remote sensing

MDPI

Review

Deep-Learning for Change Detection Using Multi-Modal Fusion of Remote Sensing Images: A Review

Souad Saidi [1], Soufiane Idbraim [1], Younes Karmoude [1], Antoine Masse [2] and Manuel Arbelo [3,*]

[1] IRF-SIC (Image et Reconnaissance de Formes–Systèmes Intelligents et Communicants) Laboratory, Faculty of Science Agadir, Ibn Zohr University, Agadir 80 000, Morocco; souad.saidi@edu.uiz.ac.ma (S.S.); s.idbraim@uiz.ac.ma (S.I.); younes.karmoude@edu.uiz.ac.ma (Y.K.)
[2] IGNFI (Institut Géographique National France International), 7 rue Biscornet, 75012 Paris, France; amasse@ignfi.fr
[3] Departamento de Física, Universidad de La Laguna, 38206 San Cristobal de La Laguna, Spain
* Correspondence: marbelo@ull.es

Abstract: Remote sensing images provide a valuable way to observe the Earth's surface and identify objects from a satellite or airborne perspective. Researchers can gain a more comprehensive understanding of the Earth's surface by using a variety of heterogeneous data sources, including multispectral, hyperspectral, radar, and multitemporal imagery. This abundance of different information over a specified area offers an opportunity to significantly improve change detection tasks by merging or fusing these sources. This review explores the application of deep learning for change detection in remote sensing imagery, encompassing both homogeneous and heterogeneous scenes. It delves into publicly available datasets specifically designed for this task, analyzes selected deep learning models employed for change detection, and explores current challenges and trends in the field, concluding with a look towards potential future developments.

Keywords: change detection; deep learning; remote sensing images; data fusion; multi-source; multi-sensor; multi-modal

check for updates

Citation: Saidi, S.; Idbraim, S.; Karmoude, Y.; Masse, A.; Arbelo, M. Deep-Learning for Change Detection Using Multi-Modal Fusion of Remote Sensing Images: A Review. *Remote Sens.* **2024**, *16*, 3852. https://doi.org/10.3390/rs16203852

Academic Editors: Byeungwoo Jeon, Yuhui Zheng, Guoqing Zhang and Le Sun

Received: 5 September 2024
Revised: 10 October 2024
Accepted: 11 October 2024
Published: 17 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Remote sensing captures Earth's surface data without direct contact. It employs sensors on satellites, airplanes, drones, or ground-based devices [1]. This non-invasive technique has significantly influenced geography, geology, agriculture, and environmental management [2]. It aids in investigating natural resources, the environment, and weather patterns, enabling informed decision-making for long-term growth [3].

With advancements in remote sensing technology, collecting diverse images using various sensors is now feasible, which improves our ability to analyze the Earth's surface. Thus, our review focuses on exploring deep learning (DL) methods for change detection through the fusion of multi-source remote sensing data, emphasizing their role in integrating information from different sensors. The optical remote sensing systems, including high-resolution (VHR) images and multispectral data, provide detailed views and valuable information for applications such as urban planning [4] and land cover mapping. VHR optical offers excellent spatial resolution, while multispectral images enable analysis of vegetation health [5], water quality [6], and mineral exploration [7]. On the other hand, microwave remote sensing systems, particularly synthetic aperture radar (SAR), offer several unique advantages: penetrating cloud cover and providing data regardless of weather conditions or daylight [8].

Each system has its limitations. VHR optical images, despite their high spatial resolution, have a limited range of spectral bands, restricting their applicability in studies that require a broader spectrum of wavelengths [9]. Multispectral images are sensitive to

atmospheric interference and cloud cover, which can significantly impact data accuracy. Moreover, due to the coherent structure of radar waves, SAR data frequently experience speckle noise [10]. Relying on a single data type can result in incomplete or biased insights, as each sensor captures different aspects of the observed environment. Also, in long-term studies, such as monitoring deforestation over decades, depending on a single dataset becomes impractical. For example, using only one dataset would fail to provide the necessary temporal depth, requiring multi-source data like combining Landsat (30 m, 1989) and Sentinel-2 (10 m, 2024).

To overcome these limitations, multi-source data fusion has emerged as a vital technique. It combines complementary information from multiple sensors to create a more complete and reliable representation of the target area. Multi-source fusion improves robustness by combining data from optical, SAR, LiDAR, and hyperspectral images, providing more detailed feature sets. This can improve the accuracy of land classification and object detection [11].

One of the applications of multi-source data fusion in remote sensing is change detection, which is the process of identifying and analyzing differences in the state of an object or phenomenon by comparing images at different times. This technique is essential for monitoring transformations in various fields, including urban planning [12], environmental monitoring [13], and disaster management [14]. Multi-source data fusion provides a robust approach to change detection, where detecting and analyzing changes between images taken at different times is crucial. By combining data from diverse remote sensing modalities, we can improve the adaptability and precision of change detection, facilitating the discrimination of diverse change patterns [15].

Over the last several decades, researchers have developed numerous change detection methods. Before deep learning, pixel-based classification methods [16–23] progressed significantly. Most traditional approaches focused on identifying changed pixels and classifying them to create change maps. Despite achieving notable performance on certain image types, these methods frequently encountered limitations regarding accuracy and generalization. Furthermore, their performance was dependent on the classifier and threshold parameters used [24]. Few studies have focused on applying multi-source data fusion for the change detection task [16,20–23].

In recent years, deep learning has revolutionized change detection tasks, primarily when used for homogeneous data. For images acquired from the same sensor type (e.g., optical-to-optical or SAR-to-SAR), DL models like convolutional neural networks (CNNs) [25], recurrent neural networks (RNNs) [26], and generative adversarial networks (GANs) [27] have significantly outperformed traditional methods. These models excel at automatically extracting hierarchical features from raw data, eliminating the need for manual feature engineering.

Moreover, deep learning techniques have made substantial strides in change detection using multi-modal data fusion, enabling the effective integration of diverse remote sensing data [28] and allowing researchers to gain a comprehensive and accurate understanding of land cover, environmental changes, and natural disasters. By seamlessly integrating diverse data sources, DL simplifies the creation of detailed depictions of what's happening on Earth. It also enhances the efficiency and precision of fusing remote sensing data, contributing to improved decision-making, environmental monitoring, and land management practices. The capability to automatically extract meaningful insights from heterogeneous data sources is a notable advancement in remote sensing data fusion.

Our state-of-the-art review distinguishes itself from the others carried out so far (between 2022 and 2023) [29–32]. While these studies focused on classification based on the type of classification (supervised, unsupervised, or semi-supervised) [29], the type of deep learning model used (CNN, RNN, GAN, transformer, etc.) [30], the level of analysis (scene, region, super-pixel) [31], and the class of the model (UNet and non-UNet) [32]. However, our approach stands out by being based on the nature of the satellite data available to the user. We consider the difference between homogeneous and heterogeneous data, such as optical

data at different scales or multi-modal data combining optical and SAR (Synthetic Aperture Radar). This review guides users towards the most suitable approaches for their data. It offers specific recommendations for multi-scale optical data (e.g., Landsat, Sentinel, WorldView-2) and multi-modal optical-SAR data (e.g., Sentinel-1A, Sentinel-2A).

This review is structured as follows: Section 2 outlines the literature review methodology used to collect articles for this review, including the search strategy and selection criteria. Section 3 presents the key findings obtained through statistical analysis of the data. Section 4 explores various multi-modal datasets used in remote sensing change detection, discussing their quality and limitations. Section 5 examines approaches and techniques used for multi-modal data fusion to enhance change detection accuracy. Section 6 raises future research trends and discussions. Finally, Section 7 concludes this survey. By adopting this review structure, we aim to provide a comprehensive and accessible resource illustrating the transformative potential of data fusion in remote sensing. It presents valuable insights for researchers and practitioners alike.

2. Literature Review Methodology

This section outlines the comprehensive search strategy and rigorous study selection process employed to identify relevant literature for this survey. It uses the Preferred Reporting Items for Systematic and Meta-Analyses (PRISMA) [33].

2.1. Search Strategy

This study collected articles from three high-impact online databases: Web of Science, IEEE Xplore, and Science Direct. We selected these databases as our primary sources due to their comprehensive coverage of scholarly literature in engineering, technology, and applied sciences. Each database is well-regarded for its extensive collection of peer-reviewed articles, conference papers, and research studies. It makes them suitable resources for identifying relevant literature concerning deep learning and change detection. Our search strategy incorporated specific keywords to target relevant studies within deep learning and change detection. The search utilized keywords such as "data fusion", "deep learning", "remote sensing", "neural networks", "multimodal", "multisource", "optical and SAR", "homogeneous", "heterogeneous", and "Change Detection", which were carefully chosen to capture the essential concepts and methodologies associated with this research area. To maximize the search results, we combined these terms using Boolean operators (AND, OR). We also employed truncation and wildcards to capture variations of the keywords. The specific search strings used were as follows:

- ("data fusion" OR "multisource" OR "multimodal") AND ("deep learning" OR "neural networks") AND ("remote sensing" OR "satellite images") AND ("change detection").
- ("homogeneous" OR "heterogeneous") AND ("deep learning" OR "neural networks") AND ("remote sensing" OR "satellite images").
- ("optical and SAR") AND ("deep learning" OR "neural networks") AND ("remote sensing" OR "satellite images") AND ("change detection").

Additionally, we applied search filters to narrow down the results based on publication date, restricting them to articles published from 2017 to 2024. Furthermore, we limited the search to English-language articles to ensure consistency in language comprehension.

2.2. Study Selection

We collected 160 search records from various search engines, including 70 from IEEE Xplore, 50 from the Web of Science, 20 from ScienceDirect, and 20 from other sources. Initially, we checked the article titles to remove duplicates. After that, we reviewed the abstracts of the collected publications to select the most relevant ones based on the alignment with our research focus, methodological strength, knowledge contribution, evidence quality, and potential study impact. We then thoroughly examined the full text of the selected articles and applied exclusion criteria (Figure 1) to filter them. In the end, we have

120 studies left for analysis in this survey. A complete selection process flow is shown in (Figure 1).

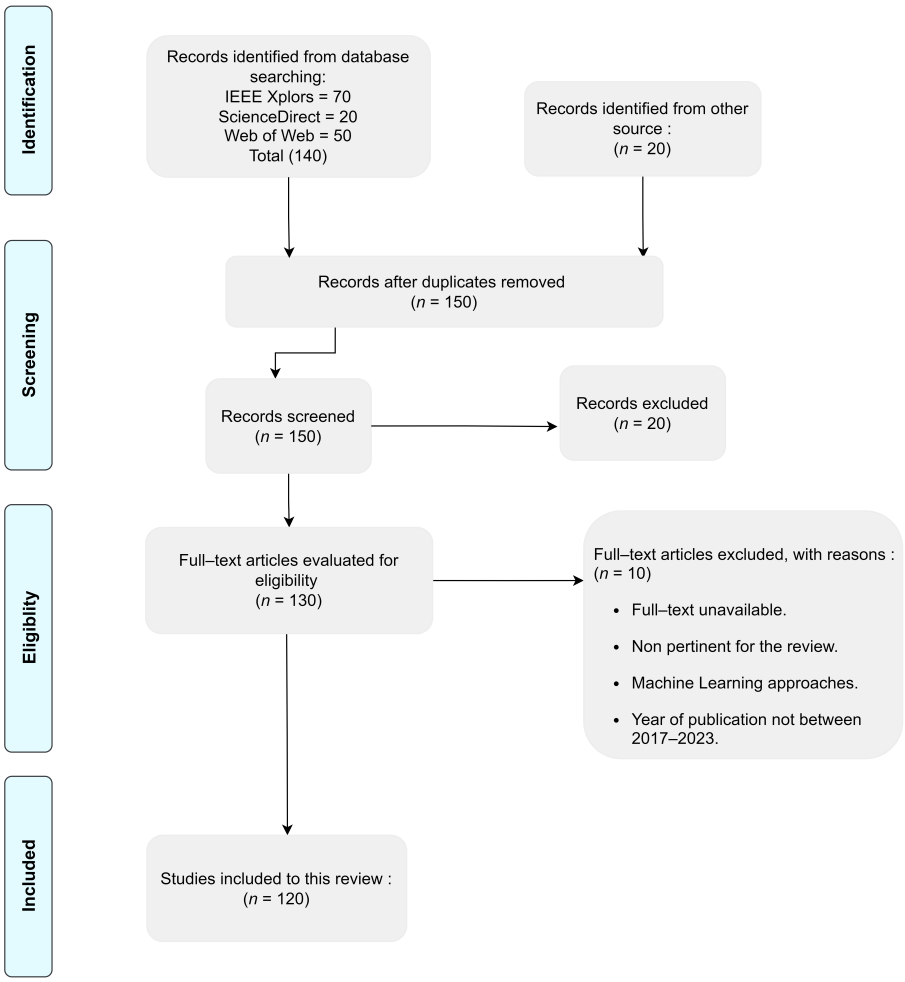


Figure 1. PRISMA flow diagram.

