difference of deep features. To overcome this limitation, weight-shared Siamese encoders are introduced in MFA, where multi-scale deep bi-temporal features can be extracted separately. However, how to effectively fuse bi-temporal features for change information extraction remains challenging. In contrast, Siamese encoders and decoders with shared-weight are used in LFA, which easily leads to the increase of network complexity, and the representation capability of fused features is reduced due to the shortage of detailed information in the decoder features. Consequently, LFA is less commonly used in CD community.

To facilitate readability, a summary of typical CNN-based IB-DLCD methods is presented in Table 1. It is worth noting that, in CNN-based IB-DLCD, a significant amount of detailed information is lost during encoding stage, resulting in failure to capture local information such as small-sized objects and edges. Additionally, during the decoding process, the reliability of the fused features is reduced due to undifferentiated fusion of different levels of deep features.

2.3.2. Attention-based IB-DLCD methods

Attention mechanism is capable of enhancing the focus on important areas while disregarding irrelevant information by updating attention weights, whereby improving the feature localization and representation capability. As an effective and flexible plug-in module, it has been extensively introduced to CD networks. The illustration of attention

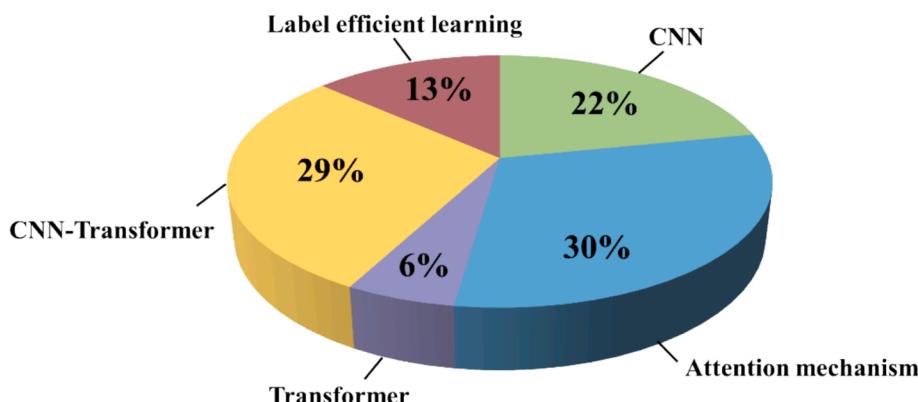


Fig. 8. Pie chart of different IB-DLCD methods.

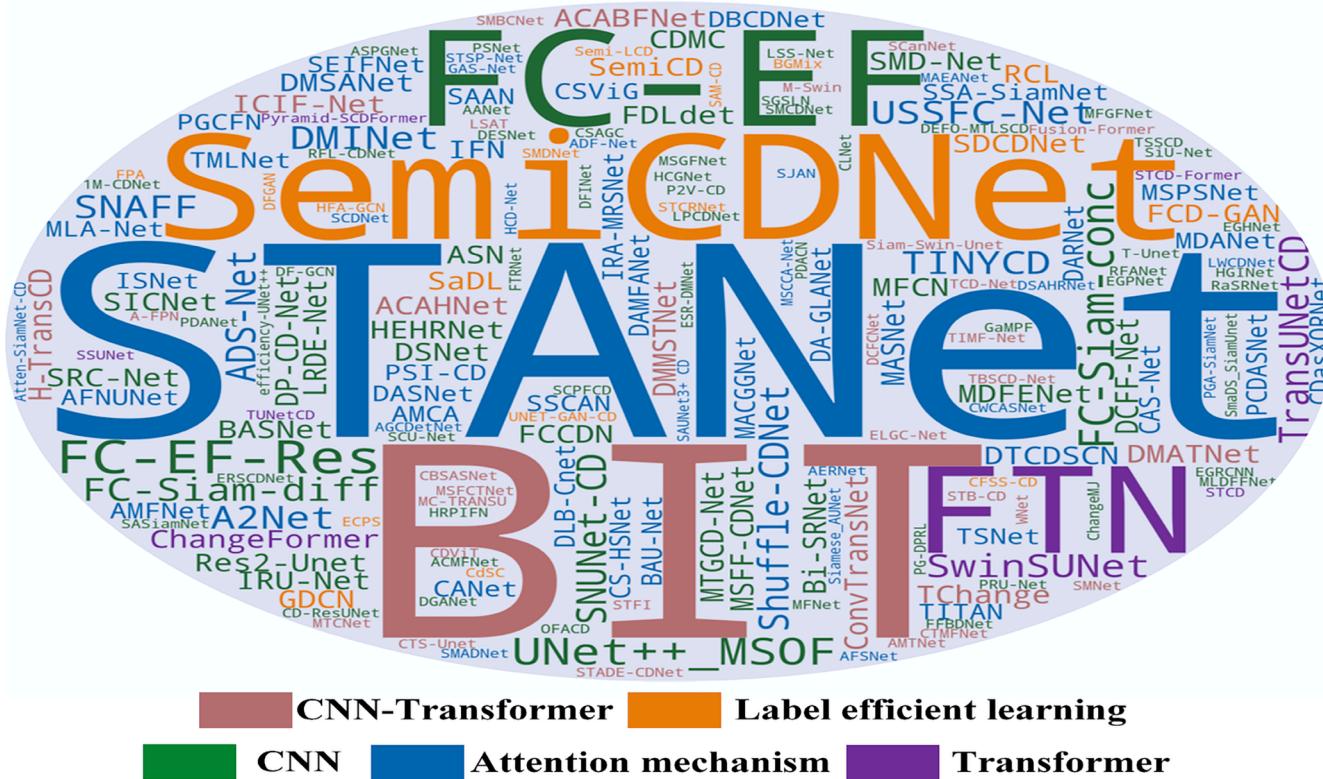


Fig. 9. Keyword cloud map.

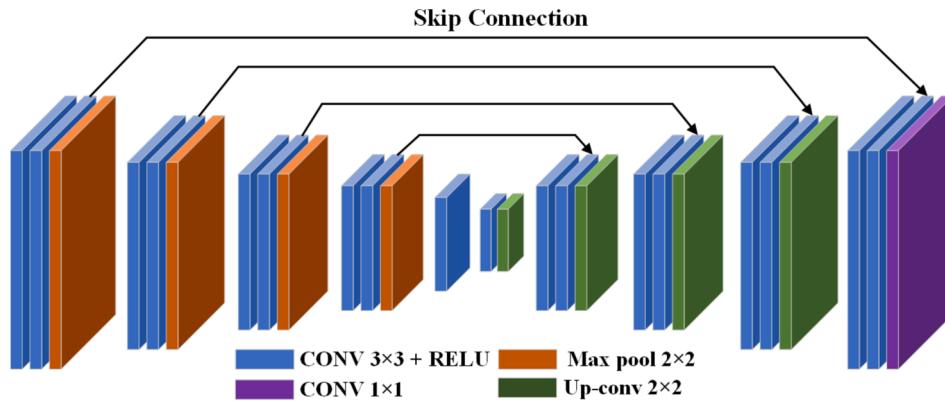


Fig. 10. Illustration of UNet.

mechanism is shown in Fig. 12, where Query, Key and Value are generated through a convolutional layer based on the input features. First, a similarity matrix is calculated using Query and Key matrix, resulting in attention scores S_i . Then, a softmax normalization function is adopted to generate attention weights a_i . Finally, based on these weights, the features (Values) are weighted and summed to produce final refined features. In particular, channel and spatial attention are two commonly used attention mechanisms in CD task. Among them, channel attention is used to calculate weights for each channel, thus guiding the model to focus on important channel information (Song and Jiang, 2021; Peng et al., 2023). On the other hand, spatial attention is used to weight or select the features in the spatial dimension, whereby constraining the model to focus on the regions of interest (Li et al., 2023d). Additionally, to exploit the complementary advantages of the two attention mechanism, an integrated attention mechanism is widely used to improve CD performance (Liu et al., 2020; Wang et al., 2021b; Cheng et al., 2022; He et al., 2023b; Ning et al., 2024).

For example, to enhance the feature representation capabilities, Zhang et al. (2020) cascaded spatial attention and channel attention modules to fuse features at different decoder layers, thereby enhancing the representation capability of the fused features. However, global contextual information in the feature space is ignored, leading to sub-optimal CD performance. To overcome the drawback, Chen and Shi (2020) proposed a Spatial-temporal Attention Neural Network (STA-Net), which aims to learn spatial-temporal dependencies of different sub-regions. Through an adaptive spatial and channel attention mechanism, Wang et al. (2021a) proposed to incorporate feature maps from different levels into different branches of a deep supervision network, thus effectively capturing change information at different scales. In addition, multi-scale attention modules (Guo et al., 2022; Zhang et al., 2022c; Lei et al., 2023a) and global attention modules (Feng et al., 2023; Peng et al., 2023; Zhang et al., 2023d) were further proposed to capture long-range multi-scale dependencies and global semantic information. It is worth noting that the introduction of the aforementioned attention

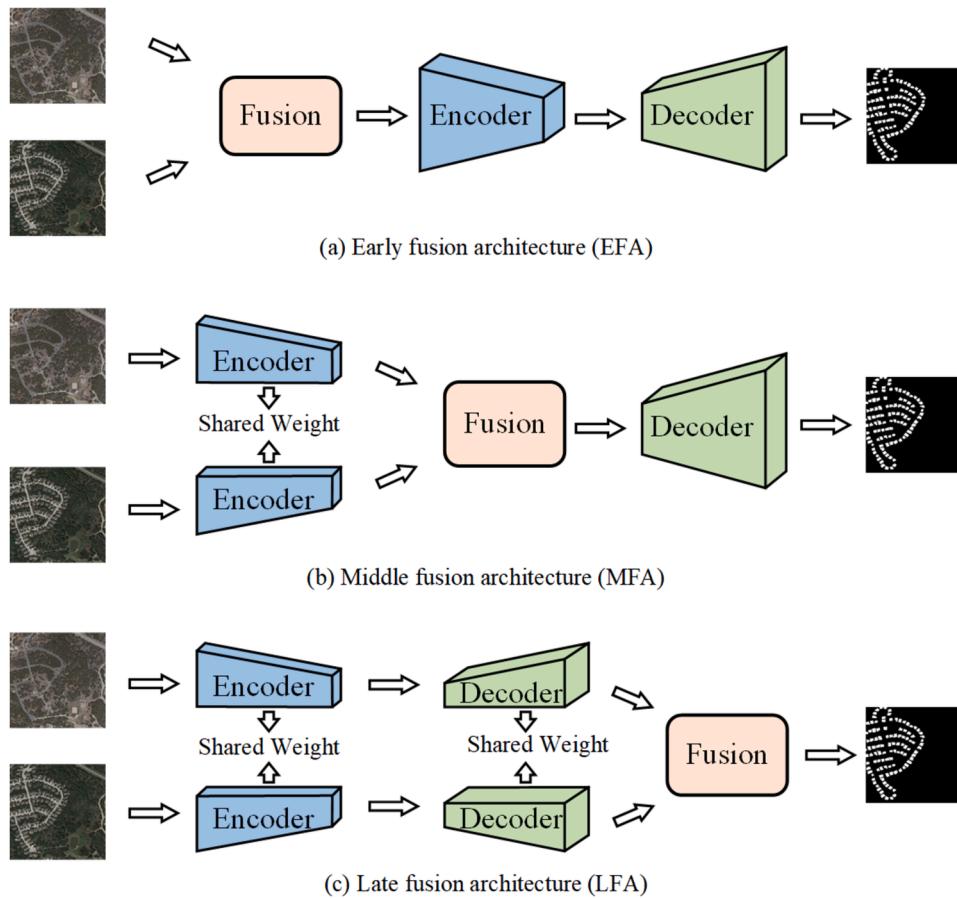


Fig. 11. Illustration of three fusion strategies.

modules often leads to a significant increase in model parameters and model complexity, which in turn results in large computational costs.

2.3.3. Transformer-based IB-DLCD methods

Compared to CNN networks with limited receptive fields, Transformer networks have a global receptive field (Vaswani et al., 2017), allowing them to effectively capture global contexts. As presented in Fig. 13, Transformer architecture includes basic modules of Multi-Head Self-Attention (MSA) and Multi-Layer Perceptron (MLP), where layer normalization (LN) layer and residual connections are employed before and after the MSA and MLP modules for better information flow. Due to its powerful feature representation and contextual modeling capabilities, Transformer model has been widely applied to improve CD networks in various aspects such as global information enhancement and multi-scale information processing. Among them, three kinds of architectures are mostly investigated, namely pyramid architectures (Yan et al., 2022; Zhu et al., 2023a), dual-branch architectures (Yuan et al., 2022), and UNet architectures (Li et al., 2022a; Liang et al., 2022). For instance, based on Swin Transformer backbone, Zhang et al. (2022b) designed the first pure Transformer CD model termed SwinSUNet, where a merging module is introduced to compensate for the loss of detailed information. To capture temporal correlations and dependencies, Wang et al. (2022e) designed a Temporal Transformer model by using a residual structure with three cross-attention mechanism units. In addition, multi-scale feature enhancement can be achieved by introducing multiple Transformer models to perform information propagation and contextual modeling at different levels (Bandara and Patel, 2022b; Yan et al., 2023; Chen et al., 2024a; Zhang et al., 2024a). For example, instead of applying standard self-attention unit, which struggles to capture inductive biases when trained from scratch, Norman et al. (2024b) proposed a shuffled sparse-attention module to focus on the

sparse changed areas. A change-enhanced feature fusion module (CEFF) is further included to facilitate bi-temporal feature fusion by enhancing the semantic changes while suppressing the noisy areas.

In particular, compared to convolutional networks, CD models based on Transformer modules heavily rely on self-attention mechanisms, leading to a significant increase in model complexity and parameter size. This inevitably leads to huge training costs and higher dependence on labeled samples. Notably, more recently, Mamba architectures based on state space models (SSM) have achieved great success in vision task due to their stronger global information modeling ability, lower computational complexity and higher scalability over Transformer architectures. In such context, Mamba-based CD models have been gaining increasing attention (Chen et al., 2024c; Paranjape et al., 2024). For instance, Zhang et al. (2024b) proposed a novel CD method termed CDMamba, where a scaled residual ConvMamba block was employed to enhance local information as well as capturing global receptive field. Notably, pixel-wise prediction and classification are mostly employed in existing Transformer-based CD architectures, which, however, easily leads to imprecise boundaries and incomplete object delineation at various scenes. To overcome this limitation, some pioneer work seeks to tackle CD task from the perspectives of mask prediction and classification (Ma et al., 2024; Yu et al., 2024a), leading to improved CD performance. For instance, Yu et al. (2024a) proposed MaskCD, where a cross-level change representation perceiver (CLCRP) was introduced to learn multi-scale change aware features in the encoding stage, while a masked-attention-based detection transformers (MA-DETR) decoder was proposed to locate and identify the changed objects accurately.

2.3.4. CNN-transformer hybrid-based IB-DLCD methods

Traditional CNN models possess powerful local modeling capabilities but fail to capture global context correlations due to limited receptive

Table 1
A summary of typical CNN-based IB-DLCD methods.

Method	Source	Highlights
Early fusion:		
FC-EF (Daudt et al., 2018a)	ICIP2018	concatenates the bi-temporal images along the channel dimension to form the network input for dense prediction
FC-EF-Res (Daudt et al., 2019)	CVIU2019	proposes a multi-task learning network, which is capable of achieving change detection and land cover mapping simultaneously
UNet++_MSOF (Peng et al., 2019)	RS2019	proposes an improved U-Net++ network and further introduces a multi-sided output fusion strategy to capture multi-scale change information
Res2-Unet (Chen et al., 2022a)	JSTARS2022	proposes a novel deep architecture that combines fine-grained multi-scale learning and a hierarchical convolutional group structure, a boundary loss function is further incorporated.
Middle fusion:		
FC-Siam-conc (Daudt et al., 2018a)	ICIP2018	fuses bi-temporal deep features through concatenation operation
FC-Siam-diff (Daudt et al., 2018a)	ICIP2018	fuses bi-temporal deep features through differential operation
SNUNet-CD (Fang et al., 2021)	GRSL2021	utilizes dense skip connections and introduces an Integrated Channel Attention Module (ICAM) for deep supervision
BASNet (Wei et al., 2022)	GRSL2021	proposes a boundary-aware Siamese network (BASNet) by introducing a multiscale paired fusion module (MPFM), a location guidance module (LGM), and a multilevel feature aggregation module (MFAM) for enhanced boundary perception
DSNet (Yang et al., 2021b)	RS2021	presents an end-to-end Siamese change detection network by considering the change detection problem as a binary semantic segmentation task
SMD-Net (Zhang et al., 2022d)	RS2022	designs a Siamese residual multi-kernel pooling module (SRMP) and a feature difference module (FDM) to extract multiscale and multi-depth differential features
Bi-SRNet (Ding et al., 2022)	TGRS2022	improves single-temporal semantic representation and cross-temporal semantic correlation by using two types of semantic reasoning blocks
P2V-CD (Lin et al., 2022)	TIP2022	considers change detection as a video understanding problem and introduces a temporal aggregation module to enhance the discriminative ability of features across different periods
DESNet (Wang et al., 2023d)	GRSL2023	embeds a Difference Enhancement (DE) Module and a multiscale parallel sampling spatial-spectral nonlocal (SSN) to enable semantic change detection (SCD) in remote sensing images
EGPNet (Zhu et al., 2023b)	JSTARS2023	by parallel coding framework and edge guidance, bi-temporal difference features are extracted at the same time for fusion
Late fusion:		
FCCDN (Chen et al., 2022c)	JPRS2022	proposes a strategy based on self-supervised learning to constrain the extraction and fusion of bi-temporal features
RaSRNet (Liang et al., 2023)	TIM2023	designs a multilevel semantic reasoning encoder-decoder (ED) structure and incorporates a relation-aware (Ra) module to perceive contextual semantic information, enabling accurate localization of changed objects
SGSLN (Zhao et al., 2023b)	TGRS2023	designs a dual encoder-decoder structure with feature exchange and fuses bi-temporal features at the decision level

field. In contrast, Transformer models have a global receptive field to effectively capture global dependencies. However, they tend to overlook local features and come with a much higher computational cost. It is therefore natural to combine the advantages of both architectures to further improve CD performance and efficiency (Chen et al., 2023b; Zhang et al., 2023b).