3. Statistical Analysis and Results

This section presents a comprehensive analysis of publication trends in deep learning for both homogeneous and heterogeneous remote sensing change detection (RSCD). The first is the histogram of scientific production in RSCD using DL over the years. The second is the leading journals and publishers contributing to this field. Finally, we examine the global distribution of these research publications.

Figure 2 depicts the publication history from 2017 to 2024. Before 2022, the field witnessed a modest research output. However, a pivotal shift towards DL applications in change detection emerged in 2022, with a significant increase in published articles (31 papers). Several factors have contributed to this surge. The emergence of transformers and hybrid models, along with advancements in deep learning algorithms, has played a significant role. Additionally, increased access to high-resolution remote sensing data and a growing interest in tackling complex change detection challenges.

We also identify journals with a consistent publication record of articles relevant to our research focus. These journals constitute a platform for sharing cutting-edge knowledge, making them indispensable resources for researchers, academics, and professionals. A concise summary is provided in Table 1, presenting essential data related to these journals.

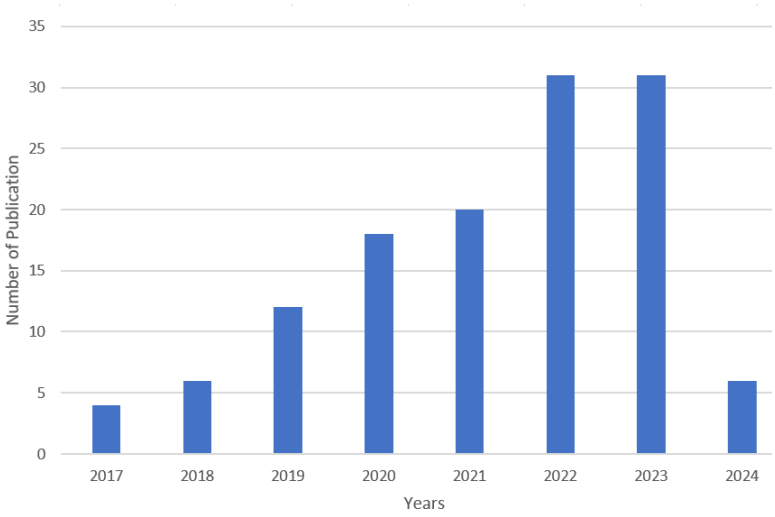


Figure 2. Year-wise publications from 2017 to 2024.

Table 1. The most productive journals.

Journal Name	Total Publications	Impact Factor (2023)	Publisher	Cite Score (2023)
IEEE Transactions on Geoscience and Remote Sensing	28	8.2	IEEE	10.9
Remote Sensing	25	5	MDPI	7.9
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	17	5.5	IEEE	7.8
IEEE Geoscience and Remote Sensing Letters	9	4.8	IEEE	6.4
ISPRS Journal of Photogrammetry and Remote Sensing	9	12.7	Elsevier	19.2
IEEE Transactions on Neural Networks and Learning Systems	4	10.4	IEEE	21.9
International Journal of Remote Sensing	3	3.5	Taylor & Francis	6.5

Furthermore, we explore the global distribution of publications, highlighting China's strong influence in the field, contributing a substantial number of 70 publications. However, other countries such as France, Germany, Japan, and the United Kingdom also contribute to the research effort. Each of these countries publishes numerous articles. A map visualizes (Figure 3) these various study origins, emphasizing the international nature of the research discipline. These findings collectively provide a comprehensive overview of the field's dynamics and worldwide impact.

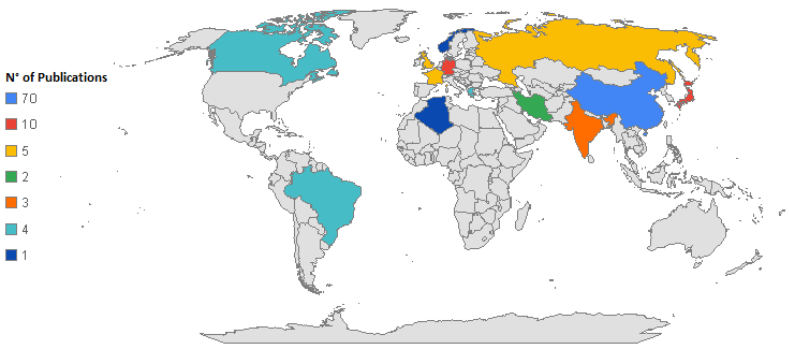


Figure 3. Global distribution of publications.

4. Multi-modal Datasets

Datasets are crucial for the performance of DL models. They influence accuracy, which measures prediction success. They also affect efficiency, reflecting speed and resource usage. High-quality datasets improve the model's reliability and enhance its capacity to generalize to new data [34]. This section delves into three critical categories of remote sensing datasets used in DL applications: single-source, multi-source, and multi-sensor data. Each of these dataset categories presents its own unique challenges and opportunities. A summary of available datasets for each category is provided in Table 2.

Table 2. Multi-modal remote sensing datasets.

	DataSet	Data Type	Resolution (m)	Satellite Types
Single Source	OSCD [35]	Optical	10-20-60	Sentinel 2
	Lake overflow [36]	Optical	30	Landsat 5(NIR/RGB)
	Farmland [37]	SAR	3	Radarsat-2 (single/four look)
	CLCD	Optical	0.5 to 2	Gaofen-2
	LEVIR-CD [38]	Optical	0.5	Google Earth
	WHU-CD [39]	Optical	0.3	QuickBird/Worldview
Multi-sensor	Wang, M [40]	Optical	5.8/4	ZY-3/GF-2
	S2Looking [41]	Optical	0.5/0.8	GF, SV, and BJ-2
	CCD [42]	Optical	0.03/1	Google Earth
	MRCDD	Optical	0.5/2	Google Earth
	Mengxi Liu [43]	Optical	4/1	Google Earth
	Bastrop [44]	Optical	30	Landsat-5/EO-1 ALI
	Saha, S [45]	Optical	0.5/0.6	Quickbird/Pleiades
	Reunion	Optical	10/2	Sentinel-2/SPOT6/7
	EV-CD building [46]	Optical	0.2/2	Variety of sensors
Multi-source	HTCD [47]	UAV/Optical	0.5971/0.07	Google Earth/Open Aerial Map
	MSBC	Optical/SAR	2/20	GF-2/Sentinel1-2A
	MSOSCD	Optical/SAR	-	Sentinel 2/Google Earth
	Hunan [48]	Optical/SAR	10/30	Sentinel-1/2, SRTM
	DFC2020 [49]	Optical/SAR	10/20	Sentinel-1/2
	Potsdam [50]	Optical/LiDar	0.05	-
	California dataset [51]	Optical/SAR	20/30	Landsat 8/Sentinel-1A
	Houston2018 [52]	HS/LiDAR/RGB	0.5/1	ITRES CASI 1500/Titan MW
	Berlin data [53]	HS/SAR	13.89	HyMap HS/Sentinel-1
	MUUFL Gulfport [54]	HS/LiDAR	0.54/1	-
	Gloucester I [55]	Optical/SAR	0.65	QuickBird 2/TerraSAR-X
	Gloucester II [55]	Optical/SAR	≈25	SPOT/ERS-1

4.1. Single-Source Data

Single-source data are collected from a single sensor of a specific kind, most commonly optical sensors, renowned for their ability to capture high-resolution imagery of the Earth's surface [56]. When we talk about single-source data, we are referring to scenarios where the entirety of the information originates from a sole sensor. A prime example is the Onera Satellite Change Detection (OSCD) dataset [35] from Sentinel-2, which provides optical imagery at resolutions of 10-20-60 m. Another example is the Lake Overflow dataset [36] from Landsat 5, which captures optical data in NIR/RGB bands at a 30 m resolution. Finally, the Farmland dataset [37] from Radarsat-2 offers SAR data at a 3 m resolution.

4.2. Multi-Sensor Data

Multi-sensor data typically involve using data from multiple sensors of the same modality. Each sensor has its own set of characteristics and capabilities. These sensors share a common purpose, such as capturing optical imagery, but they differ in specifications like spatial resolution, spectral bands, or radiometric sensitivity. By combining data from these diverse sensors, we access an extensive quantity of information that improves our understanding of

the target area or phenomenon [57]. For instance, the Bastrop dataset [44] focuses on observing the effects of forest fires in Bastrop County, Texas, USA, using pre-event images from Landsat-5 and post-event images from EO-1 ALI. Another example is the S2Looking dataset [41], a collection of data spanning the years 2017 to 2020 from various satellites, including GaoFen (GF), SuperView (SV), and Beijing-2 (BJ-2), specifically designed for satellite-side-looking change detection.

4.3. Multi-Source Data

Multi-source data typically involve combining data from various sensors and integrating their diverse sources of information. Each sensor type has unique and significant capabilities. Optical sensors are better at detecting visible changes, whereas radar may break through cloud cover to uncover subsurface changes. LiDAR, which uses laser-based technology, provides extremely detailed three-dimensional data, while hyperspectral sensors give a wide range of spectral bands for complete material characterization [58]. Various datasets combine optical and SAR imagery to provide detailed insights into distinct phenomena. For example, the California dataset [51] captures land cover changes caused by floods in 2017, combining SAR images from Sentinel-1A and optical images from Landsat 8. Similarly, the Gloucester I dataset [55] provides a pre- and post-flood image pair acquired by Quickbird 2 and TerraSAR-X, respectively. Other multi-source datasets include the HTCD dataset [47], which focuses on urban change using satellite imagery from Google Earth and UAV imagery. Additionally, the Houston2018 dataset [52] provides a collection of HS-LiDAR-RGB data, offering a comprehensive view of urban environments through the fusion of hyperspectral, LiDAR, and RGB imagery.

4.4. Data Quality and Limitations

The datasets utilized for change detection vary significantly in quality, diversity, and representativeness. Their variability impacts the performance of DL models. Single-source data, typically represented by high-resolution imagery, are invaluable for detecting small-scale land cover changes. However, their limited geographic coverage can restrict the analysis of broader environmental patterns. In contrast, datasets like the OSCD [35,36] offer extensive spatial coverage and diverse spectral bands. Yet, they are impacted by atmospheric conditions. These conditions can lead to potential data quality issues. In addition, datasets such as LEVIR-CD and WHU Building Change Detection provide detailed insights into specific changes. Regardless, their narrow focus may limit generalizability and introduce biases. This limitation can hinder the detection of other significant alterations, such as those related to vegetation or infrastructure.

While multi-sensor datasets offer a wealth of information, they also present challenges. Aligning data from different sources can be complex due to variations in pixel size, which may introduce distortions or require advanced resampling techniques. Furthermore, multi-sensor data often exhibit gaps in temporal and spatial coverage, leading to uneven data availability. For example [40], one dataset might be complete while another has missing data due to cloud cover or other factors. This can necessitate interpolation or imputation methods that can introduce potential biases or inaccuracies.

Finally, multi-source datasets provide a more comprehensive view by leveraging the strengths of both modalities (optical data and radar data). While this approach improves detection accuracy, especially in environments with frequent cloud cover, it presents significant challenges in data fusion. The differences in sensor characteristics (e.g., radar's sensitivity to surface roughness vs. optical spectral information) can lead to misalignment, noise, and inconsistent interpretations. Furthermore, such fusion requires sophisticated models to harmonize these diverse inputs effectively. However, the resulting datasets may still introduce biases depending on the availability and quality of source images across different regions and times.

5. Multi-Modal Data Fusion for Change Detection

This section explores the use of DL for detecting changes in various types of data. We will begin by discussing different fusion techniques. Next, we will examine methods designed for homogeneous data. Finally, we will explore methods developed to address the complexities of heterogeneous change detection.

5.1. Feature Fusion Strategy

Feature fusion is an essential technique in DL, particularly for tasks involving multi-modal inputs. Combining features from several sources or modalities improves the informational content and separable power of the data representation [59], resulting in better overall performance in DL applications. There are three types of feature fusion: early fusion, late fusion, and multiple fusion, as shown in Figure 4.

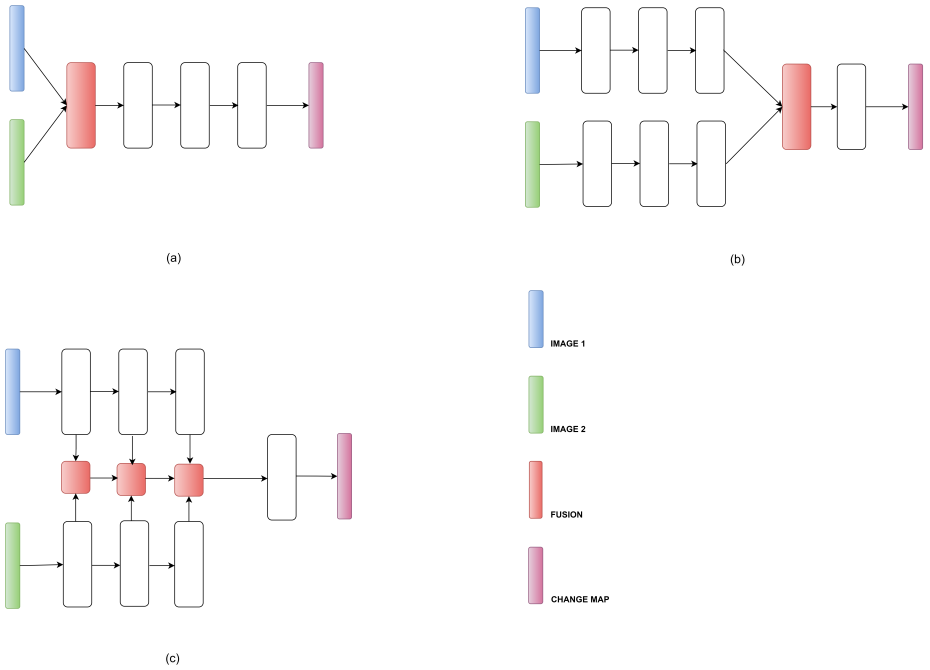


Figure 4. Feature extraction strategy. (**a**) Early fusion; (**b**) Late fusion; (**c**) Multiple fusion.

Early fusion combines features at the input layer, leading to a unified input representation before the DL model processes it, as shown in Figure 4a. Several studies, including [60–62], explore this method. Regardless, early fusion methods may only utilize partial information for the change detection task, potentially impacting the detection performance [63].

Late fusion [64] combines features at the output layer, making a final decision using the outputs of individual models trained on each modality as shown in Figure 4b. This method allows each data source to be processed in a manner suited to its unique characteristics before combining the extracted features. It is especially effective when the data sources are heterogeneous.

Multiple fusion combines features from different stages of the DL model, allowing for deeper information fusion as shown in Figure 4c. Recent studies, such as [12,63,65], have demonstrated that multi-level fusion outperforms both early and late fusion by leveraging the strengths of each. However, it is computationally complex and requires careful tuning. This method is ideal for highly complex change detection tasks, where both broad and detailed changes need to be captured.

Deep learning methods have adapted these fusion strategies through various architectural designs. Early fusion, also known as single-stream networks, often apply CNN architectures such as encoder–decoder models, as illustrated in Figure 5a. While they excel at

capturing the overall context, they may overlook subtle or minor changes. Furthermore, they may struggle when dealing with noisy or irrelevant variations in the input images. Late or multiple fusion usually utilizes Siamese network architectures (Figure 5b,c) [66]. This architecture used separate feature extraction branches with shared or unshared weights, extracting features independently from input images. The network merges after convolutional layers have processed the input. Extracted features are fused using techniques like concatenation or addition in some cases. In other cases, an attention mechanism is employed to focus on informative elements. The fused features are fed into an algorithm that compares them and produces a change map.

Siamese networks have a flexible general structure (Figure 5a) that can accommodate a variety of models, including a Siamese feature extractor, feature fusion, and a decision-making module. This adaptable architecture allows for diverse applications and task flexibility. An alternative approach is a UNet structure (Figure 5b), where the encoder processes each image separately, and fusion occurs through skip connections. This architecture is more adept at managing multi-scale features than a traditional Siamese network. However, it requires more computational resources and memory.

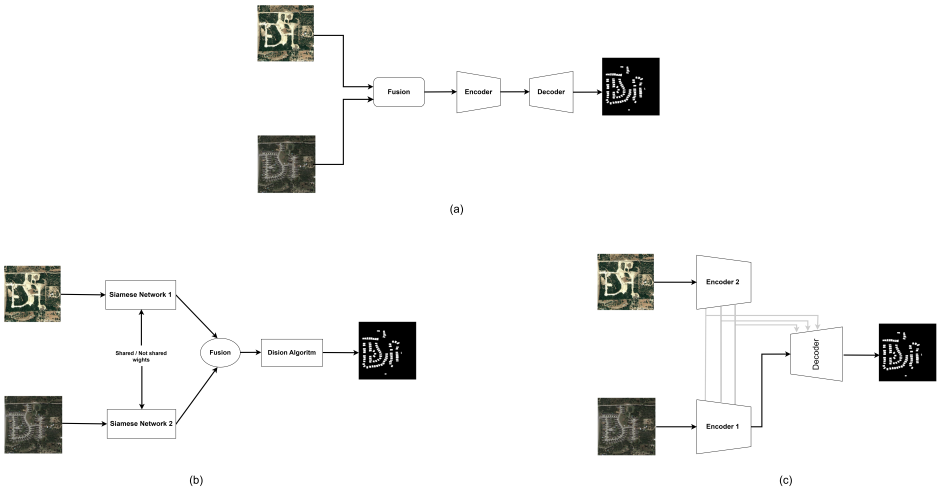


Figure 5. Structures of models. (**a**) Single Stream Network. (**b**) General Siamese network structure. (**c**) Double-Stream UNet.

5.2. Homogeneous-RSCD

The Homogeneous-RSCD (Hom-RSCD) method involves analyzing data from a single sensor. These data could be optical imagery captured by satellites, providing valuable insights into changes in land cover over time. Hom-RSCD has various real-world applications across different fields. For example, it is utilized to monitor deforestation [67–71] to identify areas with reduced forest cover caused by illegal logging or fires. In rapidly urbanizing countries like China, Hom-RSCD is used to track urban expansion and the conversion of agricultural land into urban areas [12,22,72–77]. Additionally, in Bangladesh, it is employed for flood monitoring [78,79] by comparing pre- and post-event imagery to assess flood impacts. These applications highlight the versatility and importance of Hom-RSCD in addressing critical environmental issues.

Several DL approaches are making a powerful impact on Hom-RSCD:

5.2.1. CNN-Based

Standard CNNs

In recent years, CNNs have established themselves as a versatile approach for extracting information from remote sensing images for CD. Recent research has focused on pushing the boundaries even further.

The majority of change detection studies employ double-stream structures as the primary approach. Some studies [60–62,80,81] have explored single-stream architectures, where both input images are processed sequentially through a single network, usually based on UNet. However, these remain less common in contrast with the more widely adopted Siamese network models.

Most articles investigating double-stream methods primarily use Siamese UNets. For instance, DSMS-FCN [82] utilizes a modified convolution unit for extracting multi-scale features and uses change vector analysis to make the changes maps more precise. The FDCNN [83] approach by Zhang et al., which leverages sub-VGG16 for feature extraction and dedicated networks for generating feature difference maps and fusion. The work of [84] is a fully convolutional Siamese network. It employs a modified long skip connection, incorporating concatenated absolute differences and Euclidean distances to enhance the extraction of spatial details. The ESCNet [85] incorporates Siamese networks for pre-processing the input images to extract superpixels (groups of pixels with similar properties). This information is then used to reduce noise and improve edge detection. The RFNet [86] used SE-ResNet50s as the backbone for feature extraction. It includes multiscale feature fusion that fuses features across scales and compares local features to take into account potential spatial offsets between the images. Simillary, The SMD-Net [87] employs a Siamese network (ResNet-34) that includes modules for feature interaction and region-based feature fusion to account for potential misalignments and improve CD accuracy. The Siam-FAUNet [88] utilizes an improved VGG16 encoder, Atrous Spatial Pyramid Pooling (ASPP) for capturing multi-scale context, and a Flow Alignment Module to improve semantic alignment within the network. It specifically addresses issues like blurred change boundaries and missing small targets. SSCFNet [89] emphasizes incorporating both low-level and high-level features. It achieves this via a novel combined enhancement module that constructs semantic feature blocks and a semantic cross-fusion module that utilizes different convolution operations to extract features at various levels. Lately, DETNet [90] utilizes a triplet feature extraction module with a "triple CNN" backbone to extract spatial-spectral features. Additionally, a difference feature learning module analyzes the variations in the learned features to identify subtle changes. While standard UNets offer a strong foundation, advancements like UNet++ make use of several hierarchical and dense skip paths instead of relying solely on links between encoder and decoder networks. Using the difference absolute value operation, [91] enhances the dense skip connection module based on Siamese UNet++ to process features at many scales. The DifUNet++ [92] employs a side-out fusion approach and a differential pyramid of two input images as the input. SNUNet-CD [93] incorporates upsampling modules and strategically placed skip connections between corresponding semantic levels in the encoder and decoder. This approach facilitates a more condensed information transfer within the network. BCD-Net [76] takes another approach, drawing inspiration from full-scale UNet3+ but modifying it with subpixel convolution layers instead of upsampling layers.

Beyond UNets, encoder–decoder methods like [94] demonstrate success by combining early fusion and Siamese modules to extract features from both individual and different images. The SSJLN [95] goes beyond simply combining spectral and spatial information. It actively learns their relationship, refines the fused features for change-specific information, and optimizes the learning process through a tailored loss function. Other methods leverage edge detection for enhanced performance [96,97] by incorporating edge separation and boundary extraction modules within their Siamese networks. The Approbation of Visual Foundation Models in change detection has been the subject of recent research efforts. In their supervised learning model for feature extraction in RS imagery, Ding et al. [98] included FastSAM [99] as an encoder, investigating its possible benefits in semi-supervised CD tasks.