Based on differences between the integrated encoder and decoder, five integration styles can be obtained, namely sequential integration (SI), multi-path CNN integration (MCI), parallel integration (PI), unilateral mixing (UM) and bilateral mixing (BM), as illustrated in Fig. 14. Specially, in SI, CNN and Transformer are simply integrated in a sequential manner (as illustrated in Fig. 14(a)), where a CNN network is used in the encoder to capture local information, while a Transformer network is used in the decoder to include global context (Liu et al., 2022b; Wang et al., 2022b; Wang et al., 2022d; Jiang et al., 2023; Gao et al., 2024; Yu et al., 2024c). For instance, Chen et al. (2021b) designed a network called Bidirectional Image Transformation (BIT), which utilizes semantic tokens to perform global context modeling on the bi-temporal features extracted by a ResNet18 pre-trained network. However, the encoder output only consists of coarse-grained semantic features, making them less sensitive to small-scale changed regions. To address this issue, MCI is utilized to extract both coarse and fine-grained features (Ding et al., 2024a), as illustrated in Fig. 14(b). For example, Song et al. (2022b) proposed a dual feature extraction approach, where a simple CNN was utilized to extract coarse-grained features, while another CNN based on progressive sampling was used to extract fine-grained features. However, only CNN is used in the encoder, leading to feature maps with poor global representation capability. To overcome this limitation, PI is proposed by combining CNN and Transformer in the encoder (Feng et al., 2022; Song et al., 2022a), as shown in Fig. 14(c). In such context, Tang et al. (2023) combined Siamese CNN and Siamese Transformer in parallel to extract features from both local and global perspectives. However, two limitations exist for such a simple strategy. First, the network complexity will increase sharply due to include separate CNN and Transformer networks in the encoder. Second, complex fusion algorithms have to be carefully designed to effectively fuse two different kinds of features from CNN and Transformer backbones. To address the issues, another integrated paradigm of UM is proposed (as illustrated in Fig. 14(d)), where a novel encoder is constructed by embedding CNN and Transformer units into a powerful feature extractor, thus directly capturing global-local feature representations without further fusion steps (Niu et al., 2023; Li et al., 2023b; Zhao et al., 2023b; Shi et al., 2023b). For instance, Li et al. (2023b) generated multi-scale global-local feature representations by constructing ConvTrans modules that integrate the characteristics of both CNN and Transformer. In addition, to improve the interaction between CNN and Transformer features, a hybrid approach of BM (as illustrated in Fig. 14(e)) is employed by combining CNN and Transformer in both the encoding and decoding stages (Ke and Zhang, 2022; Deng et al., 2023; Zhang et al., 2023c; Zhang et al., 2023e).

For the benefit of readability, a summary of typical CNN-Transformer hybrid-based IB-DLCD methods is presented in Table 2. Overall, the CNN-Transformer hybrid-based IB-DLCD methods are capable of combining the advantages of powerful local feature from CNN and global contextual modelling from Transformers, thus effectively improving CD performance as well as reducing model complexity. However, how to effectively integrate these two heterogeneous architectures remains challenging.

2.3.5. Label-efficient learning-based IB-DLCD methods

Although existing IB-DLCD methods have achieved satisfactory CD performance, they tends to come with complex network structures, massive model parameters, and heavy reliance on labeled samples. As a result, it is difficult to apply such CD algorithms in resource-constrained scenarios. To overcome the limitation, a series of label-efficient learning-based CD algorithms have been developed to reduce the

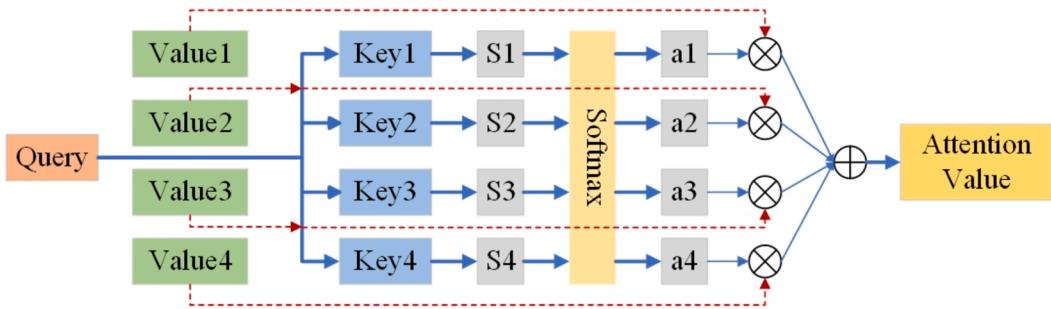


Fig. 12. Illustration of attention mechanism.

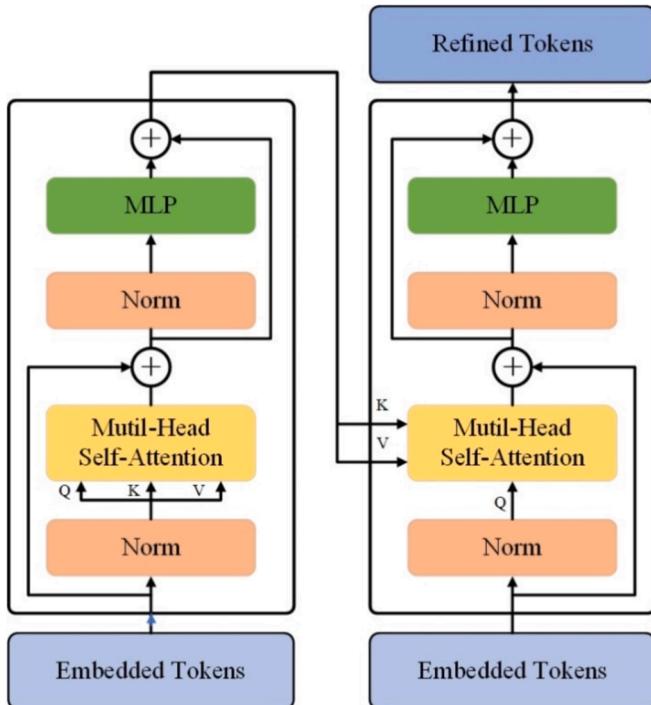


Fig. 13. Illustration of Transformer architecture.

reliance on labeled samples and improve the robustness and generalizability of CD algorithms. In particular, by exploiting information from large-scale unlabeled samples and limited labeled samples, semi-supervised CD method is capable of improving the generalization performance of CD models significantly. In the literature, many efforts are made on semi-supervised CD task. These methods are mainly based on 1) **adversarial learning**, which is used to generate new samples (Gong et al., 2019) or enforce feature distribution constraints (Peng et al., 2020); 2) **consistency learning**, which works on the basis that the output remain stable when imposing data perturbations, feature perturbations or model perturbations (Bandara and Patel, 2022a; Chen et al., 2022b; Ding et al., 2023; Han et al., 2024a); and 3) **pseudo learning**, where pseudo-labels can be generated in an offline manner (Chen et al., 2021a; Wang et al., 2022c; Wang et al., 2023c; Yuan et al., 2024) or online manner (Sun et al., 2022a; Sun et al., 2023; Zhang et al., 2023f; Zou and Wang, 2023; Yang et al., 2024). Note that, adversarial learning methods often suffer from training instability, which usually achieves inferior performance against consistency learning or pseudo learning. In addition, consistency constraints and pseudo label learning are often combined to achieve better performance (Zhang et al., 2023g).

Apart from semi-supervised learning, active learning (AL) and weakly-supervised learning (WSL) are also widely used. In AL, the most informative samples are generated for manual annotation, leading to

effective labeled data augmentation with small annotation cost. Hence, the core of such methods is to estimate model uncertainty so as to pick the uncertain or informative samples (Zhang et al., 2018; Růžička et al., 2020). Differently, in WSL-based methods, easier-to-obtain sparse labels such as point-level labels (Fang et al., 2023), patch-level labels (Li et al., 2023c; Wang et al., 2023a), or image-level labels (Huang et al., 2023; Wu et al., 2023; Zhao et al., 2023c), are first generated with prior information, which are then used to produce pixel-level dense predictions. Among them, image-level labels are mostly used by producing class activation maps (CAMs), which serve as dense pixel-level labels (Dai et al., 2023b; Zhao et al., 2024b). In addition, by integrating high-resolution images, incomplete labels such as low-resolution coarse labels can be used to generate fine labels (Zheng et al., 2021), which facilitates to capture semantic change patterns.

For the benefit of clear comparison, a summary of existing label-efficient learning based IB-DLCD methods is systematically outlined and presented in Table 3.

3. Benchmark analysis of ORSCD datasets

From the perspective of semantic types, ORSCD datasets can be divided into two major categories: binary change detection (BCD) datasets and semantic change detection (SCD) datasets. In BCD datasets, the focus is solely on the location where the changes occur, while SCD datasets typically include “from-to” change information, indicating both the locations and attributes of the changes. This section provides a comprehensive summary of ORSCD datasets in the past few years, which greatly facilitates to select datasets according to specific needs.

Fig. 15 presents the publication trends of ORSCD dataset since 2017. It can be observed that the number of CD datasets have been steadily increasing over the past few years, where it reaches a peak in 2023, indicating the significant role of CD in remote sensing areas. Particularly, BCD datasets have shown an upward trend followed by a decline, while SCD datasets have presented a stable upward trend. This indicates the fact that, compared with BCD, SCD is gaining an increasing attention due to its capability of providing more comprehensive change information. Additionally, based on the targets of interest, existing optical remote sensing image datasets for BCD and SCD are comprehensively summarized in Table 4 and Table 5, respectively. In particular, detailed information is summarized such as data source, imaging time, image size, image resolution and download weblinks. One can observe that, in terms of data source, apart from related literatures, plenty of datasets are from CD competitions or events, demonstrating the significance of CD in both academic and industrial community. In addition, in both BCD and SCD, the majority of the datasets focus on typical targets such as buildings and land cover, while a growing focus is observed on natural resource elements such as farmland, agricultural fields, water bodies, mining sites, as well as natural disasters such as earthquake, flooding and landslide. This further demonstrates the significant role of CD techniques in the dynamic monitoring of natural resources and natural disasters. To contribute to CD community, we will continuously update the datasets via <https://github.com/daifeng2016/Awesome-Optical-Re>

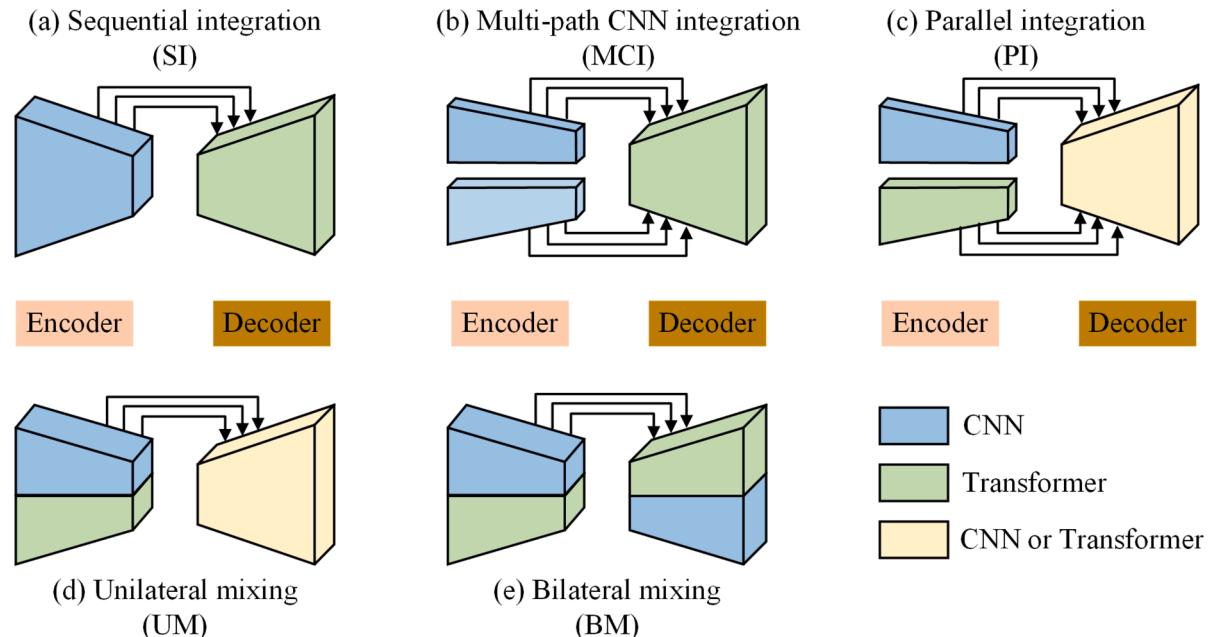


Fig. 14. Illustration of integrated framework of CNN-Transformer.

mote-Sensing-Datasets-and-Methods.

4. Comprehensive analysis of typical change detection algorithms

In this section, three typical BCD datasets (WHU-CD, LEVIR-CD, SYSU-CD) and two typical SCD datasets (SECOND and Landsat-SCD) are used as benchmarks to conduct a comprehensive comparative analysis of mainstream CD algorithms in terms of accuracy, efficiency, and model complexity. These analysis aims to provide important guidelines for practical deployment and application of CD models.

4.1. Quantitative evaluation metrics

Generally, existing accuracy metrics for BCD include Precision (P), Recall (R), Intersection over Union (IoU), Overall Accuracy (OA), F1-score and Kappa coefficient. The definition of these metrics are as follows:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (3)$$

$$OA = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

$$F1 = \frac{2}{Recall^{-1} + Precision^{-1}} \quad (5)$$

$$Kappa = \frac{OA - PRE}{1 - PRE} \quad (6)$$

$$PRE = \frac{(TP + FN) \times (TP + FP) + (TN + FN) \times (TN + FP)}{(TP + TN + FP + FN)^2} \quad (7)$$

where TP and TN denote the count of pixels accurately identified as changed and unchanged, respectively. FP indicates the count of un-

changed pixels incorrectly classified as changed, and FN denotes the count of changed pixels incorrectly classified as unchanged. Precision represents the proportion of correctly predicted positive samples among all predicted positive samples. Recall represents the proportion of correctly predicted positive samples among all actual positive samples. Different from single evaluation metrics, OA, F1, IoU and Kappa are overall evaluation metrics, which are capable of indicating overall CD performance. Specifically, F1 is the harmonic results of precision and recall, providing a balanced measure of both precision and recall; IoU measures the ratio of the intersection of the predicted and ground truth regions to their union; Kappa indicates the difference between observed agreement and expected agreement; while OA represents the proportion of correctly predicted samples (both positive and negative) among all the samples.

In particular, traditional BCD accuracy evaluation metrics are not suitable for SCD due to the severe class imbalance issue (Peng et al., 2021). To overcome the limitations, mean Intersection over Union (mIoU) and Separate Kappa (Sek) coefficient (Yang et al., 2021a) are proposed to consider semantic information. To be specific, mIoU and Sek are capable of evaluating SCD results from the perspective of two-class change and multi-class change, respectively. The detailed definition of mIoU is as follows:

$$IoU_1 = \frac{TN}{TP + FN + FP} \quad (8)$$

$$IoU_2 = \frac{TP}{TP + FN + FP} \quad (9)$$

$$mIoU = \frac{IoU_1 + IoU_2}{2} \quad (10)$$

where IoU_1 and IoU_2 represent the IoU for the unchanged and changed pixels, respectively. Based on IoU_1 and IoU_2 , mIoU is defined by calculating their mean value, which effectively considers the CD performance of both the changed class and unchanged class. To further consider multi-class change information, a confusion matrix of M_{ij} ($0 \leq i, j \leq C - 1$) is calculated, where C denotes the number semantic change label categories. In particular, the unchanged category is excluded by setting $M_{i0} = M_{0j} = 0$. Then, a Kappa coefficient is calculated based on filtered confusion matrix M_{ij} . Finally, Sek score is obtained by combining

Table 2

A summary of typical CNN-Transformer hybrid-based IB-DLCD methods.

Category	Method	Source	Highlights
Sequential integration (SI)	BIT (Chen et al., 2021b)	TGRS2021	utilizes semantic tokens to perform global context modeling on the bi-temporal features extracted by ResNet18 pre-trained network
	UVACD (Wang et al., 2022b)	RS2022	combines CNN and Transformer to achieve feature fusion and temporal information interaction
	MTCNet (Wang et al., 2022d)	JSTARS2022	proposes a Multi-Scale Transformer with CBAM incorporated in both pre-processing and post-processing stages
	MSCANet (Liu et al., 2022b)	JSTARS2022	designs a three-branch Transformer encoding-decoding block to aggregate hierarchical features
	GCFFormer (Yu et al., 2024c)	TGRS2024	proposes a Context-Aware Relative Position Encoding (CARPE) scheme, which encodes positional information by comparing the relative position relationships between different objects
Multi-path CNN integration (MCI)	DMATNet (Song et al., 2022b)	TGRS2022	uses two distinct CNNs to extract coarse and fine-grained features separately
	SCanNet (Ding et al., 2024a)	TGRS2024	introduces the Semantic Change Transformer (SCanFormer) to learn semantic transformations between bi-temporal images and further enhances the performance of SCD by incorporating spatio-temporal constraints
Parallel integration (PI)	ICIF-Net (Feng et al., 2022)	TGRS2022	designs a scale-intra and scale-inter feature fusion module and implements feature extraction and fusion at different resolutions
	ACABFNet (Song et al., 2022a)	JSTARS2022	applies global attention along the height and width dimensions of the image separately and uses a cross-attention mechanism to fuse the global feature information from both dimensions.
	WNet (Tang et al., 2023)	TGRS2023	utilizing both Siamese CNN and Siamese Transformer to extract bi-temporal features from both local and global perspectives
Unilateral mixing (UM)	SUT (Chen et al., 2023b)	RS2023	utilizes a single layer CNN and a three-layer structure that integrates CNN and Transformer with full-scale skip connections
	CTD-Former (Zhang et al., 2023b)	TGRS2023	proposes Cross-Temporal Difference (CTD) attention and Consistency-Perception Blocks (CPBs).
	SMNet (Niu et al., 2023)	RS2023	integrates global and local information through a backbone composed of pre-activated residual blocks (PR) and transformation blocks known as the PRTB backbone.
	ConvTransNet (Li et al., 2023b)	TGRS2023	achieves global-local feature extraction and fusion through parallel ConvTrans blocks
	GateFormer (Li et al., 2023a)	JSTARS2023	designs a Gate Attention mechanism that specifically attends to the change regions
	GeoFormer (Zhao et al., 2023a)	TGRS2023	utilizes geometric prior knowledge to improve detection performance
	DAHT-Net (Shi et al., 2023a)	ACCESS2023	employs an alternating usage of Swin Transformer blocks and Deformable Attention Transformers
Bilateral mixing (BM)	H-TransCD (Ke and Zhang, 2022)	JPRS2022	improves performance by introducing Transformer structures and hybrid multi-scale representations
	ACAHNet (Zhang et al., 2023e)	TGRS2023	adopts a mid-level fusion strategy and proposes an Asymmetric Multiheaded Cross Attention (AMCA) module
	DMMSTNet (Zhang et al., 2023c)	RS2023	proposes a Siamese network structure with Context Transformer Module (CoT) module
	TChange (Deng et al., 2023)	RS2023	proposes a Change MSA module and a Inter-scale Transformer Module (ISTM) to enable the interaction of global information and local information across different scales

Kappa coefficient and IoU₂:

$$Sek = Kappa^* e^{IoU_2 - 1} \quad (11)$$

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \quad (12)$$

$$p_0 = \frac{\sum_{i=0}^{C-1} M_{ii}}{\sum_{i=0}^{C-1} \sum_{j=0}^{C-1} M_{ij}} \quad (0 \leq i \leq C-1, 0 \leq j \leq C-1) \quad (13)$$

$$p_e = \frac{\sum_{i,j=0}^{C-1} M_{i+} * M_{+j}}{\left(\sum_{i=0}^{C-1} \sum_{j=0}^{C-1} M_{ij} \right)^2} \quad (14)$$

where M_{i+} represents the row sum of the confusion matrix M_{ij} , and M_{+j} represents the column sum of the confusion matrix M_{ij} . Note that, to address the class imbalance issue caused by unchanged category in SCD, only changed categories is considered to calculate the Sek score.