CNNs with Attention Mechanisms

Following the exploration of traditional CNN-based methods for change detection, current studies have increasingly integrated attention mechanisms into the architecture to enhance performance. Attention modules dynamically highlight important features while suppressing irrelevant information. This approach offers significant benefits for improving both spatial and channel-wise feature extraction [100], which is essential for enhancing change detection. For example, AFSNet [101] is adopting a Siamese UNet architecture with a VGG16 as the backbone. Its core strength lies in the enhanced full-scale skip connections that facilitate the fusion of features from different scales. An attention module is inserted between the encoder and decoder to refine side outputs generated at various scales, integrating spatial and channel attention. Thus, Adriano et al. [102] proposed a Siamese UNet that integrated attention gates (AGs) into skip connections. These AGs guide the network to concentrate on pertinent data and filter out irrelevant or noisy regions. The network experienced extensive training on real-world scenarios of emergency disaster response data availability. This training included single-mode, cross-modal, and combined optical and SAR data. Feng et al. [40] proposed a multi-modal conditional random field and a multiscale adaptive kernel network. They used a weight-sharing Siamese encoder for feature extraction and an adaptive convolution kernel block for selective weighting. An attention-based upsampling module in the decoder enhances variation data expression, and multi-modal conditional random fields improve detection results. The HARNU-Net [103] consists of an Improved UNet++ as a Siamese network. It introduces an ACON-Relu Residual Convolutional Block (A-R) structure, a remodeled convolution block, and an adjacent feature fusion module (AFFM). These components work together to integrate multi-level features and context information, improving the regularity of change boundaries. The hierarchical attention residual (HARM) module reduces false positives brought on by pseudo-changes and enhances feature refinement for better recognition of small objects. To further understand the correlation within the input images, the PGA-SiamNet [46] uses a co-attention module between the encoder and decoder. The PGA-SiamNet is capable of locating objects with displacement in other images, as well as identifying object changes of varying sizes with the aid of the pyramid change module. DASNet [104] leverages a fully convolutional architecture built upon two streams, often VGG16 or ResNet50, to extract image features. It includes a dual attention module that analyzes both channel-wise and spatial information within the features. In the same way, IFNet [65] utilizes a fully convolutional two-stream architecture based on VGG16. To address the disparity between change features and bi-temporal deep features. It incorporates dual attention (channel and spatial) and deep supervision to improve feature recognition and training of intermediate layers. The CANet [105] utilizes a Siamese architecture with ResNet18 as the backbone and incorporates a combined attention module that combines channel, spatial, and position attention mechanisms. It further enhances feature representation by incorporating an asymmetric convolution block (ACB), which replaces standard convolution with a combination of different kernel sizes, effectively enriching the feature space. To efficiently capture interchannel interactions in feature maps, the MBFNet [106] suggested a novel channel attention method. The network integrates second-order attention-based channel selection modules and a pseudo-Siamese CNN (AlexNet). In order to achieve more precise location and channel correlations, Ma et al. [107] developed a multi-attentive cued feature fusion network with a Feature Enhancement Module (FEM) that includes coordinate attention (CA). Chen et al. [108] successfully suppresses unnecessary features by fusing contextual data with an attention mechanism to provide extensive, global contextual knowledge about a building. Lately, to improve the detailed feature representation of buildings, [109] proposed an attention-guided high-frequency feature extraction module. More recently, this work [110] introduces the triplet UNet (T-UNet), which has a three-branch encoder that extracts object features and changes information simultaneously, ensuring that important details are retained during feature extraction. Furthermore, a Multi-Branch Spatial-Spectral Cross-Attention

(MBSSCA) module refines these features by leveraging details from pre- and post-images. The T-UNet outperforms other approaches like early fusion and Siamese networks.

5.2.2. Deep Belief Network-Based

Deep Belief Networks (DBNs) are a type of artificial neural network that has been explored for use in the change detection of remote sensing images. However, they are not as widely used as some other deep learning architectures.

Recent research explores using DBNs for various change detection tasks, including land cover clustering [111], change detection in SAR images using a Generalized Gamma Deep Belief Network [112], and building detection with high-resolution imagery [74]. Additionally, novel training approaches based on morphological processing of SAR images have been proposed to improve DBN performance [113]. While DBNs can be computationally expensive and require substantial data, they are a promising approach for understanding how our planet's land cover is changing.

5.2.3. RNN-Based

Change detection tasks are a good fit for RNNs, especially LSTMs (Long Short-Term Memories) since they can examine data sequences from various periods. Each RNN cell considers both the current data and information about the past stored in its hidden state, allowing the network to learn how data evolves over time. This makes them effective in identifying changes between multiple data periods. Various change detection methods [72,114–119] employed LSTM as a temporal module. In [73], the authors combined UNet and RNN architecture (BiLSTM), which is an LSTM development. UNet extracts the spatial features from input images with varying capture times, and then BiLSTM will analyze them to examine the temporal change pattern. Similarly, [120,121] also integrated LSTM networks with a fully convolutional neural network (FCN).

5.2.4. Transforms

Building on the success of attention mechanisms in understanding relationships between images, researchers are now exploring transformers for even more powerful results. Unlike attention mechanisms, which focus on specific image regions, transformers can analyze the entire image. This capability allows them to capture complex relationships between pixels across different time points. When using ViTs for CD of VHR RSIs, there are two strategies. Initially, temporal features are extracted by substituting ViTs for CNN backbones, such as ChangeFormer [122], Pyramid-SCDFormer [123], FTN [124], SwinSUNet [125], M-Swin [126], MGCDT [77,127], TCIANet [128], and EATDer [129]. Meanwhile, ViTs excel not only at feature extraction but also at modeling temporal dependencies. BiT [130] leverages a transformer encoder to pinpoint changes and employs two siamese decoders to create the change maps. [131] incorporated the token sampling strategy into the BIT framework to concentrate the model on the most beneficial areas. CTD-Former [132] proposes a novel cross-temporal transformer to analyze interactions between images from different times. Additionally, SCanFormer [133] offers a joint approach, modeling both the semantic information and change information in a single model. Zhou et al. [134] introduced the Dual Cross-Attention transformer (DCAT) method. This innovation lies in a novel dual cross-attention block that leverages a dual branch that combines convolution and transformer. Noman et al. [135] replaced conventional self-attention with a shuffled sparse-attention mechanism, focusing on selective, informative regions to capture CD data characteristics better. Additionally, they introduce a change-enhanced feature fusion (CEFF) module, which fuses features from input image pairs through per-channel re-weighting, enhancing relevant semantic changes and reducing noise.

5.2.5. Multi-Model Combinations

Recently, combining deep learning architectures has started gaining popularity in detecting changes, especially combining CNNs and transformers. These networks have a

strong ability to learn both local and global features within data. It makes them suitable for those tasks. CNNs are experts at identifying specific details within images, while transformers excel at determining how these details interconnect across the whole scene. By merging these capabilities, we can more accurately detect and analyze changes in remote sensing images over time. A lot of research uses this hybrid approach [75,136,137]. Wang et al. [138] introduce UVACD, which combines CNNs and transformers for change detection. A CNN backbone is used to extract high-level semantic features, while transformers are employed to capture the temporal information interaction for generating better change features. The work of [139] employs a hybrid architecture (TransUNetCD). The encoder in this architecture utilizes features extracted from CNNs and augments them with global contextual information. These enhanced features are then upsampled and merged with multi-scale features to generate global-local features for precise localization. Similarly, to collect and aggregate multiscale context information from features of various sizes, the CNN-transformer network MSCANet [140] presents a Multiscale Context Aggregator with token encoders and decoders. However, several methods have begun to include attention mechanisms in hybrid CNN-transformer networks. Authors in [141–143] integrate CBAM to bridge the gap between different types of features extracted from the data. In [144], a gated attention module (GAM) is employed in a layer-by-layer fashion. The work in [145] incorporates multiple attention mechanisms at different levels. On the other hand, some research employs transformer and CNN structures in parallel [146,147]. Tang et al. [148] proposed WNet, which combines features from a Siamese CNN and a Siamese transformer in the decoder. Furthermore, ACAHNet [149] combines CNN and transformer models in a series-parallel manner to create an asymmetric cross-attention hierarchical network. This reduces computational complexity and enhances interaction between the two models' features. To try to capture multiscale local and global features, Feng et al. [150] use a dual-branch CNN and transformer structure. They then employ cross-attention to fuse the features. To dynamically integrate the CNN and transformer branches' interaction. Fu et al. [151] built a semantic information aggregation module. One alternative approach involves combining CNNs with Graph Neural Networks [152].

5.3. Heterogeneous-RSCD

Heterogeneous RSCD (Het-RSCD) breaks free from the limitations of a single sensor. It can combine optical data from different resolutions or leverage the strengths of both optical and SAR data. By combining diverse sources, Het-RSCD creates a more complete view of Earth's surface changes, resulting in better accuracy and robustness in change detection tasks.

5.3.1. Multi-Scale Change Detection (Optical–Optical)

Multi-Scale Change Detection addresses the challenges of varying spatial resolutions in optical images. This process involves comparing images of the same type of data (optical) at different scales. The differences in scale can complicate the detection of changes, necessitating specialized approaches to ensure accurate results.

CNN-Based Methods

Remote sensing data are mostly image-based, and CNNs have shown impressive success. In addition to their application to individual data sources, CNNs find application in multi-scale optical change detection in several recent publications. As an early attempt, Lv et al. [51] introduces a multi-scale convolutional module within the UNet model to enhance change detection in heterogeneous images. Shao et al. [47] introduced a novel approach called SUNet, which employs two distinct feature extractors to generate feature maps from the two heterogeneous images. These extracted feature maps are then combined and fed into the decoder. Additionally, SUNet [47] utilizes a Canny edge detector and Hough transforms to extract edge auxiliary information from the heterogeneous two-phase images. The study conducted by Wang et al. [43] proposes a novel Siamese network archi-

tecturenamed OB-DSCNH, which includes a hybrid feature extraction module to extract more robust hierarchical features from input image pairs. Using a group convolution, the SepDGConv [81] allows embedding multi-stream structure into a single-stream CNN. Upgrade that to a dynamic. Zhu et al. [153] proposed a multiscale network with a chosen kernel-attention module and a non-parameter sample-enhanced method utilizing the Pearson correlation coefficient. Despite requiring few training samples, this approach excels at finding changes.

GAN-Based Methods

GANs have emerged as a powerful tool in deep learning. These fascinating architectures consist of two separate neural networks: a generator and a discriminator. The generator always aims to create realistic data samples, while the discriminator attempts to differentiate real data from the generator's creations. This ongoing action leads to the generator learning to produce increasingly high-quality outputs that closely resemble real data.

The ability of GANs to generate high-resolution (HR) images from lower-resolution (LR) inputs holds immense potential for Het-RSCD. As Het-RSCD depends on data from multiple sensors, these sensors may have varying resolutions. LR data can lack important details for accurate change detection. GANs can help by employing super-resolution strategies, as shown in Figure 6.

Super-resolution (SR) plays a crucial role in multi-modal change detection (CD) by enhancing the resolution of low-resolution (LR) images. This enhancement allows for more accurate and detailed analysis of changes. SR techniques lend themselves to both individual data modalities and fused images that combine information from multiple modalities.

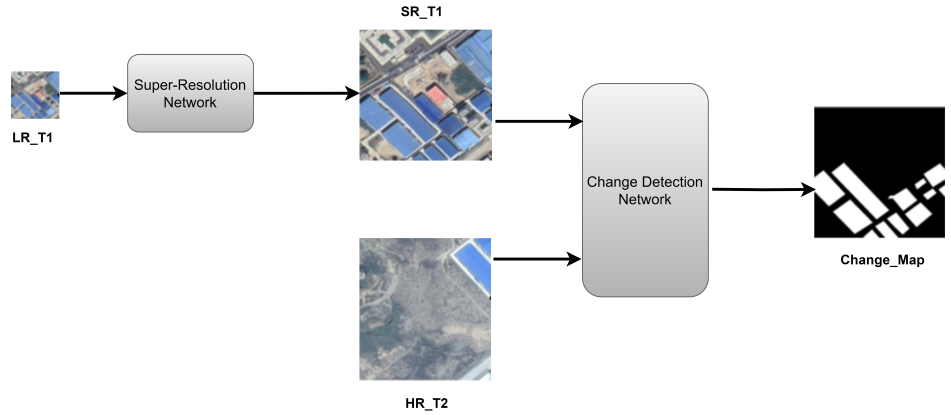


Figure 6. Structures of super resolution change detection methods.

Building upon the demonstrated effectiveness of SR in multi-modal CD tasks, SRCD-Net [154] tackles the challenge of change detection. It does this by utilizing a GAN-based SR module that generates HR images from LR ones, making it possible to compare images with similar resolutions. Simultaneously, both images are processed by parallel ResNet-based feature extractors. It applies a stacked attention module to augment the extraction of pertinent information from multiple layers. Similary, the RACDNet [155] proposed network comprises a light-weighted SR network (GAN) based on WDSR that recovers high-frequency detailed information by assigning gradient weights to different regions. The network also uses a novel Siamese-UNet architecture for effective change detection, which includes a deformable convolution unit (DCU) for aligning bi-temporal deep features and an atrous convolution unit (ACU) to increase the receptive field. An attention unit (AU) is embedded to fill the gaps between the encoder and decoder. The SiamGAN [156] is an end-to-end generative adversarial network that combines a SR network and a siamese structure to detect changes at various resolutions. A channel-wise operation was added, which allows different information scales to be combined and provides a richer repre-

sentation of the input data. Prexl et al. [157] proposed an unsupervised CD approach that extends the DCVA framework to handle pre- and post-change imagery with different spatial resolutions and spectral bands. The approach employs a self-supervised SR method to enhance lower-resolution images and a set of trainable convolution layers to address spectral differences. The MF-SRCDNet [158] proposed SR comprises an image transformation network and a loss network module based on Res-UNet. This method leverages the strengths of residual networks and UNet. It uses Res-UNet for image transformation and VGG-16 for loss. Followed by a multi-feature fusion strategy that extracts Harris-LSD visual features, morphological building index (MBI) features, and non-maximum suppressed Sobel (NMS-Sobel) features. Finally, a change detection module uses a modified STANet-PAM model with a Siamese structure, enhancing the detection of building changes using spatial attention mechanisms.

Transformers

Transformers have become increasingly popular in computer vision [159], including change detection. This rise in popularity follows their success in natural language processing [160]. In 2022, many new models published based transformers, especially in handling heterogeneous data sources.

The MM-Trans [161] involves a multi-modal transformer framework. It initially extracts features from bi-temporal images of varying resolutions using a Siamese feature extractor (ResNet18) with unshared weights. Next, with the help of a token loss, a spatial-aligned transformer (sp-Trans or SPT) is utilized to learn and shrink these bi-temporal characteristics to a constant size. To enhance interaction and alignment, a semantic-aligned transformer is then applied to the high-level bi-temporal characteristics. Ultimately, a prediction head is used to determine the altered result.

The STCD-Former [162] is a pure transformer model consisting of a spectral token transformer and a spectral token guidance spatial transformer. It encodes bi-temporal images, generates spectral tokens, and learns change rules. It includes a difference amplification module for discriminative features and an MLP for binary CD results.

Lastly, SILI [163] is an object-based method that utilizes a ResNet-18 Siamese CNN backbone to extract multilevel features from bi-temporal images. Local window self-attention establishes a feature interaction at different levels, capturing spatial-temporal correlations rather than encoding images independently. This process improves feature alignment by considering local texture variances. The refined features, obtained through a transformer encoder, contribute to enhanced feature extraction. The decoder utilizes implicit neural representation (INR) and coordinate information to generate a change map.

Multi-Model Combinations

The use of multi-model deep learning networks for multi-scale optical CD remains limited, potentially due to the challenges of data fusion and network architecture design. Moreover, convolutional multiple-layer recurrent neural networks are further proposed for CD with multi-sensor images. Chen et al. [164] proposed an innovative and universal deep siamese convolutional multiple-layer recurrent neural network (SiamCRNN), which combines the benefits of RNN and CNN. Its overall structure consists of three sub-networks that are highly connected, have a clear division of labor, and can be used to extract picture attributes, mine change information, and predict change likelihood. The M^3Fusion [165] uses a two-branch network. The CNN branch extracts patch-based features from a SPOT 6/7 image, and the RNN branch extracts temporal information from Sentinel-2 time-series images. The extracted features are the input for three classifiers, with two independent classifiers and a third applied to the fused features.

5.3.2. Multi-Modal Change Detection (Optical–SAR)

Multi-modal change detection integrates data from various sensor types, especially optical imagery and synthetic aperture radar (SAR). This approach aims to leverage each sensor type's unique strengths to enhance change detection capabilities.

CNN-Based Methods

Encoder–decoder architectures, leveraging the power of CNNs, extract features from multi-source data at various resolutions. These features are then compressed into a latent representation, effectively capturing the core of the changes. The decoder utilizes this latent representation to reconstruct an image, highlighting the areas where changes have occurred.

Early fusion methods, like M-UNet [51], employ multiscale convolutional modules within the UNet architecture to enhance change detection in heterogeneous images containing data from multiple sensors. More recent advancements include multi-modal Siamese architectures, such as the one proposed by Ebel et al. [166]. In this approach, two separate encoder branches process SAR and optical data individually. A multi-scale decoder then combines the extracted features from these branches to create a more comprehensive understanding of the changes. Similar to this, research by Hafner et al. [167] utilizes separate UNet models for SAR and optical data before fusing the extracted features at the final stage. In contrast to other research, which primarily employed pseudo-Siamese networks to extract features, [168] utilized two distinct encoder networks. Specifically, ResNet50 was used for optical data, while EfficientNet-B2 was used for SAR data. Finally, the MSCDUNet [169] architecture utilizes a pseudo-Siamese UNet++ structure. Each branch independently processes SAR and multispectral optical data using a UNet++ network to extract features. These features are then fused, and a deep supervision module leverages information from both branches to generate accurate change maps.

Alternatively, autoencoders significantly improve change detection (CD) with multi-source data by learning a unified latent space representation for data from different sources. Autoencoders handle differences between data sources (like sensor types) by finding common patterns. This lets the model identify changes regardless of the source and works well even with entirely new data sources. This makes them ideal for unsupervised change detection tasks, which are a perfect fit for domain adaptation methods that improve performance across different data distributions. DSDANet [170] stands as the first method to introduce unsupervised domain adaptation into change detection. The DAMSCDNet [171] suggests a domain adaptation-based network to treat optical and SAR images, which employs feature-level transformation to align unstable deep feature spaces. To align similar pixels from input images and minimize the impact of changed pixels, authors in [172] combined autoencoders and domain-specific affinity matrices. CAE [173] proposes an unsupervised change detection method. It contains only a convolutional autoencoder for feature extraction and the commonality autoencoder for commonalities exploration. Farahani et al. [174] propose an autoencoder-based technique to achieve fusion of features from SAR and optical data. This method aligns multi-temporal images by reducing spectral and radiometric differences, making features more similar, and improving accuracy in CD. Additionally, domain adaptation with an unsupervised autoencoder (LEAE) helps discover a shared feature space between heterogeneous images, further enhancing the fusion process. The DHFF [175] used an unsupervised CD approach, which utilizes image style transfer (IST) to achieve homogeneous transformation. The model separates semantic content and style features extracted from the images using the VGG network. The IIST strategy is employed, which iteratively minimizes a cost function to achieve feature homogeneity. A novel topology-coupling-based heterogeneous network called TSCNet [36] introduces wavelet transform, channel, and spatial attention methods in addition to transforming the feature space of heterogeneous images utilizing an encoder–decoder structure. Touati et al. [176] introduced a novel approach for detecting anomalies in image pairs using a stacked sparse autoencoder. The method works by encoding the input image into a latent space, computing reconstruction errors based on the L2 norm. It then generates a classi-

fication map indicating changes and unchanged regions by grouping the reconstructed errors using a Gaussian mixture. Zheng et al. [177] introduced a cross-resolution difference to detect changes in images with distinct resolutions. They segmented images into homogeneous regions and used a CDNN with two autoencoders to extract deep features. They defined a distance to assess semantic links, computed pixel-wise difference maps, and merged them to generate a final change map.

Transformers

CNNs have historically been used for CD across optical and SAR pictures by mapping both images into a common domain for comparison. CNNs, however, have difficulty identifying long-range dependencies in the data. A recent study by Wei et al. [178] suggests a solution to this issue by utilizing transformers. Even though the features acquired from each type of image are derived from distinct sensors, their Cross-Mapping Network (CM-Net) uses transformers to discover correlations between them. As a result, CM-Net can build a common representation space that is stronger and more reliable, enabling more precise change detection. Another approach is mSwinUNet [179], which utilizes a Swin transformer-based architecture to directly capture global semantic information from SAR and optical images. This method splits images into patches, encodes them with positional information, and employs a self-attention mechanism to learn global dependencies.

GAN-Based Methods

In remote sensing applications, GANs have become an effective tool for utilizing the complementary information of optical and synthetic aperture radar (SAR) data. Studies like [42,45,180–182] have successfully employed GAN-based image translation to enable the use of established optical CD methods on SAR data. For instance, Saha et al. [45] utilize a CycleGAN model for transcoding between different data domains. Deep features are extracted using an encoder–transformer–decoder architecture. In the same way, DTCDN [55] employs a cyclic structure to map images from one domain to another, effectively translating them into a shared feature space. The translated images are then fed into a supervised CD network. It leverages deep context features to identify and classify changes across different sensor modalities. Research by [180] translated SAR images into "optical-like" representations, enabling the use of established burn detection methods on post-fire SAR imagery. Similarly, [182] proposed a Deep Adaptation-based Change Detection Technique (DACDT) that utilizes image translation via an optimized UNet++ model to improve CD in challenging weather conditions. However, limitations exist with separate image translation and CD steps. Works like [183,184] address this by proposing frameworks that integrate both tasks within a single deep-learning architecture. Du et al. [183] introduced a Multitask Change Detection Network (MTCDN) that utilizes a concatenated GAN structure with separate generators and discriminators for optical and SAR domains. In contrast, [184] presented a Twin-Depthwise Separable Convolution Connect Network (TDSCCNet) that employs CycleGAN for front-end image domain transformation. Additionally, it uses a single-branch encoder–decoder for change feature extraction in the back-end. Recently, EO-GAN [185] employed edge information for indirect image translation via a cGAN. It extracts edges and reconstructs the corresponding optical image from a SAR image based on those edges. To further improve the learning process, a super-pixel method helps the network build a link between edge changes and actual content changes.

6. Discussion

The growing variety of remote sensing images has brought new challenges to RSCD, including analyzing changes between images of different resolutions and sources. Due to the limited availability of data in many CD scenarios, the occurrence of DRCD tasks is becoming increasingly unavoidable. For example, in regions that experience regular rainfall, floods, or storms, generating images with the same spatial resolutions over a long period poses considerable difficulties for annual land cover change monitoring. These

scenarios show the inefficiency of the typical CD method built for bi-temporal images with similar spatial resolution.

Deep learning's ability to learn autonomously from complex data has made it a popular choice for CD. However, the type of imagery used represents a major challenge. In its early stages, the field has focused on scenarios with homogeneous images. This simplifies CD, as the focus is only on identifying changes within the same data type. However, this approach has its limitations, as real-world scenarios often involve heterogeneous images. These images come from a variety of sources, such as optical and radar sensors, and have distinct characteristics.

6.1. Quantitative Evaluation of Hom-RSCD Models

The reviewed models showcase the dominance of CNNs with Siamese architecture for Hom-RSCD. These techniques have produced impressive results, frequently achieving accuracy over 90% and barely hitting 95%. Nevertheless, UNet's performance declined on challenging datasets, with a level of precision of less than 50%. Due to its incapacity to capture long-range dependencies. Researchers looked through several attention mechanisms to address this problem. These include hierarchical attention [103] to detect tiny target and pseudo-change, co-attention [46], channel and spatial attention [86,101,104], and combining multiple attentions [107] to enhance focus on changes. Moreover, SMD-Net [87], CANet [169], MSAK-Net [40], and MFPNet [186] employ diverse techniques to capture multiscale features in bi-temporal images, leading to noticeable performance improvements. As well, RFNet [86] intends to decrease the effects of spatially offset bi-temporal, which reached a precision of 74%. Although Siamese networks are good at preserving object features, they struggle to utilize change information, leading to inaccurate edge detection. To overcome this, authors in [110] were the first to propose a triple encoder capable of simultaneously extracting and synthesizing object features and changing features. This approach aims to improve change region detection accuracy, reaching an OA of 99%.

Despite their outstanding results, the CNN-based methods remain restricted to not being able to obtain the distant context information hidden in RS images. Thus, researchers have turned to transformer-based models, which excel at modeling these long-range dependencies. The DSIFN dataset yielded poor precision (68%) for BIT [130] due to the limitations of using ResNet18 for feature extraction at different scales. It lacks finesse during image restoration and different labeling in the decoding stage. For feature information extraction, Change Former [122] uses multi-head self-attention modules as the backbone network and achieves a high precision (88%) for the same dataset. It also significantly enhances the utilization of resources. Furthermore, CTD-Former [132] incorporates consistency perception blocks to preserve the shape information of changed areas. It's enhanced by deformable convolution and extracting information at bigger scales. Despite the success of the aforementioned techniques, the self-attention mechanism causes their computational costs to always be high. Using SwinT blocks in place of the traditional transformer encoder/decoder blocks and Self-Adapting Vision transformer (SAVT) blocks in the encoder, the authors in [125,129] reduce computational costs and reach a precision above 95%.