4.2. Algorithm analysis for BCD

To quantitatively compare the performance of different BCD methods, five metrics of Precision (P), Recall (R), Overall Accuracy

(OA), F1-score (F1) and Intersection over Union (IoU) for 22 typical IB-DLCD methods are summarized and presented for WHU-CD, LEVIR-CD and SYSU-CD in Table 6. These metrics are derived by three ways, namely: 1) selected directly from the published literature, 2) reproduced using the released model, and 3) reproduced using the released code. Note that WHU-CD dataset focuses on building changes before and after an earthquake, LEVIR-CD dataset covers building changes of various types such as villa residences, tall apartments, small garages and large warehouses. Differently, SYSU-CD dataset contains more diverse changed types, including urban and suburban building change, land change, vegetation change, road expansion, and sea construction. Particularly, average metrics across the three different datasets are also calculated for benefit of overall performance evaluation.

It can be observed that simple CNN-based methods achieve lower accuracy caused by negligence of semantic differences from different levels of feature maps. In contrast, in UNet++_MSOF and SNUNet-CD, by introducing dense connection operations, the semantic differences between the encoder and decoder are suppressed, thus enhancing the fusion feature representation capability. As a result, the CD accuracy is significantly improved. It is worth noting that feature discriminability can be enhanced by introducing attention mechanisms such as spatial and channel attention modules, leading to improved CD performance.

Table 3

A summary of typical Label-efficient learning-based IB-DLCD methods.

Method	Category/Examples	Pros/Cons
Semi-supervised	Adversarial learning GDCN (Gong et al., 2019) and SemiCDNet (Peng et al., 2020)	Pros: Capable of generating new samples or enforcing distribution constraints Cons: GAN suffers from training instability
	Consistency constraints SaDL (Chen et al., 2022b), SemiCD (Bandara and Patel, 2022a), C2F-SemiCD (Han et al., 2024a), Semi-LCD (Ding et al., 2023) and FPA (Zhang et al., 2023g)	Pros: High flexibility and effectiveness, easily to introduce data perturbations, feature perturbations or model perturbations Cons: The degree of constraints should be carefully controlled
	Pseudo-label learning ST-RCL (Zhang et al., 2023f), SemiSANet (Sun et al., 2022a), STCRNet (Wang et al., 2023c), RCL (Wang et al., 2022c), DCSS (Yuan et al., 2024), IAug_CDNet (Chen et al., 2021a), SemiBuildingChange (Sun et al., 2023), SSCD (Zou and Wang, 2023) and ECPS (Yang et al., 2024)	Pros: Pseudo label can be generated flexibly in offline manner or online manner, easily to be combined with consistency learning Cons: High training cost for pseudo labels generated in offline mode
Active learning	DALCD (Růžička et al., 2020), CFSS-CD (Zhang et al., 2018)	Pros: Capable of reducing annotation cost by picking the most informative samples Cons: Human annotation is still needed during training
Weakly-supervised	Point-level labels CARGNet (Fang et al., 2023)	Pros: Easily to obtain Cons: Difficult model training due to sparse supervised signal
	Patch-level labels MS-Former (Li et al., 2023c) and SDCDNet (Wang et al., 2023a)	Pros: Easily to obtain Cons: The weak label quality is difficult to control
	Image-level labels BGMix (Huang et al., 2023), FCD-GAN (Wu et al., 2023), TransWCD-DL (Zhao et al., 2023c), WSCD (Dai et al., 2023b), and WSLCD (Zhao et al., 2024b)	Pros: Easily to obtain, CAMs can be used to produce high-quality pseudo labels Cons: Difficulty in generating high-quality CAMs
	Low-resolution labels RFWSCD (Zheng et al., 2021)	Pros: Capable of exploiting coarse label information directly without annotation cost Cons: Difficult to obtain, cumbersome training steps

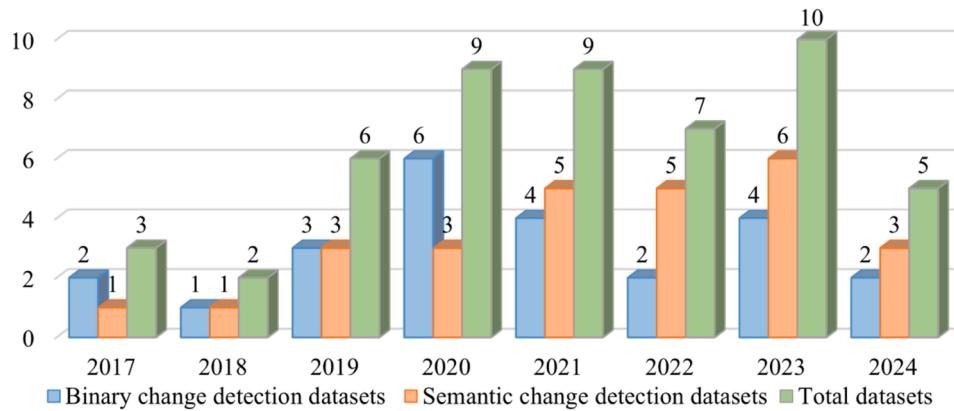


Fig. 15. Publication trends of ORSCD datasets.

Among them, A2Net achieves the best average performance, USSFC-Net yields the second-best average performance. Unlike convolutional networks, Transformer networks are capable of utilizing self-attention mechanisms to learn global contextual relationships, thus effectively enhancing feature representation capabilities. It can be observed that FTN, SwinSUNet, ChangeFormer, and BiFA, which are based on the Transformer architecture, achieve high accuracy metrics. Among them, FTN obtains the best average performance. Furthermore, the advantages of both convolutional networks and Transformer networks can be effectively integrated by using CNN-Transformer hybrid CD methods. It significantly facilitates to extract local detailed features extraction and capture global contextual relationships, leading to improved CD performance with relatively lower computational costs. Among them, RCTNet achieved the best average performance, VcT obtained the second-best average performance. In particular, the performance of different BCD algorithms may vary on specific datasets. For instance, A2Net and RCTNet observe superior performance against BiFA and VcT on WHU-CD and LEVIR-CD datasets centered on building changes but inferior performance on SYSU-CD dataset with more diverse change categories.

4.3. Algorithm analysis for SCD

Table 7 shows the comparison of quantitative results of different SCD algorithms in terms of Sek and mIoU on the two SCD datasets of SECOND and Landsat-SCD, where SECOND consists of six changed types of non-vegetated ground surface, tree, low vegetation, water, buildings, and playgrounds, while Landsat-SCD consists of four changed types of farmland, desert, building, and water. The metrics in the table are derived from the published literature. It is worth noting that HRSCD.str4 first introduces a three-branch multi-task framework for SCD. However, the interactions between different branches are insufficient. In addition, the simple UNet backbone fails to capture enough semantic features. As a result, HRSCD.str4 achieves poor quantitative metrics. Through improving individual-temporal semantic information and cross-temporal semantic interaction, Bi-SRNet observes an obvious gain in accuracy metrics. In SCDNet, by adopting ResNet34 backbone and effective feature enhancement strategies including attention mechanism, deep supervision, and multi-scale parallel sampling, more semantic information can be obtained to improve SCD performance. Furthermore, in SMNet, to overcome the limitation of feature extraction

Table 4

BCD datasets of optical remote sensing images ('-' indicates no corresponding results are provided).

Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink	
Land cover	AICD (Bourdin et al., 2011)	2011	1000	800×600	0.5m	—	
	ZY3 (Zhang and Shi, 2020)	2017	1	458×559	5.8m	https://poles.tpd.ac.cn/zh-hans/data/b9748f8d-c06e-42a4-a2bc-7a185927f025/	
	BTCD (Song et al., 2021b)	2021	5281	256×256	—	—	
	TZ-CD (Deng et al., 2023)	2023	1	31307×40620	1m	—	
	Event	Year	Track	Image Pairs	Image size	Resolution	Event link
	Remote sensing image sparse characterization and intelligent Processing algorithm competition	2019	Remote sensing image change detection	103	960×960	—	https://autdatamotion.github.io/RSC2019/#/home
	Artificial Intelligence Remote sensing Interpretation Competition	2020	Change detection	4662	512×512	0.5-3m	https://rs.sensetime.com/
	SN7: Multi-Temporal Urban Development Challenge	2020	Multi-Temporal Urban Development Challenge	12	1024×1024	4m	https://spacenet.ai/sn7-challenge/
	“Shengteng Cup” remote sensing image intelligent processing algorithm competition	2021	Remote sensing image change detection	4194	512×512	1-2m	https://rsipac.whu.edu.cn/subject_two_2021
	Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink
Building	ABCD (Fujita et al., 2017)	2017	16,950	128×128	0.4m	https://github.com/gistairc/ABCDdataset	
	CDD Dataset (Lebedev et al., 2018)	2018	16,000	256×256	0.03-1m	https://drive.google.com/file/d/1GX656JqqOyBi_Ef0w65kDGVto-nHrNs9/edit	
	WHU Building CD (Ji et al., 2018)	2019	1	32507×15354	0.2m	https://study.rsgis.whu.edu.cn/pages/download/building_dataset.html	
	LEVIR-CD (Chen and Shi, 2020)	2020	637	1024×1024	0.5m	https://justchenhao.github.io/LEVIR/	
	LEVIR-CD+ (Shen et al., 2021)	2020	1970	1024×1024	0.5m	https://github.com/S2Looking/Dataset	
	CD_Data_GZ (Peng et al., 2020)	2020	19	1006×1168-4936×5224	0.55m	https://github.com/daifeng2016/Change-Detection-Dataset-for-High-Resolution-Satellite-Imagery	
	DSIFN Dataset (Zhang et al., 2020)	2020	3940	512×512	—	https://github.com/GeoZcx/A-deeply-supervised-image-fusion-on-network-for-change-detection-in-remote-sensing-images/tree/master/dataset	
	SYSU-CD (Shi et al., 2021)	2021	20,000	256×256	0.5m	https://github.com/liumency/SYSU-CD	
	LEVIR-CC (Liu et al., 2022a)	2022	10,077	1024×1024	0.5m	https://github.com/Chen-Yang-Liu/RSICC	
	SI-BU dataset (Liao et al., 2023)	2023	4932	512×512	0.2m	https://github.com/liaochengcsu/BCE-Net	
Riverway	EGY-BCD (Holail et al., 2023)	2023	6091	256×256	0.25m	https://github.com/oshholail/EGY-BCD	
	HRCUS-CD (Zhang et al., 2023a)	2023	11,388	256×256	0.5m	https://github.com/zjd1836/AERNet	
	TUE-CD (Liu et al., 2024c)	2024	1656	256×256	1.8m	https://github.com/RSMagneto/MSI-Net	
	Event	Year	Track	Image Pairs	Image size	Resolution	Event link
	The 5th “Sino-Keke Star Map Cup” International high-resolution remote sensing image interpretation Competition	2021	Building survey and change detection in high resolution visible light images	2000	512×512	2m	https://www.gaofen-challenge.com/challenge
Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink	
Riverway	The River Data Set (Wang et al., 2018)	2019	1	463×241	30m	https://share.weiyun.com/5xdge4R	

(continued on next page)

Table 4 (continued)

Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink
Cropland	CLCD (Liu et al., 2022b)	2022	600	512×512	0.5-2m	https://github.com/liumency/CropLand-CD
Mine	MineNetCD (Yu et al., 2024b)	2024	71,711	256×256	1.2m	https://huggingface.co/datasets/ericyu/MineNetCD256

Table 5

SCD datasets of optical remote sensing images ('-' indicates no corresponding results are provided).

Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink	
	HCCD (López-Fandiño et al., 2018)	2018	3	390×200	30m	–	
	Mts-WH (Wu et al., 2017b)	2019	1	7200×6000	1m	https://sigma.whu.edu.cn/newspage.php?q=2019_03_26	
	HRSCD (Daudt et al., 2019)	2019	291	10000×10000	0.5m	https://ieeegroups.ieee.org/open-access/hrscd-high-resolution-semantic-change-detection-dataset	
	SECOND (Yang et al., 2021a)	2020	4662	512×512	–	https://captain-whu.github.io/SCD/	
	Hi-UCD (Tian et al., 2020)	2020	1293	1024×1024	0.1m	https://github.com/Daisy-7/Hi-UCD-S	
	S2MTCP (Leenstra et al., 2021)	2021	1520	600×600	10m	https://zenodo.org/record/4280482	
	DynamicEarthNet (Toker et al., 2022)	2022	600	1024×1024	3m	https://mediatum.ub.tum.de/1650201	
	Landsat-SCD (Yuan et al., 2022)	2022	8468	416×416	30m	https://figshare.com/articles/figure/Landsat-SCD_dataset_zip/19946135/1	
	ChangNet (Ji et al., 2024)	2023	31,000	1900×1200	0.3m	https://github.com/jankyeec/ChangNet	
	CNAM-CD (Zhou et al., 2023)	2023	2503	512×512	0.5m	https://github.com/Silvestezhou/CNAM-CD	
	WUSU dataset (Shi et al., 2023b)	2023	3	6358×6382/ 7025×5500	1m	https://rsidea.whu.edu.cn/resource_wusu_sharing.htm	
	Hi-CNA dataset (Sun et al., 2024)	2024	6797	512×512	0.8m	https://rsidea.whu.edu.cn/Hi-CNA_dataset.htm	
Land cover	CropSCD (Feng et al., 2024)	2024	4141	512×512	0.5-2m	https://github.com/lsmlyn/CropSCD	
Target of interest	Event	Year	Track	Image Pairs	Image size	Resolution	Event link
	National Artificial Intelligence Competition	2020	AI + Remote sensing track	–	256×256	–	https://naicpcl.ac.cn/landingpage/2020/index.html
	“Space Map Cup” remote sensing image intelligent processing algorithm competition	2022	Remote sensing image change detection	6000+	512×512	1-2m	https://rsipac.whu.edu.cn/
	“Guofeng East Eye Cup” remote sensing image intelligent processing algorithm competition	2023	Object level change detection				
	“Jilin-1” Cup Satellite Remote Sensing Application Youth Innovation and Entrepreneurship Competition	2023	Cultivated land change detection based on high-resolution satellite images	8000	256×256	<0.75m	https://www.jl1mall.com/contest/match
	2024 “Jilin-1” Cup Satellite Remote Sensing Application Youth Innovation and Entrepreneurship Competition	2024	High resolution remote sensing image total element change detection	5000	512×512	<0.75m	https://www.jl1mall.com/contest/matchMenu
Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink	
	QFabric (Verma et al., 2021)	2021	2520	120×120 –12000×1200	0.31–0.70m	https://sagarverma.github.io/qfabric	
	S2Looking (Shen et al., 2021)	2021	5000	1024×1024	0.5–0.8m	https://github.com/S2Looking/Dataset	
	NanjingDataset (Shen et al., 2022)	2022	2519	256×256	0.3m	https://github.com/SianGIS/NanjingDataset	
	BANDON (Pang et al., 2023)	2023	2283	2048×2048	0.6m	https://github.com/fitzpchao/BANDON	
Target of interest	Event	Year	Track	Image Pairs	Image size	Resolution	Event link
Building	xView2 Challenge	2019	Building Damage Assessment	11,034	1024×1024	<0.8m	https://xview2.org/dataset
	“Tianzhi Cup” artificial intelligence challenge	2021	Visible light image building intelligent change detection	5000	1024×1024	0.5–0.7m	https://rsaicp.com/portal/contestList
	Remote Sensing Image Intelligent Interpretation Technology Challenge	2021	Building change detection in remote sensing images	10,000	512×512	–	https://captain-whu.github.io/PRCV2021_RS/tasks.html
Flood	SpaceNet8: Flood Detection Challenge	2022	Flood Detection Challenge Using Multiclass Segmentation	12	1300×1300	0.3–0.8m	https://spacenet.ai/sn8-challenge/
Target of interest	Dataset	Year	Image Pairs	Image size	Resolution	Weblink	
Landslide	GVLM (Zhang et al., 2023h)	2023	17	1748×1748- 10808 × 7424	0.59m	https://github.com/zxk688/GVLM	

Table 6

Comparison of quantitative results for different IB-DLCD algorithms (Bold denotes the best result, '#' indicates reproduced results using the released model, '*' indicates reproduced results using the released code).

Method	Source	Backbone	P/R/OA/F1/IoU(%)			
			WHU-CD	LEVIR-CD	SYSU-CD	Average
CNN-based IB-DLCD methods:						
FC-EF(Daudt et al., 2018a)	ICIP2018	UNet	71.63/67.25/97.61/69.37/53.11	86.91/80.17/98.39/83.40/71.53	74.32/75.84/86.02/75.07/60.09	77.62/74.42/94.00/75.94/61.57
FC-Siam-Diff(Daudt et al., 2018a)	ICIP 2018	UNet	47.33/77.66/95.63/58.81/41.66	89.53/83.31/98.67/86.31/75.92	89.13 /61.21/82.11/72.57/56.96	75.33/74.05/92.13/72.56/58.18
FC-Siam-Conc(Daudt et al., 2018a)	ICIP 2018	UNet	60.88/73.58/97.04/66.63/49.95	91.99/76.77/98.49/83.69/71.96	82.54/71.03/86.17/76.35/61.75	78.47/73.79/93.90/75.55/61.22
UNet++_MSOF(Peng et al., 2019)	RS2019	UNet++	88.96/82.27/98.92/85.48/74.65	89.08/85.37/97.31/87.19/77.28	81.36/75.39/86.39/78.26/62.14	86.46/81.01/94.20/83.64/71.35
SNUNet-CD(Fang et al., 2021)	GRSL2021	UNet++	85.60/81.49/98.71/83.50/71.67	89.18/87.17/98.82/88.16/78.83	83.49/76.37/90.87/79.77/66.35	86.09/81.67/96.13/83.81/72.11
Attention Mechanism-based IB-DLCD methods:						
DTCDSCN(Liu et al., 2020)	GRSL2020	SE-ResNet	78.65/81.24/98.05/79.92/66.56	89.25/85.68/98.75/87.43/77.67	83.19/77.25/90.96/80.11/66.82	87.03/81.39/95.92/82.49/70.35
IFN(Zhang et al., 2020)	JPRS2020	VGG16	92.24/75.78/98.59/83.11/71.12	94.39 /82.42/98.86/88.01/78.58	79.59/73.58/89.17/76.53/61.91	88.74/77.26/95.54/82.55/70.53
STANet(Chen and Shi, 2020)	RS2020	ResNet18	79.37/85.50/98.52/82.32/69.95	83.81/91.00/98.66/87.26/77.40	70.76/85.33/87.96/77.37/63.09	77.98/87.28/95.05/82.32/70.14
DMINet(Feng et al., 2023)	TGRS2023	ResNet18	93.84/86.25/98.97/88.69/79.68	92.52/89.95/99.07/90.71/82.99	85.03/79.86/91.67/82.08/69.60	90.46/85.35/95.05/87.16/77.42
TINYCD(Codogni et al., 2023)	NCA2023	EfficientNetb4	91.72/91.76/99.34/91.74/84.74	92.68/89.47/99.10/91.05/83.57	86.12/79.21/91.52/81.58/69.14	85.17/88.63/95.91/86.70/79.15
USSFC-Net(Lei et al., 2023a)	TGRS2023	U-Net	87.58*/95.75*/99.33*/91.48*/84.30*	89.37*/ 91.53* /99.04*/90.44*/82.54*	80.41*/78.72*/90.25*/79.56*/66.05*	85.79*/88.67*/96.21*/87.16*/77.63*
AZNet(Li et al., 2023d)	TGRS2023	MobileNetV2	97.89* /95.85*/99.75*/96.85*/93.90*	92.93*/90.62*/99.17*/91.76*/84.78*	85.77*/82.24*/92.60*/83.97*/72.37*	92.19* / 89.57* / 97.17* / 90.86* / 83.68*
Transformer-based IB-DLCD methods:						
FTN(Yan et al., 2022)	ACCV2022	Swin transformer	93.09/91.24/99.37/92.16/85.45	92.71/89.37/99.06/91.01/83.51	86.86/76.82/91.79/81.53/68.82	90.89/85.81/96.74/88.23/79.26
SwinUNet(Zhang et al., 2022b)	TGRS2022	Swin transformer	95.00/92.60/99.40/93.80/88.34	90.51/89.72/98.99/90.11/82.00	84.09/72.67/90.31/77.97/63.89	89.87/85.00/96.23/87.29/78.07
ChangeFormer(Bandara and Patel, 2022b)	IGARSS2022	Transformer	92.06/83.46/98.96/87.55/77.86	92.05/88.80/99.04/90.40/82.48	81.39/75.18/90.09/78.16/64.15	86.22/80.97/94.57/83.48/74.83
BiFA(Zhang et al., 2024a)	TGRS2024	SegformerB0	95.15/93.60/99.56/94.37/89.34	91.52/89.86/99.06/90.69/82.96	86.98/81.44/ 92.75 /84.12/72.59	88.27/86.45/97.12/87.35/81.63
CNN-Transformer Hybrid-based IB-DLCD methods:						
BIT(Chen et al., 2021b)	TGRS2021	ResNet18	86.64/81.48/98.75/83.98/72.39	89.24/89.37/98.92/89.31/80.68	82.18/74.49/90.18/78.15/64.13	86.02/81.78/95.95/83.81/72.40
ICIF-Net(Feng et al., 2022)	TGRS2022	ResNet18+PVTv2-B1	92.98/85.56/98.96/88.32/79.24	91.32/88.64/98.99/89.96/81.75	83.37/78.51/91.24/80.74/68.12	89.22/84.24/96.40/86.34/76.37
H-TransCD(Ke and Zhang, 2022)	JPRS2022	ResNet18	93.85/88.73/99.24/91.22/83.85	91.45/88.72/99.00/90.06/81.92	83.05/77.40/90.95/80.13/66.84	89.45/84.95/96.40/87.14/77.53
ACAHNet(Zhang et al., 2023c)	TGRS2023	Hybrid Transformer	85.51/90.16/99.00/87.77/78.21	92.36/90.68/99.14/91.51/84.35	83.96/81.91/97.82/73.70/55	87.28/87.46/96.70/87.34/77.70
VcT(Jiang et al., 2023)	TGRS2023	ResNet18	89.39/89.77/99.18/89.58/81.12	92.57/87.65/99.01/90.04/81.89	86.10*/ 86.22* /90.00*/ 86.16* / 76.41*	89.35/87.88/96.06/88.59/79.80
RCTNet(Gao et al., 2024)	Arxiv2024	RegNet	96.04*/94.09*/99.61*/95.06*/90.58*	92.94*/90.89*/ 99.18* / 91.91* / 85.02*	84.33*/81.74*/92.11*/83.01*/70.96*	91.10*/88.90*/96.96*/89.99*/82.18*

Table 7

Comparison of quantitative results of different SCD algorithms (Bold denotes the best result).

Method	Source	Backbone	Sek/mIoU(%)		
			SECOND	Landsat-SCD	Average
HRSCD-str.4(Daudt et al., 2019)					
HRSCD-str.4(Daudt et al., 2019)	CVIU2019	UNet	13.09/67.69	13.57/78.54	13.33/73.16
SCDNet(Peng et al., 2021)	JAG2021	ResNet34	18.99/70.63	50.05/85.23	34.52/77.93
Bi-SRNNet(Ding et al., 2022)	TGRS2022	ResNet34	17.96/70.56	38.82/80.32	28.39/75.44
SMNet(Niu et al., 2023)	RS2023	PRTB	20.29/71.95	51.14/85.65	35.72/78.80
STSP-Net(He et al., 2023b)	RS2023	ResNet50	20.91/72.03	51.71/85.54	36.31/78.79
DEFO-MTLSCD(Li et al., 2024b)	TGRS2024	ResNet34	23.73/73.76	59.89/89.44	41.81/81.60
SCanNet(Ding et al., 2024a)	TGRS2024	ResNet34 + SCanFormer	23.94/73.42	60.53/88.96	42.24/81.19
SIC-Net(Ning et al., 2024)	JSTARS2024	DCP + SCP	23.96/73.68	61.29/90.44	42.62/82.06

using CNN, global and local information are integrated through a PRTB backbone, leading to an obvious performance gain. To deal with limited sensitivity to details and class-imbalance issue in SCD, STSP-Net aims to enhance spatial contexts and balance the two temporal branches and CD branch. Consequently, it observes a large performance gain. In DEFO-MTLSCD, a decoder feature interaction module (DFIT) and a feature aggregation module (FAM) are proposed to deal with the feature interaction across SCD subtasks effectively. As a result, it achieves a competitive SCD performance on both the SECOND dataset and Landsat-SCD dataset. In particular, SCanNet also achieves SOTA SCD performance, where a semantic change transformer (SCanFormer) is introduced to learn the semantic transitions. By incorporating a detail capture and spatial-temporal semantic coordination module as well as a pseudo-label growth algorithm, the missing details can be better recovered, and the class-imbalance issues can be better mitigated. Consequently, SIC-Net achieved the best SCD performance on both SECOND and Landsat-SCD datasets. Note that due to include more changed types, more severe class-imbalance issues exist for SECOND dataset, leading to an inferior SCD performance against Landsat-SCD

dataset for all SCD algorithms. Specially, compared to DEFO-MTLSCD, SCanNet is more robust against class-imbalance issues, where it obtains inferior performance on Landsat-SCD dataset but superior performance on SECOND dataset. In addition, the average performance is consistent with per-dataset performance, where the top three algorithms are SIC-Net, SCanNet and DEFO-MTLSCD, respectively.

4.4. Model complexity analysis

To compare the computational efficiency of different algorithms, this study takes the LEVIR-CD dataset as an example and provides a comprehensive comparative analysis of typical algorithms in terms of network parameters, Floating-point Operations Per Second (FLOPs), and F1-score. The visual results are shown in Fig. 16. One can conclude that the simple encoder-decoder methods (FC-EF, FC-Siam-Diff, FC-Siam-Conc) possess the lowest parameters and FLOPs. However, their accuracy metrics are relatively lower compared to other methods. Based on the UNet++ network architecture, UNet++_MSOF and SNUNet-CD achieve stronger feature representation capability and higher CD

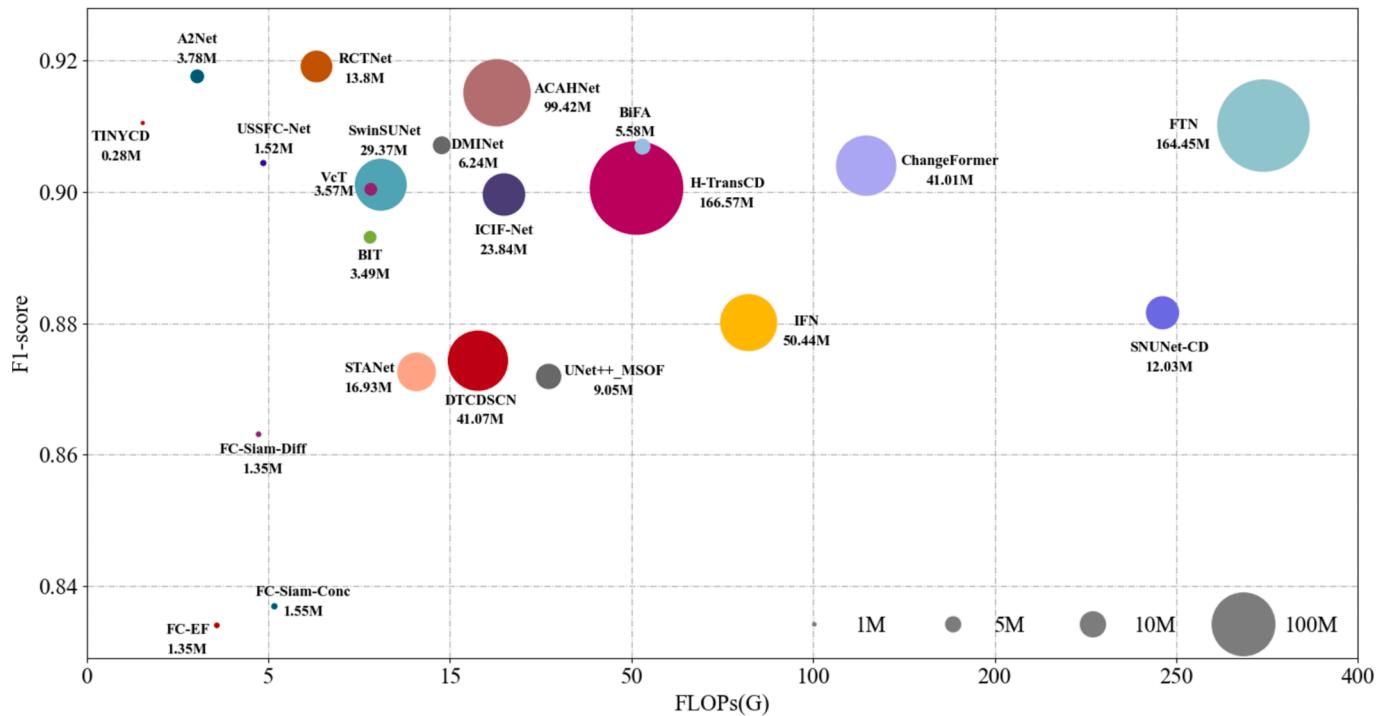


Fig. 16. Model complexity comparison of different BCD methods.

accuracy. However, these methods also come with a significant increase in model complexity. Furthermore, by incorporating spatial-channel attention mechanisms, DTCDSN, IFN, STANet, and DMINet are capable of learning discriminative feature representations, thereby further improving CD accuracy. However, the introduction of numerous attention modules leads to a significant increase in model complexity. In particular, lightweight CD methods such as TINYCD, USSFC-Net, and A2Net achieve high CD accuracy with low computational costs, striking a better balance between accuracy and efficiency.

Additionally, in FTN, SwinSUNet, ChangeFormer, due to their strong ability to capture global contextual relationships, CD performance can be improved significantly. However, the network parameters and computational complexity witness a sharp increase due to extensive usage of Transformer modules. For instance, the model parameters of FTN exceeds 164 million. Note that, due to the usage of lightweight Transformer of SegFormerB0, BiFA achieves high accuracy metrics with much less parameters and computational complexity. Furthermore, CNN-Transformer hybrid methods can effectively leverage the

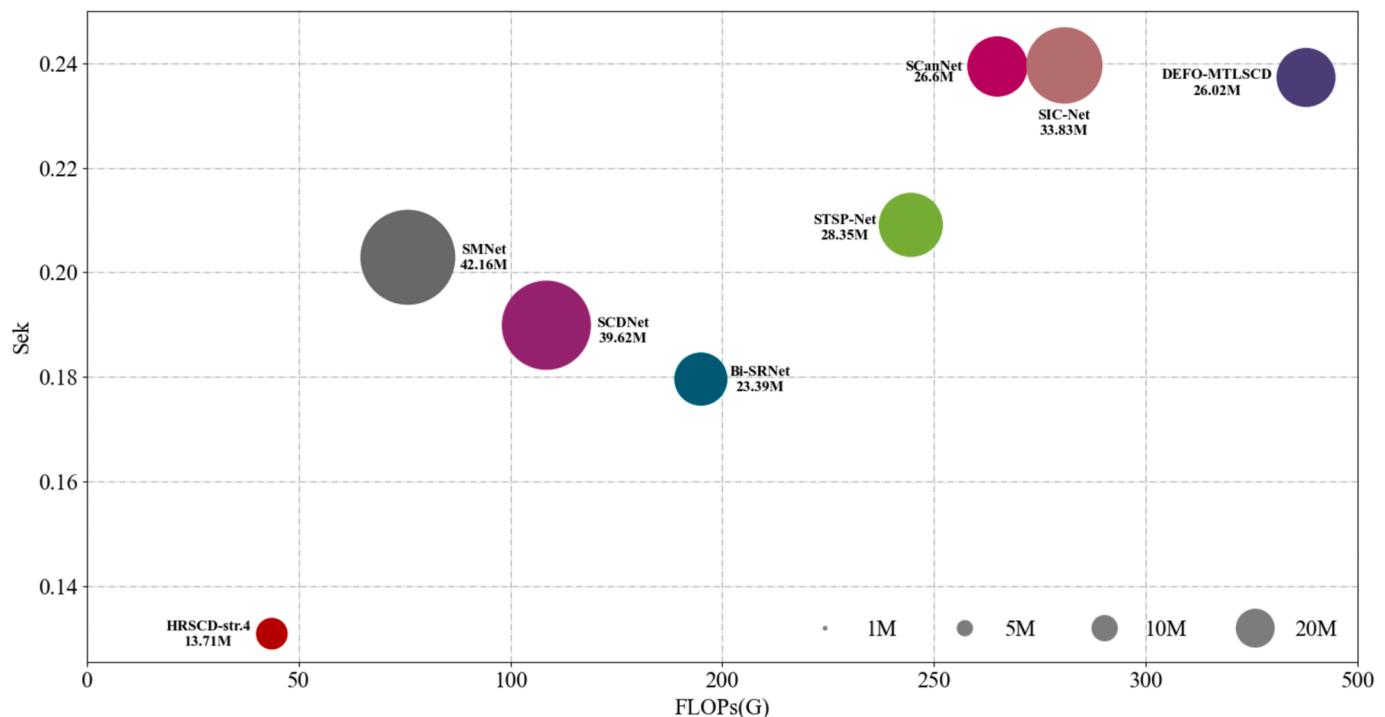


Fig. 17. Model complexity comparison of different SCD methods.

advantages of both architectures. Consequently, BIT, ICIF-Net, VcT, and RCTNet achieve competitive performance with reduced network parameters. Note that, due to the complexity of the Transformer modules, hybrid methods such as H-TransCD and ACAHNet tend to involve much larger network parameters.

In addition, based on SECOND dataset, a detailed complexity analysis is conducted from the perspectives of network parameters, Floating-point Operations Per Second (FLOPs), and Sek for different SCD methods. As presented in Fig. 17, one can observe that due to the usage of simple UNet backbone, HRSCD.st4. possesses the smallest model parameters and computational complexity but the worst quantitative performance. To improve feature extraction ability, ResNet34 is employed in SCDNet, Bi-SRNet and DEFO-MTLSCD, which leads to different degrees of performance gains at the cost of increased model parameters and computational complexity. Similarly, by introducing ResNet50 backbone, STSP-Net also observes a gain of SeK but an increase of model parameters and computational complexity. Particularly, due to including of a complicate decoder for feature interaction and aggregation, DEFO-MTLSCD achieves a much higher Sek score but with the largest computational complexity. To overcome the limitations of CNN backbone, a PRTB backbone is proposed in SMNet by integrating global and local information. As a result, it achieves a better Sek score but with the most model parameters. Differently, in SCanNet, a hybrid of ResNet and SCanFormer backbone is introduced, which is capable of obtaining competitive SCD performance with smaller gains in model parameters but larger gains in computational complexity. Note that CNN backbone with detail capture path (DCP) and semantic context path (SCP) is proposed in SIC-Net, a detail guidance module and a spatial-temporal semantic coordination module are also introduced to improve feature interaction for both encoding and decoding stage. Consequently, it obtains the best quantitative performance with a moderate increase in model parameters but a large increase in computational complexity.