However, the transformer's global focus ignores details in low-resolution images, leading to poor segmentation and problems with decoder recovery. Combining the transformer with CNN can handle this issue. This combination is embedded into a UNet by TransunetCD [139] to enhance performance. Conversely, ICIF-Net [150] implements a simultaneous technique, extracting features from both CNN and transformer backbone networks, which yielded remarkable results (precision above 80%). However, the two feature extraction processes operate independently. Thus, WNet [148] joins to bring a deformable idea into the dual-Siamese branch encode to overcome the effects of the fixed convolutional kernel in CNNs and the regular patch generation in transformers. It raises the precision to around 90%. Table 3 shows the performance of homogeneous RSCD methods on different datasets.

Table 3. Summary table of performance homogeneous CD methods.

Method Name/Ref	Network Structure	DataSet	Precision (%)	F1 (%)	OA (%)
DSMS-FCN [82]	Siamese UNet	SZTAKI-Szada SZTAKI-Tiszadob	52.78 89.18	57.72 88.86	94.57 96.20
ESCNet [85]	Siamese UNet	SZTAKI-Tiszadob SZTAKI-Szada	76.33 48.89	74.56 53.73	93.95 94.07
RFNet [86]	Siamese CNN	WHU-CD	95.72	92.49	-
SMD-Net [87]	Siamese UNet	CDD BCDD OSCD	96.6 94.80 96.6	97 94.33 97.0	99.3 99.48 99.3
SSCFNet [89]	Siamese UNet	LEVIR-CD SZTAKI	93.71 96.54	95.31 96.58	- -
Siam-FAUNet [88]	Siamese UNet	CDD WHU-CD	95.62 44.47	94.58 55.50	98.14 94.95
DASNet [104]	Siamese UNet + Attention	CDD	92.2	92.7	98.2
DifUNet++ [92]	Siamese UNet++	SVCD LEVIR-CD	92.15 92.15	92.37 89.6	- -
SNUNet-CD [93]	Siamese UNet++	CDD	96.3	96.2	-
TCDNet [94]	Siamese CNN	Google Earth	71.18	-	-
SSJLN [95]	Siamese CNN	GF-1 Data EMT+ Data	- -	94.94 98.75	- -
SAM-CD [98]	Siamese CNN	LEVIR-CD CLCD WHU-CD S2Looking	95.87 88.25 97.97 72.80	95.50 86.89 97.58 65.13	99.14 96.26 99.60 -
NestNet [91]	Siamese UNet++ ,Attention	CDD OSCD	88.26 49.01	88.62 49.32	- -
HARNU-Net [103]	Siamese UNet, Attention	CDD	97.10	97.20	99.34
AFSNet [101]	Siamese UNet, Attention	CDD	98.44	95.56	98.94
CANet [105]	Siamese UNet, Attention	CDD	93.2	93.2	98.4
PGA-SiamNet [46]	Siamese UNet, Attention	EV-CD building	94.01	91.74	99.68
MFPNet [186]	Siamese UNet, Attention	SVCD Zhang dataset		97.54 68.45	- -
MAFF-Net [107]	Siamese UNet, Attention	CDD LEVIR-CD WHU-CD	96.5 89.7 92.4	99.2 98.9 99.4	- - -
MSF-Net [108]	Siamese UNet, Attention	LEVIR-CD	90	88.66	-
FERA-Net [109]	Siamese UNet, Attention	LEVIR-CD WHU-CD	91.57 93.51	89.58 92.48	- -
T-UNet [110]	Triple UNet, Attention	LEVIR-CD WHU-CD DSIFN	92.60 95.44 70.86	91.63 91.77 69.52	99.16 99.42 89.83
ChangeFormer [122]	Siamese Transformer	LEVIR-CD DSIFN	92.05 88.48	90.40 86.67	99.04 95.56
SwinSUNet [125]	Siamese Transformer	CDD OSCD WHU	95.7 55.0 95.0	94.0 54.5 93.8	98.5 95.3 99.4

Table 3. *Cont.*

Method Name/Ref	Network Structure	DataSet	Precision (%)	F1 (%)	OA (%)
BiT [130]	Siamese Transformer	LEVIR-CD	89.24	89.31	98.92
		DSIFN	68.36	69.26	89.41
EATDer [129]	Siamese Transformer	LEVIR-CD	91.74	91.20	98.75
		CDD	96.83	95.97	98.97
		WHU-CD	91.32	90	98.58
CTD-Former [132]	Siamese Transformer	LEVIR-CD	91.85	92.71	98.62
		WHU-CD	96.74	96.86	99.5
		CLCD	87.29	85.08	96.11
SCanFormer [133]	Siamese Transformer	SECOND	-	63.66	87.86
		Landsat-SCD	89.27	96.26	
TransUNetCD [139]	Siamese UNet + Transformer	CDD	93.2	93.2	98.4
		S2Looking	93.2	93.2	98.4
CTCANet [141]	Siamese CNN + Transformer	LEVIR-CD	92.19	91.21	99.11
		SYSU-CD	80.50	81.23	91.40
DCAT [134]	Siamese (CNN + Transformer)	LEVIR-CD+	84.72	84.02	-
		SYSU-CD	87.00	79.63	-
		WHU-CD	91.53	88.19	-
SMART [145]	Siamese (CNN + Transformer)	LEVIR-CD	94.29	93.04	98.69
		SYSU-CD	86.17	84.80	89.42
		WHU-CD	89.9	91.57	98.70
		DSIFN	76.89	78.7	87
WNet [148]	Siamese CNN+Siamese Transformer	LEVIR-CD	91.16	90.67	99.06
		WHU-CD	92.37	91.25	99.31
		SYSU-CD	81.71	80.64	90.98
		SVCD	97.71	97.56	99.42
ACAHNet [149]	Siamese (CNN + Transformer)	CDD	97.5	97.72	99.48
		LEVIR-CD	92.36	91.51	99.14
		SYSU-CD	83.96	82.73	91.97
ICIF-Net [150]	Siamese (CNN + Transformer)	LEVIR-CD+	87.79	83.65	98.73
		WHU-CD	92.98	88.32	98.96
		SYSU-CD	83.37	80.74	91.24
Slddnet [151]	Siamese (CNN + Transformer)	LEVIR-CD	-	91.75	-
		WHU-CD	-	92.76	-
		GZ-CD	-	86.61	-

6.2. Quantitative Evaluation of Het-RSCD Models

In the context of Het-RSCD, some methods often resolve the problem at the image level, especially when working with images of varying resolution. The easiest way is to use interpolation to upsample LR photos to HR resolution [102,187]. Upsampling VHR optical images doesn't significantly affect accuracy much, as seen in OB-DSCNH [43], which achieves high overall accuracies of 97% due to the lesser impact of the lack of spatial details. In [188], bands at 20 m resolution were resampled at 10 m using bicubic interpolation. Despite attaining an overall accuracy of 89%, they miss a lot of information details because high-resolution differences cause bicubic interpolation to perform poorly. It may result in mismatched or unclearly aligned retrieved features. By employing subpixel information through the subpixel convolution technique, SPCNet [189] aims to resolve this problem. However, the model was only tested on synthetic LR images, which raises the question about its overall generalizability. Despite their capabilities, both resampling and interpolation face inherent obstacles in preserving accuracy for change detection. Resam-

pling sacrifices precise spatial information in high-resolution images, while interpolation struggles to fully retrieve the rich semantic detail missing from low-resolution images.

Recently, DL-SR techniques have been applied to transform LR into HR, which overcomes the resolution limitations intrinsic to different sensors thanks to its powerful ability to recover semantic information from images. Most SR methods used GANs; for example, SiamGAN [156] combines a SRGAN and a siamese structure trained with a 4 and 1 m resolution image, achieving an accuracy of 69.5% and a F1-score of 76.06%. However, limitations were observed in handling complex scenes. This is due to its reliance on patch-based processing. In SRCDNet [154] and RACDNet [155], the SR model (generator) employs only residual networks, which can greatly increase the training period and make it difficult to fully maintain the spatial and contour detail information required for reconstruction. The SR module in MF-SRCDNet [158] introduces a Res-UNet to generate unified SR images and VGG-16 as a loss network. This model matches the resolution and learns similar sensor properties, such as lighting and viewing angle. It achieves an impressive result in detecting changes in images with a 4× and 8× resolution difference. However, the model's performance faced challenges in reconstructing the spatial structure of highly disparate scenes. Even though super-resolution technology is achieving good results. It remains limited by its fixed-scale upsampling ability and the high cost of obtaining paired LR–HR images for real-world SR training.

To overcome limitations in handling complex data and extracting comprehensive information, research is shifting towards fusing features from multi-modal images. For instance, most methods fuse SAR and optical images. M-UNet [51] is an early fusion method that employs a multiscale UNet, achieving an accuracy between 79% and 90% across three datasets. While single streams are initially appealing for their simplicity and efficiency, they suffer problems in capturing complicated relationships across multiple modalities, especially in dynamic and complicated environments. As a result, the focus of research has moved towards the employing of Siamese networks. Ebel et al. [166] introduced an UNet Siamese architecture that fuses SAR and optical data at several decoder depths. Following a similar concept, research in [167] handles each data modality separately using UNet before fusing the obtained features at the final decision stage. Both approaches achieved a F1 score of around 60%. However, the accuracy is tiny compared with the optical baseline. Therefore, using pure UNet may not be the most effective approach for handling multiple source images. The authors of [168] applied two different encoders to the optical and SAR, ResNet50 and EfficientNet-B2, respectively, for Flood CD, achieving an accuracy of 97%. MSCDUNet [169] fused multispectral, SAR, and VHR images by combining the strengths of dense connections and depth supervision in the pseudo-Siamese UNet++, achieving F1 scores of 92.81% on the MSOSCD and 64.21% on the MSBC. The lower score on MSBC highlights the challenge of small data.

Moreover, high-dimensional features in heterogeneous images are present in different feature spaces, making it challenging to accurately highlight changes in information between them. For this reason, SiamRNN [164] used LSTM units to process the spatial-spectral features and extract change information, achieving a precision rate of 0.8738 and an F1 of 0.8215. Fortunately, SiamRNN is suitable for multi-source VHR images with a smaller domain difference. To fuse optical and UAV images with different resolutions, SUNet [47] succeeded by adding two distinct extraction channels. The extract features were concatenated with edge information before the UNet encoder to adjust images of different sizes and to push the model to focus more on contours and shapes than colors. Achieving an impeccably high result (precision: 97%, F1: 91%).

In recent years, researchers have integrated transformers into DRCD tasks, employing them individually or concatenated with CNNs. This development shows a growing recognition of transformers' ability to capture global context and semantic relationships. While existing CNN methods often neglect physical mechanisms, STCD-Former [162] stands out by employing spectral tokens to guide patch token interaction. However, its training on images of the same resolution but different sensors (achieving 99% in OA)

limits its ability to generalize to more diverse scenarios with varying sensors or image properties. To achieve semantic alignment across resolutions (i.e., difference ratio, e.g., 4, 8), a recent study [161] used CNN-based siamese feature extraction and transformers to learn correlations between the upsampled LR features and the original HR ones, which verifies the effectiveness of the feature-wise alignment strategy. The methods mentioned are effective for fixed resolution differences but may not be suitable for situations with other resolution differences, limiting their practical applications. To fill this gap, SILI [163] offers a single model adjusted to different ratios between bi-temporal images by using local window self-attention to establish a feature interaction at different levels and capturing spatial-temporal correlations rather than encoding images independently. The decoder utilizes implicit neural representation (INR) to generate a change map.

Data fusion is also used for classification tasks, with many methods integrating LiDAR and hyperspectral images through various applications. For instance, Siamese networks are often employed, as seen in studies [190,191]. Techniques include the Squeeze-and-Excitation module for weighted feature fusion [192]. FusAtNet's cross-attention allows each modality's feature learning to benefit from the other [193]. Additionally, SepDGConv's single-stream network with Dynamic Group Convolution [81]. AMM-FuseNet [194] enhances performance using channel attention and densely connected atrous spatial pyramid pooling. Additionally, [165] fuses Sentinel-2 Time Series and Spot7 images using a GRU with Attention and a CNN branch, aided by auxiliary classifiers.

However, a notable limitation of supervised methods is that models necessitate large amounts of labeled data, which are costly and time-consuming to create, especially for change detection tasks. Interest in unsupervised networks is growing as they aim to reduce reliance on labeled datasets. Domain adaptation is a popular method that aims to project pre-change and post-change images into a shared feature space to allow for comparison. Image-to-image (I2I) translation via a conditional generative adversarial network (cGAN) [195] is a powerful technique for mapping data across domains. Particularly, the CycleGAN [42,45] approach utilizes cGANs and enforces cyclic consistency to accomplish even more powerful results. Therefore, censoring change pixels is important for applying this method in heterogeneous CD because their existence perturbs training and promotes irrelevant object transformations. Despite their capability, high training requirements, imbalanced training dynamics, and the possibility for mode failure or unstable loss functions can limit their real-world applicability. In addition, these methods [175,196] applied homogeneous transformation, which refers to transforming the heterogeneous images into a homogeneous domain based on image translation and immediately comparing them at the pixel level. Nevertheless, the homogeneous transformation characteristics rely on low-level information such as pixel values, which are likely to affect the altered products' semantic meaning, particularly in regions with many objects and intense environments. Nowadays, several research papers have begun to concentrate on self-supervised multi-modal learning. It motivates the network to acquire more meaningful and accessible feature representations. Wu et al. [173] effectively aligned related pixels from multi-modal images through domain-specific affinity matrices and autoencoders. Luppino et al. [172] suggested a commonality autoencoder capable of discovering common features within heterogeneous image representations. Nevertheless, its sensitivity to hyperparameters requires careful tuning for optimal performance. Jiang et al. [176] proposed a stacked sparse autoencoder unsupervised method for anomaly detection in image pairs. While most of the current methods focus on extracting deep features to get the full image transformation, neglecting the image's topological structure. It includes direction, edge, and texture information. Thus, TSCNet [36] proposes a new topology-coupling algorithm by introducing wavelet transform, channel, and spatial attention mechanisms. Table 4 shows the performance of heterogeneous RSCD methods on different datasets.

Table 4. Summary table of performance heterogeneous CD methods.

Method Name/Reference	Network Structure	DataSet	Precision (%)	F1 (%)	OA (%)
M-UNet [51]	Single UNet	Shuguang	-	84.73	98.69
		Sardinia	-	67	98.01
		California	-	61.33	96.66
OB-DSCNH [43]	Siamese CNN	Mengxi Liu [43]	-	-	97.92
SepDGConv [81]	Single CNN	Houston2018	56.55	-	63.74
		Berlin	54.23	-	68.21
		MUUFL	72.75	-	83.23
MM-Trans [161]	Siamese CNN + Transformer	8×/11× CCD	95.48/95.17	90.44/90.07	-
		5×/8× S2looking	65.37/64.57	58.62/56.99	-
		8× HTCD	82.13	74.99	-
MSCDUNet [169]	Siamse UNet++	MSBC Dataset	-	64.21	-
		MSOSCD Dataset	-	92.81	-
RACDNet [155]	GAN + Saimese UNet	MRCDD Dataset		91.18	96.79
SUNet [139]	Siamese UNet	HTCD dataset	97.3	91	99.6
Patrick et al. [166]	Siamese UNet	ONERA CD data	60.2	58.1	-
STCD-Former [162]	Siamese Transformer	Bastrop data	-	99.25	-
M^3Fusion [165]	Siamese CNN + RNN	Reunion Island	90.09	89.96	-
AMM-FuseNet [194]	Siamese UNet + Attention	Hunan	59.13	79.06	
		DFC2020	-	90.33	94.56
		Potsdam		79.31	85.28
MFT [197]	Siamese CNN + Transformer	Houston2013	90.56	-	89.15
		MUUFL	81	-	94.18
		Trento	95.91	-	97.76
Chen et al. [191]	Siamese CNN	Houston2013	98.57	-	98.61
		Bayview Park	99.75	-	99.41
		Recology	98.90	-	98.15
MBFNet [106]	Siamese CNN + Attention	PoDelta	-	-	82.61
		CHONGMING	-	-	93.61
TWINNS [188]	Siamse CNN, GRU	Reunion Island	89.87	89.88	-
SiamCRNN [164]	Siamese CNN + LSTM	LiDAR-Opt	87.38	82.15	82.15
MF-SRCDNet [158]	GAN + Siamese UNet	WXCD	84.5	88.1	95.3
		BCDD	96.4	96.4	98.5
SiamGAN [156]	Siamese GAN	Guangzhou	69.5	76.06	-
SRCDNet [154]	GAN + Siamese UNet, Attention	4×/8× BCDD	84.44/81.61	85.66/81.69	-
		4×/8× CDD	92.07/91.95	-	-
SILI [163]	Siamese CNN + Transformer	LEVIR-CD(4×)	90	88	98
		SV-CD(8×)	95	94	98
		DE-CD(3.3×)	61	50	-
DAMSCDNet [171]	Siamese CNN	Data1	78.89	82.17	-
		Data2	92.04	93.86	-
		Data3	71.51	-	71.71
CA_AE [172]	Autoencoders	Lake overflow	-	-	92.2
		Constructions	-	-	85.9
CAE [173]	Autoencoders	Yellow River	-	-	97.74
		Sardinia	-	-	97.47
		farmland	-	-	97.91

Table 4. *Cont.*

Method Name/Reference	Network Structure	DataSet	Precision (%)	F1 (%)	OA (%)
Farahani et al. [174]	Autoencoders	San Francisco	-	96.44	72/68
DHFF [175]	Siamese VGG (IST)	Tōhoku	84.66	-	98.63
		Haiti	58.19	-	98.23
TSCNet [36]	Autoencoders + Attention	Flood California [198]	49.4	5.74	93.9
Niu et al. [195]	Autoencoders	Yellow River	-	-	97.7
		farmland	-	-	98.26
CM-Net [178]	Autoencoder + Transformer	SARDINA	90.55		97.52
		Shuguang	95.00	-	98.57
		GLOUCESTERSHIRE	93.51	-	96.92
DTCDN [55]	CycleGAN	Gloucester I	89.96	89.95	97.98
		Gloucester II	90.78	88.67	96.33
		California	66.73	72.03	97.61
		Shuguang	92.92	91.56	99.75
DACDT [182]	CycleGAN	Gloucester I	-	-	98.67
		Gloucester II	-	-	97.68
		California	-	-	98.87
MTCDN [183]	CycleGAN	Gloucester I	88.86	88.22	97.65
		Gloucester II	89.49	88.87	96.34
		California	55.20	61.54	95.83
TDSCCNet [184]	CycleGAN	Italy	85.64	81.07	97.62
		WV-3	91.34	91.37	98.01
		Gloucester	93.29	93.75	97.36
		Shuguang	82.58	88.58	97.01
EO-GAN [185]	CGAN	Yellow River			98.01
		Shuguang	-	-	98.16

6.3. Challenges and Future Directions

Despite the advancements in homogeneous and heterogeneous change detection methods, several challenges remain. The major challenge in CD is the lack of open-source datasets, particularly for multi-source data. Despite the large quantity of RS images available, obtaining high-quality annotated (CD) datasets poses considerable difficulties because CD tasks require multiple images, making it even harder to acquire such datasets. Although homogeneous datasets are more accessible, the rarity of comprehensive multi-source datasets poses an obstacle to developing and testing robust change detection models. This limits the ability to compare approaches and slows down advancements in the field. A further challenge arises from the rarity of actual changes in RS images. This means that most pixels in a dataset remain constant. As a result, a unique strategy, such as a carefully considered loss function, is essential to address the performance issues caused by class imbalance.

Additionally, the majority of research focuses on detecting changes from two images, leaving us blind to subtle shifts and complex dynamics. This limited view can miss gradual changes, misinterpret noise, and limit our ability to model processes. Thus, by integrating multiple images, we widen our temporal window, exposing hidden trends, improving accuracy, and enabling new applications like studying slow-moving changes.

Moving forward, future studies could concentrate on semantic change detection using multi-sensor data. Models focusing on multi-sensor data, such as fusing Landsat and Sentinel-2 images, are still rare. This research will include the potential of employing multiple images as input to improve model performance and feature representation.

7. Conclusions

In many real-world remote sensing applications, change detection is an essential component. Deep learning has gained increasing traction for accomplishing this task.

This study delves into the deployment of deep learning techniques for change detection in remote sensing, particularly utilizing multi-modal imagery. It provides a summary of available datasets suitable for change detection and analyzes the effectiveness of various deep-learning models. There are two categories of models: those tailored for homogeneous change detection and those suitable for diverse data types (heterogeneous). Additionally, the paper illustrates the strengths, challenges, and possible avenues for future research in this field.

A large amount of research in change detection has focused on homogenous scenarios. Moreover, heterogeneous change detection presents a more challenging issue. Managing discrepancies in data types, specifically when dealing with varying resolutions in multi-sensor data, significantly complicates the detection process. Consequently, many research efforts try to deal with change detection problems using multi-source data with similar or near-identical resolutions, such as combining SAR and optical data.

Author Contributions: All authors contributed in a substantial way to the manuscript. S.S. and S.I. conceived the review. S.S. and S.I. designed the overall structure of the review. S.S. wrote the manuscript. All authors discussed the basic structure of the manuscript. S.S., S.I. and A.M. contributed to the discussion of the review. S.S., S.I., A.M. and Y.K. made contribution to the review of related literature. M.A. reviewed the manuscript and supervised the study for all the stages. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA), and the CNRST of Morocco (ALKHAWARIZMI/2020/29).

Data Availability Statement: No new data were created in this manuscript.

Acknowledgments: The authors are grateful to the reviewers for their constructive comments and valuable assistance in improving the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aplin, P. Remote sensing: Land cover. *Prog. Phys. Geogr.* **2004**, *28*, 283–293. [CrossRef]
2. Rees, G. *Physical Principles of Remote Sensing*; Cambridge University Press: Cambridge, UK, 2013.
3. Pettorelli, N. *Satellite Remote Sensing and the Management of Natural Resources*; Oxford University Press: Oxford, UK, 2019.
4. Yin, J.; Dong, J.; Hamm, N.A.; Li, Z.; Wang, J.; Xing, H.; Fu, P. Integrating remote sensing and geospatial big data for urban land use mapping: A review. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102514. [CrossRef]
5. Dash, J.P.; Pearse, G.D.; Watt, M.S. UAV multispectral imagery can complement satellite data for monitoring forest health. *Remote Sens.* **2018**, *10*, 1216. [CrossRef]
6. Cillero Castro, C.; Domínguez Gómez, J.A.; Delgado Martín, J.; Hinojo Sánchez, B.A.; Cereijo Arango, J.L.; Cheda Tuya, F.A.; Díaz-Varela, R. An UAV and satellite multispectral data approach to monitor water quality in small reservoirs. *Remote Sens.* **2020**, *12*, 1514. [CrossRef]
7. Shirmard, H.; Farahbakhsh, E.; Müller, R.D.; Chandra, R. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sens. Environ.* **2022**, *268*, 112750. [CrossRef]
8. Demchev, D.; Eriksson, L.; Smolanitsky, V. SAR image texture entropy analysis for applicability assessment of area-based and feature-based aea ice tracking approaches. In Proceedings of the EUSAR 2021; 13th European Conference on Synthetic Aperture Radar, VDE, Online, 29–31 April 2021; pp. 1–3.
9. Wen, D.; Huang, X.; Bovolo, F.; Li, J.; Ke, X.; Zhang, A.; Benediktsson, J.A. Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 68–101. [CrossRef]
10. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]
11. Zhang, J. Multi-source remote sensing data fusion: Status and trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24. [CrossRef]
12. Liu, Y.; Pang, C.; Zhan, Z.; Zhang, X.; Yang, X. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 811–815. [CrossRef]