5. Challenges and prospects

5.1. Lightweight change detection

Lightweight change detection aims to reduce model parameters and computational complexity while maintaining high accuracy and robustness. With the expanding application scenarios of CD, more and more researchers are focusing on how to apply CD algorithms on mobile devices and edge computing platforms to achieve online real-time CD. Currently, lightweight CD mainly employ: i) Indirect lightweight CD methods, where model pruning (Lei et al., 2023b) and knowledge distillation (Codegoni et al., 2023) strategies are used to compress the model globally without changing the main network modules. ii) Direct lightweight CD methods, which reduce the network parameters and complexity by replacing standard convolutional units with lightweight convolutional modules such as deformable convolution (Song et al., 2021a), channel-wise 1×1 convolution (Cui and Jiang, 2022), depth-wise separable convolution (Li et al., 2023d), etc. These techniques aim to reduce the computational cost and model complexity while maintaining the capability to capture meaningful features for CD. In the future, the integration of recent advancements in model compression techniques such as weight sharing, tensor decomposition (Dai et al., 2023a) and computational acceleration (Deng et al., 2020) hold significant research values for the design and training of lightweight CD models.

5.2. Label-efficient change detection

Data-driven DLCD methods often rely on large amounts of labeled samples for network parameter learning. Due to the difficulty of multi-temporal remote sensing image interpretation, acquiring the desired labeled data is challenging and costly. This significantly affects the

practical deployment and application of CD models. Hence, it is crucial to effectively enhancing the robustness and generalization capacity of CD models in scenarios with limited labeled samples. In existing approaches for label-efficient CD, semi-supervised or weakly supervised learning methods are mostly used to make full use of limited labeled data and abundant unlabeled data. These approaches, although effective in exploiting the potential of unlabeled data, remains to be improved from perspectives of samples or deep features. Future research directions in this field primarily include:

- i) Leveraging the potential of generative models of generative adversarial networks (Seo et al., 2023; Zheng et al., 2023) or diffusion models (Tang and Chen, 2024) to generate large amounts of high-quality training samples. These synthesized samples can then be used to improve model's transferability and generalization to real CD scenarios, which significantly improves zero-shot and few-shot CD performance. Note that the pre-trained generative models can also be served as a powerful feature extractor for downstream task fine-tuning. For example, Bandara et al. (2022) pre-trained a Denoising Diffusion Probabilistic Model (DDPM) to learn data distribution by using off-the-shell unlabeled data, which was then served as a feature extractor for CD task fine-tuning with a light-weight change decoder. Meanwhile, the class-imbalance issues in CD can also be well addressed by synthesizing more data of interest. In particular, diffusion models are capable of generating high-quality satellite images from thematic maps, digital maps, text, or images (Espinosa and Crowley, 2023; Khanna et al., 2023), which greatly expands the scope of sample generation. For instance, Yu et al. (2024d) proposed MetaEarth based on denoising diffusion models, which is capable of generating unbounded, multi-resolution RS images worldwide. Zheng et al. (2024a) proposed Change2, where scalable multi-temporal images and their change labels can be generated by using a generative probabilistic change model (GPCM).
- ii) By utilizing self-supervised learning techniques, it is possible to learn robust feature representations and construct generic backbone networks, thus reducing the demand for labeled data in downstream tasks. Subsequently, network fine-tuning techniques can be employed to tackle CD tasks. Due to its effectiveness, many foundation models have been trained to solve remote sensing tasks, such as RVSA (Wang et al., 2022a), RingMo (Sun et al., 2022b), CMID (Muhtar et al., 2023), SatMAE (Cong et al., 2022), Scale MAE (Reed et al., 2023), etc. Although these models have been proven effective in transferring into CD areas, a novel self-supervised foudation model which fully considers the spatial-temporal characteristics of CD remains to be explored.
- iii) By combining self-supervised and semi-supervised learning techniques, significant knowledge can be exploited from unlabeled samples in both the feature representation and model training stages. As a result, such approach enables “robust feature representation – efficient model training” throughout the entire process of label-efficient CD, making it possible to improve model generalization performance using as few labeled samples as possible.

5.3. Multi-source and multi-modal change detection

Due to the rapid development of sensor technology, the availability of multisource, multimodal remote sensing data has become increasingly abundant. Notably, here, multimodal and multi-source data refer to image data (Optical, SAR, Hyperspectral, etc.), video data, and point cloud data (LiDAR) acquired by RS-specific sensors, GIS data (Thematic maps, Digital maps, POIs, etc.) generated by GIS software and text data produced by human prior knowledge. These data are capable of providing multi-dimensional and all-weather spatial-temporal

information support for CD. For example, by combining 2D image data and 3D LiDAR data, it is possible to achieve 3DCD, which is capable of providing much more useful change information; by combining optical images and SAR images, it is possible obtain all-weather change information. In such context, it is natural to combine the advantages of these data to improve the robustness and reliability of CD. However, multi-modal data have significant organizational differences, making it challenging to establish a unified spatiotemporal representation framework. This poses a significant challenge for CD tasks that rely on precise comparisons of multi-temporal features. Future research directions in this field mainly include:

- i) Multi-modal transfer: By utilizing image translation techniques (such as GANs, contrastive learning, diffusion models, optimal transport models, etc.), it is possible to generate high-quality images sequences with unified modality. Subsequently, existing CD models can be easily introduced to deal with homogenous bi-temporal images (Li et al., 2021a). For example, Zhan et al. (2024) proposed a cross-domain separable translation network for multimodal CD by integrating within-domain self-reconstruction, cross-domain translation and cycle-reconstruction. However, the performance of CD heavily relies on the quality of image transfer, which easily leads to cumbersome processing steps, high computational costs, and error accumulation effect.
- ii) Multi-modal fusion and alignment: Instead of explicitly generating images of the same modality, images from different modality are mapped into a unified latent feature space for change analysis (Jain et al., 2022; Zhang et al., 2022a; Han et al., 2024b; Liu et al., 2024a). In such a way, model complexity and computational costs can be significantly reduced, and error accumulation effect can be suppressed to a large extent. For example, He et al. (2023a) proposed a cross-modal change detection network (CMCDNet) based on CNN, where multi-scale feature fusion is performed by using gating and self-attention modules. Chen et al. (2024b) proposed to fuse OpenStreetMap and optical imagery by introducing an object-guided Transformer. Particularly, text data is capable of providing valuable semantic information, effective fusion of text-image data is crucial to improving CD performance by exploiting textual cues and image features (Dong et al., 2024; Wang et al., 2024b; Zhao et al., 2024a). To summarize, how to effectively fuse and align multi-modal data while suppressing redundancy information remain to be further explored to deal with multi-modal CD task.
- iii) Versatile latent feature learning: With the development of self-supervised learning (such as contrastive learning, masked image modeling, etc.) and large foundation models, it is becoming increasingly popular to learning versatile and discriminative latent features directly from unlabeled multi-modal data, which is then used for direct feature comparison (Chen and Bruzzone, 2022) or downstream CD task fine-tuning (Wang et al., 2023e; Fuller et al., 2024). It simultaneously reduces the need of annotated data for multi-modal CD. For example, inspired by the concept of neural plasticity in brain science, Xiong et al. (2024) proposed a Dynamic One-For-All (DOFA) model, which is capable of obtaining a single versatile Transformer jointly trained on data from five sensors. On the whole, the large multi-modal foundation models have opened up new avenues for multi-modal data processing in a unified architecture, which paves the way for versatile and intelligent RS information processing. However, massive multi-modal training data and huge training cost are needed to train such a large model, which may be inaccessible for most common users.

5.4. Integration of geographic knowledge into change detection

Existing CD networks only consider multi-temporal image features

and borrow ideas from semantic segmentation to extract change patterns. Although these approaches achieve tremendous success in CD areas from image processing perspective, the incorporation of remote sensing prior knowledge is ignored, resulting in poor model interpretability and weak transferability. In general, remote sensing images contain rich geospatial information such as spatial location, orientation, topology. It is also easy to incorporate geographic knowledge such as terrain, slope, aspect, as well as textual semantic descriptions based on prior rules of the objective world. Therefore, it is significant to effectively integrate the aforementioned geospatial knowledge to enhance the accuracy and reliability of CD models. In particular, knowledge graphs can be used to construct structure graph patterns that represent the concepts and relationships among entities in the objective world, making it possible to create large-scale knowledge repositories that align with human cognition and facilitate machine understanding. In a pioneer work, Zhang et al. (2023i) introduced knowledge graphs to the field of remote sensing imagery intelligent interpretation. By mining knowledge from remote sensing domain, remote sensing imaging mechanisms, and geographical knowledge, they constructed a large-scale knowledge graph specific to remote sensing. Such knowledge graph was successfully applied in various domains such as remote sensing scene understanding (Li et al., 2021b), land cover classification (Li et al., 2022b), and marine oil spill monitoring (Liu et al., 2023a). It is worth noting that spatial-temporal knowledge graphs can enhance the reliability of CD, and the CD results can, in turn, update the knowledge base of the spatial-temporal knowledge graph, forming a mutually reinforcing iterative updating mechanism. However, the coupling of knowledge graphs and deep learning in CD methods is still in its early stages. It is therefore crucial to exploring how to better incorporate remote sensing imaging mechanisms and geospatial knowledge into DLCD models, making it possible to simultaneously achieve data-driven and knowledge reasoning-driven intelligent change detection.

5.5. Large foundation model-assisted change detection

With the booming of artificial intelligence techniques, the emergence of large foundation models with billions of parameters, such as GPT, BERT, and CLIP (Devlin et al., 2018; Radford et al., 2021; Liu et al., 2023b), have significantly propelled the practical application of AI technology, ushering in the era of large-scale AI models. In particular, visual foundation models represented by SAM (Segment Anything Model) have greatly promoted the generalization and intelligence of visual information processing. In such context, the network architectures for visual tasks have evolved from convolutional networks and Transformer networks to a new stage of pre-trained large foundation models. Overall, the techniques and methods of large foundation models will promote the intelligence of CD tasks in the following aspects:

- i) Utilizing massive remote sensing images to build a general-purpose remote sensing visual foundation model, which enables the implementation of a “plug-and-play” solution for CD tasks through the approach of “pre-training large models + downstream task fine-tuning”. Note that due to domain gap between natural images and RS images, it is impossible to directly apply pre-trained features on CD task. Meanwhile, it is difficult and costly to fine-tune the foundation models with billions of parameters. In such context, numerous parameter-efficient fine-tuning strategies are introduced by freezing the large foundation model parameters while adding small trainable adaptation parameters (Scheibenreif et al., 2024). For example, to adapt SAM to CD task, Chen et al. (2023a) proposed to introduce a Low-Rank Adaptation strategy (LORA) to add low-rank trainable parameters into the multi-head attention layers. Based on visual representations from FastSAM, Ding et al. (2024b) proposed a convolutional adaptor to aggregate task-oriented change information. However, detailed information is usually insufficient for

- ViT-based foundation models, which is essential to capture multi-scale change information. It is therefore crucial to introducing more high-frequency information into the foundation model for CD task (Mei et al., 2024). Overall, more effective fine-tuning strategies and more high-frequency information fusion algorithms remain to be further explored for CD task.
- ii) Leveraging the powerful zero-shot performance of visual language foundation models to provide high-quality pseudo labels for CD tasks, making it possible to effectively realize semi-supervised or weakly-supervised CD without the need of accurate annotation process. For example, Wang et al. (2023b) proposed a weakly-supervised CD method based on pseudo labels generated by using the localization capability of CAMs and zero-shot segmentation ability of SAM. Li et al. (2024a) proposed a novel semi-supervised CD method guided by a vision language model (VLM), where pseudo labels are produced by using a VLM-mixed change event generation strategy.
 - iii) By leveraging the technology of large foundation models, it becomes possible to explore massive multi-source remote sensing imagery to extract both low-level visual features and high-level semantic knowledge. This enables the realization of open-world CD, zero-shot CD, cross-modal CD, and open-vocabulary CD within a unified framework, which significantly improves the intelligence of CD. In such context, Guo et al. (2023) introduced a large multi-modal foundation model termed SkySense, which demonstrates remarkable generalization capabilities on both single-modal and multi-modal remote sensing tasks. Based on SAM, Zheng et al. (2024b) proposed a segment any change model for CD by introducing training-free adaptation and bitemporal latent matching, making zero-shot CD possible.
 - iv) By leveraging point prompts, rectangle prompts, and text prompts supported by large foundation models, the efficiency of CD data annotation can be greatly improved. This, in turn, drives the realization of on-demand and precise CD based on human-computer interaction. Such an approach will effectively expand the applications scenarios of CD and significantly enhance its efficiency and reliability. Specially, human language contains massive semantic information, which makes it possible for understanding the rules of objective world with ease and interacting with the others straightforwardly. By introducing large vision-language models into CD, rich semantic priors can be utilized to compensate for the ambiguity of visual information, thus improving the robustness and generalization of CD algorithms significantly. In addition, it is possible to tackle CD tasks in a much smarter interaction mode by using text prompt, where changed objects of interest can be extracted based on user requirements. For example, Zhao et al. (2024a) proposed a semantic first change paradigm (SeFi-CD), where anything you want change detection (AUWCD) was implemented by combining CLIP Surgery and SAM. Furthermore, change caption task is emerging for the benefit of in-depth change analysis using natural language, where the change map and its text description can be generated simultaneously (Wang et al., 2024b; Zhu et al., 2024). Particularly, more useful change information (such as change location, change area, change count, change type, change reason, etc.) can be continuously obtained through “question-and-answering” interaction with the large vision-language CD models, which opens the door to the ultimate embodied intelligence of CD (Noman et al., 2024a). For instance, Liu et al. (2024b) proposed an interactive Change-Agent by integrating a multi-level change interpretation model as eyes and a large language model as the brain. Through easy interaction with users, comprehensive change interpretation and insightful analysis can be conducted. Irvin et al. (2024) proposed a novel vision-language assistant model called TEOChat, which achieves

impressive zero-shot performance on change location and change question answering tasks.

6. Conclusion

This article provides an in-depth analysis of existing DLCD methods for optical remote sensing images from the perspective of algorithm granularity, namely FB-DLCD, PB-DLCD and IB-DLCD. Meanwhile, it offers a comprehensive analysis of the most widely used IB-DLCD methods in terms of model architecture, feature enhancement, and label-efficient learning strategies. Additionally, a comprehensive summary of existing optical remote sensing datasets for both BCD and SCD is provided, and a comparative analysis of typical BCD and SCD algorithms is conducted, which aims to serve as a crucial guidance for the design of CD networks and the selection of CD datasets. Finally, we conduct an in-depth exploration of future trends in CD field, including lightweight CD, label-efficient CD, multi-source and multi-modal CD, integration of geographic knowledge into CD, and large-scale foundation model-based CD. To sum up, emerging AI techniques, such as self-supervised learning, lifelong learning, knowledge graphs, and large foundation models, have been continuously developing and evolving, and opening up new opportunities for intelligent CD, smart CD, and ultimate realization of brain-inspired generic CD.

CRediT authorship contribution statement

Daifeng Peng: Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Xuelian Liu:** Writing – review & editing, Visualization, Methodology. **Yongjun Zhang:** Writing – review & editing, Funding acquisition, Formal analysis. **Haiyan Guan:** Resources, Investigation, Formal analysis. **Yansheng Li:** Writing – review & editing, Formal analysis. **Lorenzo Bruzzone:** Writing – review & editing, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (42371449, 41801386, 42030102), Technology Innovation Center for Integrated Applications in Remote Sensing and Navigation, Ministry of Natural Resources of the People's Republic of China (TICIARSN-2023-07), Key Laboratory of National Geographic Census and Monitoring, Ministry of Natural Resources (2023NGCM02) and Key Laboratory of Land Satellite Remote Sensing Application, Ministry of Natural Resources of the People's Republic of China (KLSMR-G202308).

Data availability

Data will be made available on request.

References

- Bai, T., Wang, L., Yin, D., Sun, K., Chen, Y., Li, W., Li, D., 2023. Deep learning for change detection in remote sensing: a review. *Geo-Spatial Inform. Sci.* 26 (3), 262–288. <https://doi.org/10.1080/10095020.2022.2085633>.
- Bandara W. G. C., Patel V. M., 2022a. Revisiting consistency regularization for semi-supervised change detection in remote sensing images. *arXiv preprint arXiv: 2204.08454*. Doi: 10.48550/arXiv.2204.08454.
- Bandara W. G. C., Patel V. M., 2023. Deep Metric Learning for Unsupervised Remote Sensing Change Detection. *arXiv preprint arXiv:2303.09536*. Doi: 10.48550/arXiv.2303.09536.

- Bandara, W.G.C., Nair, N.G., Patel, V.M., 2022. Ddpm-cd: Remote sensing change detection using denoising diffusion probabilistic models. arXiv preprint arXiv:11892. Doi: 10.48550/arXiv.2206.11892.
- Bandara, W.G.C., Patel, V.M., 2022. A transformer-based siamese network for change detection. In: IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 207–210. <https://doi.org/10.1109/IGARSS46834.2022.9883686>.
- Bourdis, N., Marraud, D., Sahbi, H., 2011. Constrained optical flow for aerial image change detection. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, pp. 4176–4179. <https://doi.org/10.1109/IGARSS.2011.6050150>.
- Celik, T., 2009. Unsupervised change detection in satellite images using principal component Analysis and k-Means Clustering. *IEEE Geosci. Remote Sens. Lett.* 6 (4), 772–776. <https://doi.org/10.1109/LGRS.2009.2025059>.
- Chen, Y., Bruzzone, L., 2022. A self-supervised approach to pixel-level change detection in bi-temporal RS images. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–11. <https://doi.org/10.1109/TGRS.2022.3203897>.
- Chen K., Liu C., Li W., Liu Z., Chen H., Zhang H., Zou Z., Shi Z., 2023a. Time travelling pixels: Bitemporal features integration with foundation model for remote sensing image change detection. arXiv preprint arXiv:16202. Doi: 10.48550/arXiv.2312.16202.
- Chen H., Song J., Han C., Xia J., Yokoya N., 2024c. Changemamba: Remote sensing change detection with spatio-temporal state space model. arXiv preprint arXiv: 03425. Doi: 10.48550/arXiv.2404.03425.
- Chen, H., Li, W., Shi, Z., 2021a. Adversarial instance augmentation for building change detection in remote sensing images. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–16. <https://doi.org/10.1109/TGRS.2021.3066802>.
- Chen, H., Qi, Z., Shi, Z., 2021b. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–14. <https://doi.org/10.1109/TGRS.2021.3095166>.
- Chen, H., Li, W., Chen, S., Shi, Z., 2022b. Semantic-aware dense representation learning for remote sensing image change detection. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–18. <https://doi.org/10.1109/TGRS.2022.3203769>.
- Chen, H., Hu, K., Filippi, P., Xiang, W., Bishop, T., Wang, Z., 2024a. Siamese Bi-Attention Pooling Network for Change Detection in Remote Sensing. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*. <https://doi.org/10.1109/JSTARS.2024.3373753>.
- Chen, H., Lan, C., Song, J., Broni-Bediako, C., Xia, J., Yokoya, N., 2024b. ObjFormer: Learning Land-Cover Changes From Paired OSM Data and Optical High-Resolution Imagery via Object-Guided Transformer. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–22. <https://doi.org/10.1109/TGRS.2024.3410389>.
- Chen, H., Shi, Z., 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens. (Basel)* 12 (10), 1662. <https://doi.org/10.3390/rs12101662>.
- Chen, F., Wang, N., Yu, B., Wang, L., 2022a. Res2-Unet, a new deep architecture for building detection from high spatial resolution images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 15, 1494–1501. <https://doi.org/10.1109/JSTARS.2022.3146430>.
- Chen, P., Zhang, B., Hong, D., Chen, Z., Yang, X., Li, B., 2022c. FCCDN: Feature constraint network for VHR image change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*. 187, 101–119. <https://doi.org/10.1016/j.isprsjprs.2022.02.021>.
- Chen, M., Zhang, Q., Ge, X., Xu, B., Hu, H., Zhu, Q., Zhang, X., 2023. A Full-Scale Connected CNN-Transformer Network for Remote Sensing Image Change Detection. *Remote Sens. (Basel)* 15 (22), 5383. <https://doi.org/10.3390/rs15225383>.
- Cheng G., Huang Y., Li X., Lyu S., Xu Z., Zhao Q., Xiang S., 2023. Change detection methods for remote sensing in the last decade: A comprehensive review. arXiv preprint arXiv:2305.05813. Doi: 10.48550/arXiv.2305.05813.
- Cheng, G., Wang, G., Han, J., 2022. ISNet: Towards Improving Separability for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11. <https://doi.org/10.1109/TGRS.2022.3174276>.
- Codegoni, A., Lombardi, G., Ferrari, A., 2023. TINYCD: a (not so) deep learning model for change detection. *Neural Computing Applications*. 35 (11), 8471–8486. <https://doi.org/10.1007/s00521-022-08122-3>.
- Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., Ermon, S., 2022. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Adv. Neural Inf. Proces. Syst.* 35, 197–211.
- Cui, F., Jiang, J., 2022. Shuffle-CDNet: A Lightweight Network for Change Detection of Bitemporal Remote-Sensing Images. *Remote Sens. (Basel)* 14 (15), 3548. <https://doi.org/10.3390/rs1415348>.
- Dai, W., Fan, J., Miao, Y., Hwang, K., 2023a. Deep Learning Model Compression With Rank Reduction in Tensor Decomposition. *IEEE Transactions on Neural Networks Learning Systems*. <https://doi.org/10.1109/TNNLS.2023.3330542>.
- Dai, Y., Zhao, K., Shen, L., Liu, S., Yan, X., Li, Z., 2023b. A Siamese Network Combining Multi-Scale Joint Supervision and Improved Consistency Regularization for Weakly Supervised Building Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*. 16, 4963–4982. <https://doi.org/10.1109/JSTARS.2023.3279863>.
- Daudt, R.C., Le Saux, B., Boulch, A., 2018a. Fully convolutional siamese networks for change detection. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 4063–4067. <https://doi.org/10.1109/ICIP.2018.8451652>.
- Daudt, R.C., Le Saux, B., Boulch, A., Gousseau, Y., 2018b. Urban change detection for multispectral earth observation using convolutional neural networks. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 2115–2118. <https://doi.org/10.1109/IGARSS.2018.8518015>.
- Daudt, R.C., Le Saux, B., Boulch, A., Gousseau, Y., 2019. Multitask learning for large-scale semantic change detection. *Computer Vision Image Understanding*. 187, 102783. <https://doi.org/10.1016/j.cviu.2019.07.003>.
- De Alban, J.D.T., Connnette, G.M., Oswald, P., Webb, E.L., 2018. Combined Landsat and L-Band SAR Data Improves Land Cover Classification and Change Detection in Dynamic Tropical Landscapes. *Remote Sens. (Basel)* 10 (2), 28. <https://doi.org/10.3390/rs10020306>.
- Demir, B., Bovolo, F., Bruzzone, L., 2013. Updating Land-Cover Maps by Classification of Image Time Series: A Novel Change-Detection-Driven Transfer Learning Approach. *IEEE Trans. Geosci. Remote Sens.* 51 (1), 300–312. <https://doi.org/10.1109/TGRS.2012.2195727>.
- Deng, L., Li, G., Han, S., Shi, L., Xie, Y., 2020. Model compression and hardware acceleration for neural networks: A comprehensive survey. *Proc. IEEE* 108 (4), 485–532. <https://doi.org/10.1109/JPROC.2020.2976475>.
- Deng, Y., Meng, Y., Chen, J., Yue, A., Liu, D., Chen, J., 2023. TChange: A Hybrid Transformer-CNN Change Detection Network. *Remote Sens. (Basel)* 15 (5), 1219. <https://doi.org/10.3390/rs15051219>.
- Deng, J., Wang, K., Deng, Y., Qi, G., 2008. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* 29 (16), 4823–4838. <https://doi.org/10.1080/01431160801950162>.
- Devlin J., Chang M.-W., Lee K., Toutanova K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805. Doi: 10.48550/arXiv.1810.04805.
- Ding, L., Guo, H., Liu, S., Mou, L., Zhang, J., Bruzzone, L., 2022. Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–14. <https://doi.org/10.1109/TGRS.2022.3154390>.
- Ding, Q., Shao, Z., Huang, X., Feng, X., Altan, O., Hu, B., 2023. Consistency-guided lightweight network for semi-supervised binary change detection of buildings in remote sensing images. *Giscience Remote Sensing*. 60 (1), 2257980. <https://doi.org/10.1080/15481603.2023.2257980>.
- Ding, L., Zhang, J., Guo, H., Zhang, K., Liu, B., Bruzzone, L., 2024a. Joint spatio-temporal modeling for semantic change detection in remote sensing images. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–14. <https://doi.org/10.1109/TGRS.2024.3362795>.
- Ding, L., Zhu, K., Peng, D., Tang, H., Yang, K., Bruzzone, L., 2024b. Adapting Segment Anything Model for Change Detection in VHR Remote Sensing Images. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–11. <https://doi.org/10.1109/TGRS.2024.3368168>.
- Dong, S., Wang, L., Du, B., Meng, X., 2024. ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 208, 53–69. <https://doi.org/10.1016/j.isprsjprs.2024.01.004>.
- Ertürk, A., Iordache, M.D., Plaza, A., 2017. Sparse Unmixing With Dictionary Pruning for Hyperspectral Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10 (1), 321–330. <https://doi.org/10.1109/JSTARS.2016.2606514>.
- Espinosa M., Crowley E. J., 2023. Generate your own scotland: Satellite image generation conditioned on maps. arXiv preprint arXiv:2308.16648. Doi: 10.48550/arXiv.2308.16648.
- Fang, L., Jiang, Y., Yu, H., Zhang, Y., Yue, J., 2023. Point Label Meets Remote Sensing Change Detection: A Consistency-Aligned Regional Growth Network. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–11. <https://doi.org/10.1109/TGRS.2023.3348459>.
- Fang, S., Li, K., Shao, J., Li, Z., 2021. SNUNet-CD: A densely connected Siamese network for change detection of VHR images. *IEEE Geoscience Remote Sensing Letters*. 19, 1–5. <https://doi.org/10.1109/LGRS.2021.3056416>.
- Feng, W., Guan, F., Sun, C., Xu, W., 2024. Cross-modal change detection flood extraction based on self-supervised contrastive pre-training. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 10, 75–82. <https://doi.org/10.5194/isprs-annals-X-1-2024-75-2024>.
- Feng, Y., Xu, H., Jiang, J., Liu, H., Zheng, J., 2022. ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–13. <https://doi.org/10.1109/TGRS.2022.3168331>.
- Feng, Y., Jiang, J., Xu, H., Zheng, J., Liu, H., 2023. Change detection on remote sensing images using dual-branch multilevel intertemporal network. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–15. <https://doi.org/10.1109/TGRS.2023.3241257>.
- Fujita, A., Sakurada, K., Imaizumi, T., Ito, R., Hikosaka, S., Nakamura, R., 2017. Damage detection from aerial images via convolutional neural networks. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp. 5–8. <https://doi.org/10.23919/MVA.2017.7986759>.
- Fuller, A., Millard, K., Green, J., 2024. CROMA: Remote sensing representations with contrastive radar-optical masked autoencoders. In: *Advances in Neural Information Processing Systems*, p. 36.
- Gao Y., Pei G., Sheng M., Sun Z., Chen T., Yao Y., 2024. Relating CNN-Transformer Fusion Network for Change Detection. arXiv preprint arXiv:03178. Doi: 10.48550/arXiv.2407.03178.
- Gong, M., Zhan, T., Zhang, P., Miao, Q., 2017. Superpixel-based difference representation learning for change detection in multispectral remote sensing images. *IEEE Transactions on Geoscience Remote Sensing*. 55 (5), 2658–2673. <https://doi.org/10.1109/TGRS.2017.2650198>.
- Gong, M., Yang, Y., Zhan, T., Niu, X., Li, S., 2019. A generative discriminatory classified network for change detection in multispectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*. 12 (1), 321–333. <https://doi.org/10.1109/JSTARS.2018.2887108>.
- Guo X., Lao J., Dang B., Zhang Y., Yu L., Ru L., Zhong L., Huang Z., Wu K., Hu D., 2023. Skysense: A multi-modal remote sensing foundation model towards universal

- interpretation for earth observation imagery. arXiv preprint arXiv:2312.10115. doi: 10.48550/arXiv.2312.10115.
- Guo, Q., Zhang, J., Zhu, S., Zhong, C., Zhang, Y., 2022. Deep Multiscale Siamese Network With Parallel Convolutional Structure and Self-Attention for Change Detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12. <https://doi.org/10.1109/TGRS.2021.3131993>.
- Han, T., Tang, Y., Chen, Y., Zou, B., Feng, H., 2024b. Global structure graph mapping for multimodal change detection. *Int. J. Digital Earth* 17 (1), 2347457. <https://doi.org/10.1080/17538947.2024.2347457>.
- Han, C., Wu, C., Hu, M., Li, J., Chen, H., 2024a. C2F-SemiCD: A Coarse-to-Fine Semi-Supervised Change Detection Method Based on Consistency Regularization in High-Resolution Remote-Sensing Images. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–21. <https://doi.org/10.1109/TGRS.2024.3370568>.
- He, X., Zhang, S., Xue, B., Zhao, T., Wu, T., 2023a. Cross-modal change detection flood extraction based on convolutional neural network. *International Journal of Applied Earth Observation Geoinformation*. 117, 103197. <https://doi.org/10.1016/j.jag.2023.103197>.
- He, Y., Zhang, H., Ning, X., Zhang, R., Chang, D., Hao, M., 2023b. Spatial-temporal semantic perception network for remote sensing image semantic change detection. *Remote Sens. (Basel)* 15 (16), 4095. <https://doi.org/10.3390/rs15164095>.
- Holail, S., Saleh, T., Xiao, X., Li, D., 2023. AFDE-Net: Building Change Detection using Attention-Based Feature Differential Enhancement for Satellite Imagery. *IEEE Geoscience Remote Sensing Letters*. 20, 1–5. <https://doi.org/10.1109/LGRS.2023.3283505>.
- Hou, X., Bai, Y., Li, Y., Shang, C., Shen, Q., 2021. High-resolution triplet network with dynamic multiscale feature for change detection on satellite images. *ISPRS J. Photogramm. Remote Sens.* 177, 103–115. <https://doi.org/10.1016/j.isprsjprs.2021.05.001>.
- Hua, Y., Mou, L., Zhu, X.X., 2019. Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification. *ISPRS Journal of Photogrammetry Remote Sensing*. 149, 188–199. <https://doi.org/10.1016/j.isprsjprs.2019.01.015>.
- Huang, R., Wang, R., Guo, Q., Wei, J., Zhang, Y., Fan, W., Liu, Y., 2023. Background-mixed augmentation for weakly supervised change detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 7919–7927. <https://doi.org/10.1609/aaai.v37i7.25958>.
- Irvin J. A., Liu E. R., Chen J. C., Dormoy I., Kim J., Khanna S., Zheng Z., Ermon S., 2024. TEOChat: A Large Vision-Language Assistant for Temporal Earth Observation Data. arXiv preprint arXiv:2410.06234. doi: 10.48550/arXiv.2410.06234.
- Jain, P., Schoen-Phelan, B., Ross, R., 2022. Self-supervised learning for invariant representations from multi-spectral and SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*. 15, 7797–7808. <https://doi.org/10.1109/JSTARS.2022.3204888>.
- Ji, D., Gao, S., Tao, M., Lu, H., Zhao, F., 2024. Changenet: Multi-temporal asymmetric change detection dataset. In: ICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2725–2729. <https://doi.org/10.1109/ICASSP48485.2024.10446592>.
- Ji, S., Wei, S., Lu, M., 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience Remote Sensing*. 57 (1), 574–586. <https://doi.org/10.1109/TGRS.2018.2858817>.
- Jiang, W., Sun, Y., Lei, L., Kuang, G., Ji, K., 2024. Change detection of multisource remote sensing images: a review. *Int. J. Digital Earth* 17 (1), 2398051. <https://doi.org/10.1080/17538947.2024.2398051>.
- Jiang, B., Wang, Z., Wang, X., Zhang, Z., Chen, L., Wang, X., Luo, B., 2023. VcT: Visual change transformer for remote sensing image change detection. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–14. <https://doi.org/10.1109/TGRS.2023.3327139>.
- Ke, Q., Zhang, P., 2022. Hybrid-transcd: A hybrid transformer remote sensing image change detection network via token aggregation. *ISPRS Int. J. Geo Inf.* 11 (4), 263. <https://doi.org/10.3390/ijgi11040263>.
- Khanna S., Liu P., Zhou L., Meng C., Rombach R., Burke M., Lobell D., Ermon S., 2023. Diffusionsat: A generative foundation model for satellite imagery. arXiv preprint arXiv:2312.03606. doi: 10.48550/arXiv.2312.03606.
- Khelifi, L., Mignotte, M., 2020. Deep learning for change detection in remote sensing images: comprehensive review and meta-analysis. *IEEE Access* 8, 126385–126400. <https://doi.org/10.1109/ACCESS.2020.3008036>.
- Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltasavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry Remote Sensing*. 146, 305–319. <https://doi.org/10.1016/j.isprsjprs.2018.09.018>.
- Lebedev, M., Vizilter, Y.V., Vygodov, O., Knyaz, V.A., Rubis, A.Y., 2018. Change detection in remote sensing images using conditional adversarial networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 42, 565–571. <https://doi.org/10.5194/isprs-archives-XLII-2-565-2018>.
- Leenstra, M., Marcos, D., Bovalo, F., Tuia, D., 2021. Self-supervised pre-training enhances change detection in Sentinel-2 imagery. *Pattern Recognition. ICPR International Workshops and Challenges*. 12667, 578–590. https://doi.org/10.1007/978-3-030-68787-8_42.
- Lei, T., Geng, X., Ning, H., Lv, Z., Gong, M., Jin, Y., Nandi, A.K., 2023a. Ultralightweight Spatial-Spectral Feature Cooperation Network for Change Detection in Remote Sensing Images. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–14. <https://doi.org/10.1109/TGRS.2023.3261273>.
- Lei, T., Xu, Y., Ning, H., Lv, Z., Min, C., Jin, Y., Nandi, A.K., 2023b. Lightweight Structure-aware Transformer Network for VHR Remote Sensing Image Change Detection. *IEEE Geosci. Remote Sens. Lett.* 21, 1–5. <https://doi.org/10.1109/LGRS.2023.3323534>.
- Li, X., Du, Z., Huang, Y., Tan, Z., 2021a. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS Journal of Photogrammetry Remote Sensing*. 179, 14–34. <https://doi.org/10.1016/j.isprsjprs.2021.07.007>.
- Li, Z., Tang, C., Liu, X., Li, C., Li, X., Zhang, W., 2023c. MS-Former: Memory-Supported Transformer for Weakly Supervised Change Detection with Patch-Level Annotations. arXiv preprint arXiv:2311.09726. doi: 10.48550/arXiv.2311.09726.
- Li, K., Cao, X., Deng, Y., Meng, D., 2024a. DiffMatch: Visual-Language Guidance Makes Better Semi-supervised Change Detector. arXiv preprint arXiv:2405.04788. doi: 10.48550/arXiv.2405.04788.
- Li, Y., Kong, D., Zhang, Y., Tan, Y., Chen, L., 2021b. Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification. *ISPRS Journal of Photogrammetry Remote Sensing*. 179, 145–158. <https://doi.org/10.1016/j.isprsjprs.2021.08.001>.
- Li, Y., Ouyang, S., Zhang, Y., 2022b. Combining deep learning and ontology reasoning for remote sensing image semantic segmentation. *Knowl.-Based Syst.* 243, 108469. <https://doi.org/10.1016/j.knosys.2022.108469>.
- Li, Z., Tang, C., Liu, X., Zhang, W., Dou, J., Wang, L., Zomaya, A.Y., 2023d. Lightweight Remote Sensing Change Detection With Progressive Feature Aggregation and Supervised Attention. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–12. <https://doi.org/10.1109/TGRS.2023.3241436>.
- Li, Z., Wang, X., Fang, S., Zhao, J., Yang, S., Li, W., 2024. A Decoder-Focused Multi-Task Network for Semantic Change Detection. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–25. <https://doi.org/10.1109/TGRS.2024.3362728>.
- Li, W., Xue, L., Wang, X., Li, G., 2023b. ConvTransNet: A CNN-Transformer Network for Change Detection with Multi-Scale Global-Local Representations. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–15. <https://doi.org/10.1109/TGRS.2023.3272694>.
- Li, L.-L., You, Z.-H., Chen, S.-B., Huang, L.-L., Tang, J., Luo, B., 2023a. GateFormer: Gate Attention UNet With Transformer for Change Detection of Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 17, 871–883. <https://doi.org/10.1109/JSTARS.2023.3335281>.
- Li, Q., Zhong, R., Du, X., Du, Y., 2022a. TransUNetCD: A hybrid transformer network for change detection in optical remote-sensing images. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–19. <https://doi.org/10.1109/TGRS.2022.3169479>.
- Liang, S., Hua, Z., Li, J., 2022. Transformer-based multi-scale feature fusion network for remote sensing change detection. *J. Appl. Remote Sens.* 16 (4), 046509. <https://doi.org/10.1117/1.JRS.16.046509>.
- Liang, Y., Zhang, C., Han, M., 2023. RaSRNet: An end-to-end Relation-aware Semantic Reasoning Network for Change Detection in Optical Remote Sensing Images. *IEEE Transactions on Instrumentation Measurement*. 73, 1–11. <https://doi.org/10.1109/TIM.2023.3243680>.
- Liao, C., Hu, H., Yuan, X., Li, H., Liu, C., Liu, C., Fu, G., Ding, Y., Zhu, Q., 2023. BCE-Net: Reliable building footprints change extraction based on historical map and up-to-date images using contrastive learning. *ISPRS Journal of Photogrammetry Remote Sensing*. 201, 138–152. <https://doi.org/10.1016/j.isprsjprs.2023.05.011>.
- Lin, M., Yang, G., Zhang, H., 2022. Transition is a process: Pair-to-video change detection networks for very high resolution remote sensing images. *IEEE Trans. Image Process.* 32, 57–71. <https://doi.org/10.1109/TIP.2022.3226418>.
- Lin, M., Chai, Z., Deng, H., Liu, R., 2022b. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. *IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*. 15, 4297–4306. <https://doi.org/10.1109/JSTARS.2022.3177235>.
- Liu, J., Chen, K., Xu, G., Sun, X., Yan, M., Diao, W., Han, H., 2019. Convolutional neural network-based transfer learning for optical aerial images change detection. *IEEE Geoscience Remote Sensing Letters*. 17 (1), 127–131. <https://doi.org/10.1109/LGRS.2019.2916601>.
- Liu, B., Chen, H., Li, K., Yang, M.Y., 2024a. Transformer-based multimodal change detection with multitask consistency constraints. *Inf. Fusion* 108, 102358. <https://doi.org/10.1016/j.inffus.2024.102358>.
- Liu, C., Chen, K., Zhang, H., Qi, Z., Zou, Z., Shi, Z., 2024b. Change-Agent: Towards Interactive Comprehensive Change Interpretation and Analysis from Change Detection and Change Captioning. arXiv preprint arXiv:2403.19646. doi: 10.48550/arXiv.2403.19646.
- Liu, Y., Pang, C., Zhan, Z., Zhang, X., Yang, X.J.I.G., 2020. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geoscience Remote Sensing Letters*. 18 (5), 811–815. <https://doi.org/10.1109/LGRS.2020.2988032>.
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., 2023b. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*. 1 (2), 100017. <https://doi.org/10.1016/j.metrad.2023.100017>.
- Liu, X., Zhang, Y., Zou, H., Wang, F., Cheng, X., Wu, W., Liu, X., Li, Y., 2023a. Multi-source knowledge graph reasoning for ocean oil spill detection from satellite SAR images. *International Journal of Applied Earth Observation Geoinformation*. 116, 103153. <https://doi.org/10.1016/j.jag.2022.103153>.
- Liu, Y., Zhang, K., Guan, C., Zhang, S., Li, H., Wan, W., Sun, J., 2024c. Building Change Detection in Earthquake: A Multi-Scale Interaction Network With Offset Calibration and A Dataset. *IEEE Transactions on Geoscience Remote Sensing*. 62, 1–17. <https://doi.org/10.1109/TGRS.2024.3438290>.
- Liu, C., Zhao, R., Chen, H., Zou, Z., Shi, Z., 2022a. Remote sensing image change captioning with dual-branch transformers: A new method and a large scale dataset. *IEEE Transactions on Geoscience Remote Sensing*. 60, 1–20. <https://doi.org/10.1109/TGRS.2022.3218921>.

- Long, Y., Xia, G.-S., Li, S., Yang, W., Yang, M.Y., Zhu, X.X., Zhang, L., Li, D., 2021. On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-aid. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 14, 4205–4230. <https://doi.org/10.1109/JSTARS.2021.3070368>.
- López-Fandiño, J., Garea, A.S., Heras, D.B., Argüello, F., 2018. Stacked autoencoders for multiclass change detection in hyperspectral images. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 1906–1909. <https://doi.org/10.1109/IGARSS.2018.8518338>.
- Lv, P., Zhong, Y., Zhao, J., Zhang, L., 2018. Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery. IEEE Transactions on Geoscience Remote Sensing. 56 (7), 4002–4015. <https://doi.org/10.1109/TGRS.2018.2819367>.
- Lyu, H., Lu, H., Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sens. (Basel) 8 (6), 506. <https://doi.org/10.3390/rs8060506>.
- Ma, X., Wu, Z., Lian, R., Zhang, W., Song, S., 2024. Rethinking Remote Sensing Change Detection With A Mask View. arXiv preprint arXiv:2406.15320. Doi: 10.48550/arXiv.2406.15320.
- Mei, L., Ye, Z., Xu, C., Wang, H., Wang, Y., Lei, C., Yang, W., Li, Y., 2024. SCD-SAM: Adapting Segment Anything Model for Semantic Change Detection in Remote Sensing Imagery. IEEE Transactions on Geoscience Remote Sensing. 62, 1–13. <https://doi.org/10.1109/TGRS.2024.3407884>.
- Mou, L., Bruzzone, L., Zhu, X.X., 2018. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. IEEE Transactions on Geoscience Remote Sensing. 57 (2), 924–935. <https://doi.org/10.1109/TGRS.2018.2863224>.
- Muhtar, D., Zhang, X., Xiao, P., Li, Z., Gu, F., 2023. Cmid: A unified self-supervised learning framework for remote sensing image understanding. IEEE Transactions on Geoscience Remote Sensing. 61, 1–17. <https://doi.org/10.1109/TGRS.2023.3268232>.
- Ning, X., He, Y., Zhang, H., Zhang, R., Chang, D., Hao, M., 2024. Semantic Information Collaboration Network for Semantic Change Detection in Remote Sensing Images. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 17, 12893–12909. <https://doi.org/10.1109/JSTARS.2024.3418632>.
- Niu, Y., Guo, H., Lu, J., Ding, L., Yu, D., 2023. SMNet: symmetric multi-task network for semantic change detection in remote sensing images based on CNN and transformer. Remote Sens. (Basel) 15 (4), 949. <https://doi.org/10.3390/rs15040949>.
- Noh, H., Ju, J., Kim, Y., Kim, M., Choi, D.-G., 2024. Unsupervised change detection based on image reconstruction loss with segment anything. Remote Sens. Lett. 15 (9), 919–929. <https://doi.org/10.1080/2150704X.2024.2388851>.
- Noman, M., Ahsan, N., Naseer, M., Cholakkal, H., Anwer, R. M., Khan, S., Khan, F. S., 2024a. CDChat: A Large Multimodal Model for Remote Sensing Change Description. arXiv preprint arXiv:2409.16261. Doi: 10.48550/arXiv.2409.16261.
- Noman, M., Fiaz, M., Cholakkal, H., Narayan, S., Anwer, R.M., Khan, S., Khan, F.S., 2024. Remote sensing change detection with transformers trained from scratch. IEEE Transactions on Geoscience Remote Sensing. 62 (1–14). <https://doi.org/10.1109/TGRS.2024.3383800>.
- Pang, C., Wu, J., Ding, J., Song, C., Xia, G.-S., 2023. Detecting building changes with off-nadir aerial images. SCIENCE CHINA Inf. Sci. 66 (4), 140306. <https://doi.org/10.1007/s11432-022-3691-4>.
- Paranjape, J. N., de Melo, C., Patel, V. M., 2024. A Mamba-based Siamese Network for Remote Sensing Change Detection. arXiv preprint arXiv:2407.06839. Doi: 10.48550/arXiv.2407.06839.
- Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., Huang, X., 2020. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. IEEE Transactions on Geoscience Remote Sensing. 59 (7), 5891–5906. <https://doi.org/10.1109/TGRS.2020.3011913>.
- Peng, D., Bruzzone, L., Zhang, Y., Guan, H., He, P., 2021. SCDNet: A novel convolutional network for semantic change detection in high resolution optical remote sensing imagery. International Journal of Applied Earth Observation Geoinformation. 103, 102465. <https://doi.org/10.1016/j.jag.2021.102465>.
- Peng, D., Guan, H., 2019. Unsupervised change detection method based on saliency analysis and convolutional neural network. J. Appl. Remote Sens. 13 (2), 024512. <https://doi.org/10.1117/1.JRS.13.024512>.
- Peng, D., Zhang, Y., Guan, H., 2019. End-to-end change detection for high resolution satellite images using improved UNet++. Remote Sens. (Basel) 11 (11), 1382. <https://doi.org/10.3390/rs11111382>.
- Peng, D., Zhai, C., Zhang, Y., Guan, H., 2023. High-resolution optical remote sensing image change detection based on dense connection and attention feature fusion network. Photogram. Rec. 38 (184), 498–519. <https://doi.org/10.1111/phor.12462>.
- Qing, Y., Ming, D., Wen, Q., Weng, Q., Xu, L., Chen, Y., Zhang, Y., Zeng, B., 2022. Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level. International Journal of Applied Earth Observation Geoinformation. 112, 102899. <https://doi.org/10.1016/j.jag.2022.102899>.
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., 2021. Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning, pp. 8748–8763.
- Reed, C.J., Gupta, R., Li, S., Brockman, S., Funk, C., Clipp, B., Keutzer, K., Candido, S., Uyttendaele, M., Darrell, T., 2023. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4088–4099.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Růžička V., D'Aronco S., Wegner J. D., Schindler K., 2020. Deep active learning in remote sensing for data efficient change detection. arXiv preprint arXiv:2008.11201. Doi: 10.48550/arXiv.2008.11201.
- Scheibenbrenner, L., Mommert, M., Borth, D., 2024. Parameter Efficient Self-Supervised Geospatial Domain Adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 27841–27851.
- Seo, M., Lee, H., Jeon, Y., Seo, J., 2023. Self-pair: Synthesizing changes from single source for object change detection in remote sensing imagery. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 6374–6383.
- Shafique, A., Cao, G., Khan, Z., Asad, M., Aslam, M., 2022. Deep learning-based change detection in remote sensing images: A review. Remote Sens. (Basel) 14 (4), 871. <https://doi.org/10.3390/rs14040871>.
- Shen, Q., Huang, J., Wang, M., Tao, S., Yang, R., Zhang, X., 2022. Semantic feature-constrained multitask siamese network for building change detection in high-spatial-resolution remote sensing imagery. ISPRS Journal of Photogrammetry Remote Sensing. 189, 78–94. <https://doi.org/10.1016/j.isprsjprs.2022.05.001>.
- Shen, L., Lu, Y., Chen, H., Wei, H., Xie, D., Yue, J., Chen, R., Lv, S., Jiang, B., 2021. S2Looking: A satellite side-looking dataset for building change detection. Remote Sens. (Basel) 13 (24), 5094. <https://doi.org/10.3390/rs13245094>.
- Shi, Q., Liu, M., Li, S., Liu, X., Wang, F., Zhang, L., 2021. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. IEEE Transactions on Geoscience Remote Sensing. 60, 1–16. <https://doi.org/10.1109/TGRS.2021.3085870>.
- Shi, G., Mei, Y., Wang, X., Yang, Q., 2023a. DAHT-Net: Deformable Attention-Guided Hierarchical Transformer Network Based on Remote Sensing Image Change Detection. IEEE Access 11, 103033–103043. <https://doi.org/10.1109/ACCESS.2023.3307642>.
- Shi, W., Zhang, M., Zhang, R., Chen, S., Zhan, Z., 2020. Change detection based on artificial intelligence: State-of-the-art and challenges. Remote Sens. (Basel) 12 (10), 1688. <https://doi.org/10.3390/rs12101688>.
- Shi, S., Zhong, Y., Liu, Y., Wang, J., Wan, Y., Zhao, J., Lv, P., Zhang, L., Li, D., 2023b. Multi-temporal urban semantic understanding based on GF-2 remote sensing imagery: from tri-temporal datasets to multi-task mapping. Int. J. Digital Earth 16 (1), 3321–3347. <https://doi.org/10.1080/17538947.2023.2246445>.
- Song, K., Cui, F., Jiang, J., 2021a. An efficient lightweight neural network for remote sensing image change detection. Remote Sens. (Basel) 13 (24), 5152. <https://doi.org/10.3390/rs13245152>.
- Song, X., Huo, Z., Li, J., 2022b. Remote sensing image change detection transformer network based on dual-feature mixed attention. IEEE Transactions on Geoscience Remote Sensing. 60, 1–16. <https://doi.org/10.1109/TGRS.2022.3209972>.
- Song, K., Jiang, J., 2021. AGCDNet: An attention-guided network for building change detection in high-resolution remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 14, 4816–4831. <https://doi.org/10.1109/JSTARS.2021.3077545>.
- Song, L., Xia, M., Jin, J., Qian, M., Zhang, Y., 2021b. SUACDNet: Attentional change detection network based on siamese U-shaped structure. International Journal of Applied Earth Observation Geoinformation. 105, 102597. <https://doi.org/10.1016/j.jag.2021.102597>.
- Song, L., Xia, M., Weng, L., Lin, H., Qian, M., Chen, B., 2022a. Axial cross attention meets CNN: Bibranch fusion network for change detection. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 16, 21–32. <https://doi.org/10.1109/JSTARS.2022.3224081>.
- Su, X., Wu, W., Li, H., Han, Y., 2011. Land-use and land-cover change detection based on object-oriented theory. In: 2011 International Symposium on Image and Data Fusion, pp. 1–4. <https://doi.org/10.1109/ISIDF.2011.6024300>.
- Sun, X., Wang, P., Lu, W., Zhu, Z., Lu, X., He, Q., Li, J., Rong, X., Yang, Z., Chang, H., 2022b. RingMo: A remote sensing foundation model with masked image modeling. IEEE Transactions on Geoscience Remote Sensing. 61, 1–22. <https://doi.org/10.1109/TGRS.2022.3194732>.
- Sun, C., Wu, J., Chen, H., Du, C., 2022a. SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using Siamese networks with graph attention. Remote Sens. (Basel) 14 (12), 2801. <https://doi.org/10.3390/rs14122801>.
- Sun, C., Chen, H., Du, C., Jing, N., 2023. SemiBuildingChange: A Semi-supervised High-Resolution Remote Sensing Image Building Change Detection Method With A Pseudo Bi-Temporal Data Generator. IEEE Transactions on Geoscience Remote Sensing. 60, 1–19. <https://doi.org/10.1109/TGRS.2023.3321637>.
- Sun, Z., Zhong, Y., Wang, X., Zhang, L., 2024. Identifying cropland non-agriculturalization with high representational consistency from bi-temporal high-resolution remote sensing images: From benchmark datasets to real-world application. ISPRS Journal of Photogrammetry Remote Sensing. 212, 454–474. <https://doi.org/10.1016/j.isprsjprs.2024.05.011>.
- Tan, X., Chen, G., Wang, T., Wang, J., Zhang, X., 2024. Segment change model (scm) for unsupervised change detection in vr remote sensing images: a case study of buildings. In: IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium, pp. 8577–8580. <https://doi.org/10.1109/IGARSS53475.2024.10642429>.
- Tang, K., Chen, J., 2024. ChangeAnywhere: Sample Generation for Remote Sensing Change Detection via Semantic Latent Diffusion Model. arXiv preprint arXiv:08892. Doi: 10.48550/arXiv.2404.08892.
- Tang, H., Wang, H., Zhang, X., 2021. A segmentation map difference-based domain adaptive change detection method. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 14, 9571–9583. <https://doi.org/10.1109/JSTARS.2021.3113327>.

- Tang, X., Zhang, T., Ma, J., Zhang, X., Liu, F., Jiao, L., 2023. Wnet: W-shaped hierarchical network for remote sensing image change detection. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–14. <https://doi.org/10.1109/TGRS.2023.3296383>.
- Tao, C., Qi, J., Li, Y., Wang, H., Li, H., 2019. Spatial information inference net: Road extraction using road-specific contextual information. *ISPRS Journal of Photogrammetry and Remote Sensing*. 158, 155–166. <https://doi.org/10.1016/j.isprsjprs.2019.10.001>.
- Tian, S., Ma, A., Zheng, Z., Zhong, Y., 2020. Hi-UCD: A large-scale dataset for urban semantic change detection in remote sensing imagery. *arXiv preprint arXiv:2011.03247*. Doi: 10.48550/arXiv.2011.03247.
- Toker, A., Kondmann, L., Weber, M., Eisenberger, M., Camero, A., Hu, J., Hoderlein, A. P., Şenaras, Ç., Davis, T., Cremers, D., 2022. Dynamicearthnet: Daily multi-spectral satellite dataset for semantic change segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21158–21167.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Proces. Syst.* 30.
- Verma, S., Panigrahi, A., Gupta, S., 2021. QFabric: Multi-task change detection dataset. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1052–1061.
- Wang, D., Chen, X., Jiang, M., Du, S., Xu, B., Wang, J., 2021a. ADS-Net: An Attention-Based deeply supervised network for remote sensing image change detection. *International Journal of Applied Earth Observation and Geoinformation*. 101, 102348. <https://doi.org/10.1016/j.jag.2021.102348>.
- Wang, Y., Albrecht C. M., Braham N. A. A., Liu C., Xiong Z., Zhu X. X., 2023e. DeCUR: decoupling common & unique representations for multimodal self-supervision. *arXiv preprint arXiv:205300*. Doi: 10.48550/arXiv.2309.05300.
- Wang, Y., Yu, W., Kopp M., Ghamisi P., 2024b. ChangeMinds: Multi-task Framework for Detecting and Describing Changes in Remote Sensing. *arXiv preprint arXiv:2410.10047*. Doi: 10.48550/arXiv.2410.10047.
- Wang, Y., Hong, D., Sha, J., Gao, L., Liu, L., Zhang, Y., Rong, X., 2022e. Spectral-spatial-temporal transformers for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 60, 1–14. <https://doi.org/10.1109/TGRS.2022.3203075>.
- Wang, G., Li, B., Zhang, T., Zhang, S., 2022b. A network combining a transformer and a convolutional neural network for remote sensing image change detection. *Remote Sens. (Basel)* 14 (9), 2228. <https://doi.org/10.3390/rs14092228>.
- Wang, J.-X., Li, T., Chen, S.-B., Tang, J., Luo, B., Wilson, R.C., 2022c. Reliable contrastive learning for semi-supervised change detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*. 60, 1–13. <https://doi.org/10.1109/TGRS.2022.3228016>.
- Wang, J., Liu, F., Wang, H., Liu, X., Jiao, L., Yang, H., Li, L., Chen, P., 2023a. SDCDNet: A Semi-Dual Change Detection Network Framework with Super-Weak Label for Remote Sensing Image. *IEEE Transactions on Geoscience and Remote Sensing*. 61 (1–14). <https://doi.org/10.1109/TGRS.2023.3286113>.
- Wang, W., Tan, X., Zhang, P., Wang, X., 2022d. A CBAM based multiscale transformer fusion approach for remote sensing image change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 15, 6817–6825. <https://doi.org/10.1109/JSTARS.2022.3198517>.
- Wang, L., Wang, L., Wang, Q., Atkinson, P.M., 2021b. SSA-SiamNet: Spectral-spatial-wise attention-based Siamese network for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 60, 1–18. <https://doi.org/10.1109/TGRS.2021.3095899>.
- Wang, R., Wu, H., Qiu, H., Wang, F., Liu, X., Cheng, X., 2023d. A Difference Enhanced Neural Network for Semantic Change Detection of Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/LGRS.2023.3310676>.
- Wang, Q., Yuan, Z., Du, Q., Li, X., 2018. GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 57 (1), 3–13. <https://doi.org/10.1109/TGRS.2018.2849692>.
- Wang, D., Zhang, Q., Xu, Y., Zhang, J., Du, B., Tao, D., Zhang, L., 2022a. Advancing plain vision transformer toward remote sensing foundation model. *IEEE Transactions on Geoscience and Remote Sensing*. 61, 1–15. <https://doi.org/10.1109/TGRS.2022.3222818>.
- Wang, L., Zhang, M., Shi, W., 2023b. CS-WSCDNet: Class Activation Mapping and Segment Anything Model-Based Framework for Weakly Supervised Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 61, 1–12. <https://doi.org/10.1109/TGRS.2023.3330479>.
- Wang, L., Zhang, M., Shi, W., 2023c. STCRNet: A Semi-Supervised Network Based on Self-Training and Consistency Regularization for Change Detection in VHR Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 17, 2272–2282. <https://doi.org/10.1109/JSTARS.2023.3345017>.
- Wang, L., Zhang, M., Gao, X., Shi, W., 2024. Advances and Challenges in Deep Learning-Based Change Detection for Remote Sensing Images: A Review through Various Learning Paradigms. *Remote Sens. (Basel)* 16 (5), 804. <https://doi.org/10.3390/rs16050804>.
- Wei, H., Chen, R., Yu, C., Yang, H., An, S., 2022. BASNet: A Boundary-Aware Siamese Network for Accurate Remote-Sensing Change Detection. *IEEE Geosci. and Remote Sens. Lett.* 19, 1–5. <https://doi.org/10.1109/LGRS.2021.3119885>.
- Wu, C., Du, B., Cui, X., Zhang, L., 2017a. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. *Remote Sens. Environ.* 199, 241–255. <https://doi.org/10.1016/j.rse.2017.07.009>.
- Wu, C., Zhang, L., Du, B., 2017b. Kernel slow feature analysis for scene change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 55 (4), 2367–2384. <https://doi.org/10.1109/TGRS.2016.2642125>.
- Wu, C., Du, B., Zhang, L., 2023. Fully convolutional change detection framework with generative adversarial network for unsupervised, weakly supervised and regional supervised change detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 45 (8), 9774–9788. <https://doi.org/10.1109/TPAMI.2023.3237896>.
- Wu, C., Zhang, L., Du, B., Chen, H., Wang, J., Zhong, H., 2024. UNet-Like Remote Sensing Change Detection: A review of current models and research directions. *IEEE Geoscience and Remote Sensing Magazine*. <https://doi.org/10.1109/MGRS.2024.3412770>.
- Xiong Z., Wang Y., Zhang F., Stewart A. J., Hanna J., Borth D., Papoutsis I., Saux B. L., Camps-Valls G., Zhu X. X., 2024. Neural plasticity-inspired foundation model for observing the earth crossing modalities. *arXiv preprint arXiv:15356*. Doi: 10.48550/arXiv.2403.15356.
- Xu, Q., Chen, K., Sun, X., Zhang, Y., Li, H., Xu, G., 2020. Pseudo-Siamese capsule network for aerial remote sensing images change detection. *IEEE Geoscience and Remote Sensing Letters*. 19, 1–5. <https://doi.org/10.1109/LGRS.2020.3022512>.
- Xu, L., Jing, W., Song, H., Chen, G., 2019. High-resolution remote sensing image change detection combined with pixel-level and object-level. *IEEE Access* 7, 78909–78918. <https://doi.org/10.1109/ACCESS.2019.2922839>.
- Yan, T., Wan, Z., Zhang, P., 2022. Fully transformer network for change detection of remote sensing images. In: *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 1691–1708.
- Yan, T., Wan, Z., Zhang, P., Cheng, G., Lu, H., 2023. TransY-Net: Learning Fully Transformer Networks for Change Detection of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*. 61 (1–22). <https://doi.org/10.1109/TGRS.2023.3327253>.
- Yang, L., Chen, Y., Song, S., Li, F., Huang, G., 2021b. Deep Siamese Networks Based Change Detection with Remote Sensing Images. *Remote Sens. (Basel)* 13 (17), 3394. <https://doi.org/10.3390/rs13173394>.
- Yang, Y., Tang, X., Ma, J., Zhang, X., Pei, S., Jiao, L., 2024. ECPS: Cross Pseudo Supervision Based on Ensemble Learning for Semi-Supervised Remote Sensing Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 62, 1–17. <https://doi.org/10.1109/TGRS.2024.3370236>.
- Yang, K., Xia, G.-S., Liu, Z., Du, B., Yang, W., Pelillo, M., Zhang, L., 2021a. Asymmetric siamese networks for semantic change detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*. 60, 1–18. <https://doi.org/10.1109/TGRS.2021.3113912>.
- Yu W., Zhang X., Das S., Zhu X. X., Ghamisi P., 2024a. MaskCD: A Remote Sensing Change Detection Network Based on Mask Classification. *arXiv preprint arXiv:2404.12081*. Doi: 10.48550/arXiv.2404.12081.
- Yu W., Zhang X., Zhu X. X., Gloaguen R., Ghamisi P., 2024b. MineNetCD: A Benchmark for Global Mining Change Detection on Remote Sensing Imagery. *arXiv preprint arXiv:2407.03971*. Doi: 10.48550/arXiv.2407.03971.
- Yu Z., Liu C., Liu L., Shi Z., Zou Z., 2024d. MetaEarth: A Generative Foundation Model for Global-Scale Remote Sensing Image Generation. *arXiv preprint arXiv:13570*. Doi: 10.48550/arXiv.2405.13570.
- Yu, W., Zhuo, L., Ji, J., 2024c. GCFormer: Global Context-aware Transformer for Remote Sensing Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 62, 1–12. <https://doi.org/10.1109/TGRS.2024.3381738>.
- Yuan, P., Zhao, Q., Zhao, X., Wang, X., Long, X., Zheng, Y., 2022. A transformer-based Siamese network and an open optical dataset for semantic change detection of remote sensing images. *Int. J. Digital Earth* 15 (1), 1506–1525. <https://doi.org/10.1080/17538947.2022.2111470>.
- Yuan, S., Zhong, R., Yang, C., Li, Q., Dong, Y., 2024. Dynamically updated semi-supervised change detection network combining cross-supervision and screening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*. 62, 1–14. <https://doi.org/10.1109/TGRS.2024.3369059>.
- Zhan T., Zhu Y., Lan J., Dang Q., 2024. Cross-Domain Separable Translation Network for Multimodal Image Change Detection. *arXiv preprint arXiv:16158*. Doi: 10.48550/arXiv.2407.16158.
- Zhan, Y., Fu, K., Yan, M., Sun, X., Wang, H., Qiu, X., 2017. Change detection based on deep siamese convolutional network for optical aerial images. *IEEE Geoscience and Remote Sensing Letters*. 14 (10), 1845–1849. <https://doi.org/10.1109/LGRS.2017.2738149>.
- Zhang, H., Chen, H., Zhou, C., Chen, K., Liu, C., Zou, Z., Shi, Z., 2024a. Bifa: Remote sensing image change detection with bitemporal feature alignment. *IEEE Transactions on Geoscience and Remote Sensing*. 62, 1–17. <https://doi.org/10.1109/TGRS.2024.3376673>.
- Zhang, X., Cheng, S., Wang, L., Li, H., 2023e. Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 61, 1–15. <https://doi.org/10.1109/TGRS.2023.3245674>.
- Zhang H., Chen K., Liu C., Chen H., Zou Z., Shi Z., 2024b. CDMamba: Remote Sensing Image Change Detection with Mamba. *arXiv preprint arXiv:04207*. Doi: 10.48550/arXiv.2406.04207.
- Zhang, C., Feng, Y., Hu, L., Tapete, D., Pan, L., Liang, Z., Cigna, F., Yue, P., 2022a. A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*. 109, 102769. <https://doi.org/10.1016/j.jag.2022.102769>.
- Zhang, H., Gong, M., Zhang, P., Su, L., Shi, J., 2016. Feature-level change detection using deep representation and feature change analysis for multispectral imagery. *IEEE Geosci. and Remote Sens. Lett.* 13 (11), 1666–1670. <https://doi.org/10.1109/LGRS.2016.2601930>.
- Zhang, X., He, L., Qin, K., Dang, Q., Si, H., Tang, X., Jiao, L., 2022d. SMD-Net: Siamese multi-scale difference-enhancement network for change detection in remote sensing. *Remote Sens. (Basel)* 14 (7), 1580. <https://doi.org/10.3390/rs14071580>.

- Zhang, X., Huang, X., Li, J., 2023f. Joint Self-training and Rebalanced Consistency Learning for Semi-supervised Change Detection. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–13. <https://doi.org/10.1109/TGRS.2023.3314452>.
- Zhang, X., Huang, X., Li, J., 2023g. Semisupervised Change Detection With Feature-Prediction Alignment. *IEEE Transactions on Geoscience Remote Sensing*. 61, 1–16. <https://doi.org/10.1109/TGRS.2023.3247605>.
- Zhang, W., Lu, X., Li, X., 2018. A coarse-to-fine semi-supervised change detection for multispectral images. *IEEE Transactions on Geoscience Remote Sensing*. 56 (6), 3587–3599. <https://doi.org/10.1109/TGRS.2018.2802785>.
- Zhang, W., Lu, X., 2019. The spectral-spatial joint learning for change detection in multispectral imagery. *Remote Sens. (Basel)* 11 (3), 240. <https://doi.org/10.3390/rs11030240>.
- Zhang, J., Pan, B., Zhang, Y., Liu, Z., Zheng, X., 2022c. Building Change Detection in Remote Sensing Images Based on Dual Multi-Scale Attention. *Remote Sens. (Basel)* 14 (21), 5405. <https://doi.org/10.3390/rs14215405>.
- Zhang, Y., Peng, D., Huang, X., 2017. Object-based change detection for VHR images based on multiscale uncertainty analysis. *IEEE Geoscience and Remote Sensing Letters*. 15 (1), 13–17. <https://doi.org/10.1109/LGRS.2017.2763182>.
- Zhang, J., Shao, Z., Ding, Q., Huang, X., Wang, Y., Zhou, X., Li, D., 2023a. AERNet: An Attention-Guided Edge Refinement Network and a Dataset for Remote Sensing Building Change Detection. *IEEE Trans. Geosci. Remote Sens.* 61, 1–16. <https://doi.org/10.1109/TGRS.2023.3300533>.
- Zhang, M., Shi, W., 2020. A feature difference convolutional neural network-based change detection method. *IEEE Transactions on Geoscience and Remote Sensing*. 58 (10), 7232–7246. <https://doi.org/10.1109/TGRS.2020.2981051>.
- Zhang, M., Liu, Z., Feng, J., Liu, L., Jiao, L., 2023c. Remote Sensing Image Change Detection Based on Deep Multi-Scale Multi-Attention Siamese Transformer Network. *Remote Sens. (Basel)* 15 (3), 842. <https://doi.org/10.3390/rs15030842>.
- Zhang, C., Wang, L., Cheng, S., Li, Y., 2022b. SwinSUNet: Pure transformer network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 60, 1–13. <https://doi.org/10.1109/TGRS.2022.3160007>.
- Zhang, Y., Wang, F., Li, Y., OuYang, S., Wei, D., Liu, X., Kong, D., Chen, R., Zhang, B., 2023i. Remote sensing knowledge graph construction and its application in typical scenarios. *Natl. Remote Sens. Bull.* 27, 249–266. <https://doi.org/10.11834/jrs.20210469>.
- Zhang, M.Y., Xu, G.L., Chen, K.M., Yan, M.J., Sun, X., 2019. Triplet-Based Semantic Relation Learning for Aerial Remote Sensing Image Change Detection. *IEEE Geosci. Remote Sens. Lett.* 16 (2), 266–270. <https://doi.org/10.1109/LGRS.2018.2869608>.
- Zhang, X., Yu, W., Pun, M.-O., Shi, W., 2023h. Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 197, 1–17. <https://doi.org/10.1016/j.isprsjprs.2023.01.018>.
- Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., Liu, G., 2020. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*. 166, 183–200. <https://doi.org/10.1016/j.isprsjprs.2020.06.003>.
- Zhang, R., Zhang, H., Ning, X., Huang, X., Wang, J., Cui, W., 2023d. Global-aware siamese network for change detection on remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*. 199, 61–72. <https://doi.org/10.1016/j.isprsjprs.2023.04.001>.
- Zhang, K., Zhao, X., Zhang, F., Ding, L., Sun, J., Bruzzone, L., 2023b. Relation Changes Matter: Cross-Temporal Difference Transformer for Change Detection in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*. 61, 1–15. <https://doi.org/10.1109/TGRS.2023.3281711>.
- Zhao, Z., Ru, L., Wu, C., 2023c. Exploring Effective Priors and Efficient Models for Weakly-Supervised Change Detection. *arXiv preprint arXiv:2307.10853*. Doi: 10.48550/arXiv.2307.10853.
- Zhao, L., Huang, Z., Kuang, D., Peng, C., Gan, J., Li, H., 2024a. SeFi-CD: A Semantic First Change Detection Paradigm That Can Detect Any Change You Want. *arXiv preprint arXiv:09874*. Doi: 10.48550/arXiv.2407.09874.
- Zhao, M., Hu, X., Zhang, L., Meng, Q., Chen, Y., Bruzzone, L., 2024. Beyond Pixel-Level Annotation: Exploring Self-Supervised Learning for Change Detection With Image-Level Supervision. *IEEE Transactions on Geoscience and Remote Sensing*. 62, 1–16. <https://doi.org/10.1109/TGRS.2024.3379431>.
- Zhao, J., Jiao, L., Wang, C., Liu, X., Liu, F., Li, L., Yang, S., 2023a. GeoFormer: A Geometric Representation Transformer for Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 61, 1–17. <https://doi.org/10.1109/TGRS.2023.3331751>.
- Zhao, S., Zhang, X., Xiao, P., He, G., 2023b. Exchanging Dual-Encoder-Decoder: A New Strategy for Change Detection With Semantic Guidance and Spatial Localization. *IEEE Transactions on Geoscience and Remote Sensing*. 61, 1–16. <https://doi.org/10.1109/TGRS.2023.3327780>.
- Zheng, Z., Liu, Y., Tian, S., Wang, J., Ma, A., Zhong, Y., 2021. Weakly supervised semantic change detection via label refinement framework. In: 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). pp. 2066–2069. <https://doi.org/10.1109/IGARSS47720.2021.9553768>.
- Zheng, Z., Tian, S., Ma, A., Zhang, L., Zhong, Y., 2023. Scalable multi-temporal remote sensing change data generation via simulating stochastic change process. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 21818–21827.
- Zheng, Z., Ermon, S., Kim, D., Zhang, L., Zhong, Y., 2024a. Changen2: Multi-temporal remote sensing generative change foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2024.3475824>.
- Zheng, Z., Zhong, Y., Zhang, L., Ermon, S., 2024b. Segment any change. *arXiv preprint arXiv:01188*. Doi: 10.48550/arXiv.2402.01188.
- Zhou, Y., Wang, J., Ding, J., Liu, B., Weng, N., Xiao, H., 2023. SIGNet: A Siamese Graph Convolutional Network for Multi-Class Urban Change Detection. *Remote Sens. (Basel)* 15 (9), 2464. <https://doi.org/10.3390/rs15092464>.
- Zhu, Y., Li, L., Chen, K., Liu, C., Zhou, F., Shi, Z., 2024. Semantic-CC: Boosting Remote Sensing Image Change Captioning via Foundational Knowledge and Semantic Guidance. *arXiv preprint arXiv:2407.14032*. Doi: 10.48550/arXiv.2407.14032.
- Zhu, Q., Guo, X., Li, Z., Li, D., 2022. A review of multi-class change detection for satellite remote sensing imagery. *Geo-spatial Inf. Sci.* 1–15. <https://doi.org/10.1080/10095020.2022.2128902>.
- Zhu, Y., Lv, K., Yu, Y., Xu, W., 2023b. Edge-Guided Parallel Network for VHR Remote Sensing Image Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 16, 7791–7803. <https://doi.org/10.1109/JSTARS.2023.3306274>.
- Zhu, P., Xu, H., Luo, X., 2023a. MDAFormer: Multi-level difference aggregation transformer for change detection of VHR optical imagery. *International Journal of Applied Earth Observation and Geoinformation*. 118, 103256. <https://doi.org/10.1016/j.jag.2023.103256>.
- Zou, C., Wang, Z., 2023. A New Semi-Supervised Method for Detecting Semantic Changes in Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*. 20, 1–5. <https://doi.org/10.1109/LGRS.2023.3311106>.