13. Shi, S.; Zhong, Y.; Zhao, J.; Lv, P.; Liu, Y.; Zhang, L. Land-use/land-cover change detection based on class-prior object-oriented conditional random field framework for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *60*, 1–16. [CrossRef]
14. Brunner, D.; Bruzzone, L.; Lemoine, G. Change detection for earthquake damage assessment in built-up areas using very high resolution optical and SAR imagery. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Honolulu, HI, USA, 25–30 July 2010; pp. 3210–3213.
15. You, Y.; Cao, J.; Zhou, W. A survey of change detection methods based on remote sensing images for multi-source and multi-objective scenarios. *Remote Sens.* **2020**, *12*, 2460. [CrossRef]
16. Deng, J.; Wang, K.; Deng, Y.; Qi, G. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* **2008**, *29*, 4823–4838. [CrossRef]
17. Bovolo, F.; Bruzzone, L.; Marconcini, M. A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2070–2082. [CrossRef]
18. Hao, M.; Zhou, M.; Jin, J.; Shi, W. An advanced superpixel-based Markov random field model for unsupervised change detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1401–1405. [CrossRef]
19. Zhou, L.; Cao, G.; Li, Y.; Shang, Y. Change detection based on conditional random field with region connection constraints in high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3478–3488. [CrossRef]
20. Tan, K.; Jin, X.; Plaza, A.; Wang, X.; Xiao, L.; Du, P. Automatic change detection in high-resolution remote sensing images by using a multiple classifier system and spectral–spatial features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3439–3451. [CrossRef]
21. Seo, D.K.; Kim, Y.H.; Eo, Y.D.; Lee, M.H.; Park, W.Y. Fusion of SAR and multispectral images using random forest regression for change detection. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 401. [CrossRef]
22. Wang, C.; Wang, X. Building change detection from multi-source remote sensing images based on multi-feature fusion and extreme learning machine. *Int. J. Remote Sens.* **2021**, *42*, 2246–2257. [CrossRef]
23. Touati, R.; Mignotte, M.; Dahmane, M. Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model. *IEEE Trans. Image Process.* **2019**, *29*, 757–767. [CrossRef]
24. Cheng, G.; Huang, Y.; Li, X.; Lyu, S.; Xu, Z.; Zhao, H.; Zhao, Q.; Xiang, S. Change detection methods for remote sensing in the last decade: A comprehensive review. *Remote Sens.* **2024**, *16*, 2355. [CrossRef]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Schmidt, R.M. Recurrent neural networks (rnns): A gentle introduction and overview. *arXiv* **2019**, arXiv:1912.05911.
27. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
28. Li, J.; Hong, D.; Gao, L.; Yao, J.; Zheng, K.; Zhang, B.; Chanussot, J. Deep learning in multimodal remote sensing data fusion: A comprehensive review. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102926. [CrossRef]
29. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep learning-based change detection in remote sensing images: A review. *Remote Sens.* **2022**, *14*, 871. [CrossRef]
30. Jiang, H.; Peng, M.; Zhong, Y.; Xie, H.; Hao, Z.; Lin, J.; Ma, X.; Hu, X. A survey on deep learning-based change detection from high-resolution remote sensing images. *Remote Sens.* **2022**, *14*, 1552. [CrossRef]
31. Bai, T.; Wang, L.; Yin, D.; Sun, K.; Chen, Y.; Li, W.; Li, D. Deep learning for change detection in remote sensing: A review. *Geo-Spat. Inf. Sci.* **2023**, *26*, 262–288. [CrossRef]
32. Parelius, E.J. A review of deep-learning methods for change detection in multispectral remote sensing images. *Remote Sens.* **2023**, *15*, 2092. [CrossRef]
33. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef]
34. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
35. Daudt, R.C.; Le Saux, B.; Boulch, A.; Gousseau, Y. Urban change detection for multispectral earth observation using convolutional neural networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2115–2118.
36. Wang, X.; Cheng, W.; Feng, Y.; Song, R. TSCNet: Topological structure coupling network for change detection of heterogeneous remote sensing images. *Remote Sens.* **2023**, *15*, 621. [CrossRef]
37. Chen, H.; Yokoya, N.; Wu, C.; Du, B. Unsupervised multimodal change detection based on structural relationship graph representation learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
38. Chen, H.; Shi, Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* **2020**, *12*, 1662. [CrossRef]
39. Ji, S.; Wei, S.; Lu, M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 574–586. [CrossRef]
40. Feng, S.; Fan, Y.; Tang, Y.; Cheng, H.; Zhao, C.; Zhu, Y.; Cheng, C. A change detection method based on multi-scale adaptive convolution kernel network and multimodal conditional random field for multi-temporal multispectral images. *Remote Sens.* **2022**, *14*, 5368. [CrossRef]

41. Shen, L.; Lu, Y.; Chen, H.; Wei, H.; Xie, D.; Yue, J.; Chen, R.; Lv, S.; Jiang, B. S2Looking: A satellite side-looking dataset for building change detection. *Remote Sens.* **2021**, *13*, 5094. [CrossRef]
42. Lebedev, M.; Vizilter, Y.V.; Vygolov, O.; Knyaz, V.A.; Rubis, A.Y. Change detection in remote sensing images using conditional adversarial networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 565–571. [CrossRef]
43. Wang, M.; Tan, K.; Jia, X.; Wang, X.; Chen, Y. A deep siamese network with hybrid convolutional feature extraction module for change detection based on multi-sensor remote sensing images. *Remote Sens.* **2020**, *12*, 205. [CrossRef]
44. Volpi, M.; Camps-Valls, G.; Tuia, D. Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 50–63. [CrossRef]
45. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised multiple-change detection in VHR multisensor images via deep-learning based adaptation. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5033–5036.
46. Jiang, H.; Hu, X.; Li, K.; Zhang, J.; Gong, J.; Zhang, M. PGA-SiamNet: Pyramid feature-based attention-guided siamese network for remote sensing orthoimagery building change detection. *Remote Sens.* **2020**, *12*, 484. [CrossRef]
47. Shao, R.; Du, C.; Chen, H.; Li, J. SUNet: Change detection for heterogeneous remote sensing images from satellite and UAV using a dual-channel fully convolution network. *Remote Sens.* **2021**, *13*, 3750. [CrossRef]
48. Li, Y.; Zhou, Y.; Zhang, Y.; Zhong, L.; Wang, J.; Chen, J. DKDFN: Domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. *ISPRS J. Photogramm. Remote Sens.* **2022**, *186*, 170–189. [CrossRef]
49. Robinson, C.; Malkin, K.; Jojic, N.; Chen, H.; Qin, R.; Xiao, C.; Schmitt, M.; Ghamisi, P.; Hänsch, R.; Yokoya, N. Global land-cover mapping with weak supervision: Outcome of the 2020 IEEE GRSS data fusion contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3185–3199. [CrossRef]
50. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Benitez, S.; Breitkopf, U. The ISPRS benchmark on urban object classification and 3D building reconstruction. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. I-3* **2012**, *1*, 293–298. [CrossRef]
51. Lv, Z.; Huang, H.; Gao, L.; Benediktsson, J.A.; Zhao, M.; Shi, C. Simple multiscale UNet for change detection with heterogeneous remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
52. Xu, Y.; Du, B.; Zhang, L.; Cerra, D.; Pato, M.; Carmona, E.; Prasad, S.; Yokoya, N.; Hänsch, R.; Le Saux, B. Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1709–1724. [CrossRef]
53. Hong, D.; Hu, J.; Yao, J.; Chanussot, J.; Zhu, X.X. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 68–80. [CrossRef]
54. Gader, P.; Zare, A.; Close, R.; Aitken, J.; Tuell, G. *Muufl Gulfport Hyperspectral and Lidar Airborne Data Set*; University of Florida: Gainesville, FL, USA, 2013.
55. Li, X.; Du, Z.; Huang, Y.; Tan, Z. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 14–34. [CrossRef]
56. Huang, C.; Chen, Y.; Zhang, S.; Wu, J. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. *Rev. Geophys.* **2018**, *56*, 333–360. [CrossRef]
57. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
58. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; et al. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–39. [CrossRef]
59. Gómez-Chova, L.; Tuia, D.; Moser, G.; Camps-Valls, G. Multimodal classification of remote sensing images: A review and future directions. *Proc. IEEE* **2015**, *103*, 1560–1584. [CrossRef]
60. Daudt, R.C.; Le Saux, B.; Boulch, A.; Gousseau, Y. Multitask learning for large-scale semantic change detection. *Comput. Vis. Image Underst.* **2019**, *187*, 102783. [CrossRef]
61. Peng, D.; Zhang, Y.; Guan, H. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [CrossRef]
62. Zheng, Z.; Wan, Y.; Zhang, Y.; Xiang, S.; Peng, D.; Zhang, B. CLNet: Cross-layer convolutional neural network for change detection in optical remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 247–267. [CrossRef]
63. Lei, Y.; Peng, D.; Zhang, P.; Ke, Q.; Li, H. Hierarchical paired channel fusion network for street scene change detection. *IEEE Trans. Image Process.* **2020**, *30*, 55–67. [CrossRef] [PubMed]
64. Zhang, M.; Xu, G.; Chen, K.; Yan, M.; Sun, X. Triplet-based semantic relation learning for aerial remote sensing image change detection. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 266–270. [CrossRef]
65. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [CrossRef]
66. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the Computer Vision–ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; Proceedings, Part II 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 850–865.

67. Adarme, M.O.; Feitosa, R.Q.; Happ, P.N.; De Almeida, C.A.; Gomes, A.R. Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 910. [CrossRef]
68. Zhang, J.; Wang, Z.; Bai, L.; Song, G.; Tao, J.; Chen, L. Deforestation Detection Based on U-Net and LSTM in Optical Satellite Remote Sensing Images. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, Brussels, Belgium, 11–16 July 2021; pp. 3753–3756.
69. John, D.; Zhang, C. An attention-based U-Net for detecting deforestation within satellite sensor imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102685. [CrossRef]
70. Alshehri, M.; Ouadou, A.; Scott, G.J. Deep Transformer-based Network Deforestation Detection in the Brazilian Amazon Using Sentinel-2 Imagery. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
71. Bidari, I.; Chickerur, S. Deep Recurrent Residual U-Net with Semi-Supervised Learning for Deforestation Change Detection. *SN Comput. Sci.* **2024**, *5*, 893. [CrossRef]
72. Papadomanolaki, M.; Verma, S.; Vakalopoulou, M.; Gupta, S.; Karantzalos, K. Detecting urban changes with recurrent neural networks from multitemporal Sentinel-2 data. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 214–217.
73. Khusni, U.; Dewangkoro, H.I.; Arymurthy, A.M. Urban area change detection with combining CNN and RNN from Sentinel-2 multispectral remote sensing data. In Proceedings of the 2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), Yogyakarta, Indonesia, 15–16 September 2020; pp. 171–175.
74. Huang, F.; Shen, G.; Hong, H.; Wei, L. Change detection of buildings with the utilization of a deep belief network and high-resolution remote sensing images. *Fractals* **2022**, *30*, 2240255. [CrossRef]
75. Pang, L.; Sun, J.; Chi, Y.; Yang, Y.; Zhang, F.; Zhang, L. CD-TransUNet: A hybrid transformer network for the change detection of urban buildings using l-band SAR images. *Sustainability* **2022**, *14*, 9847. [CrossRef]
76. Shafique, A.; Seydi, S.T.; Cao, G. BCD-Net: Building change detection based on fully scale connected U-Net and subpixel convolution. *Int. J. Remote Sens.* **2023**, *44*, 7416–7438. [CrossRef]
77. Xiong, J.; Liu, F.; Wang, X.; Yang, C. Siamese Transformer-Based Building Change Detection in Remote Sensing Images. *Sensors* **2024**, *24*, 1268. [CrossRef]
78. Ahmed, N.; Hoque, M.A.A.; Arabameri, A.; Pal, S.C.; Chakrabortty, R.; Jui, J. Flood susceptibility mapping in Brahmaputra floodplain of Bangladesh using deep boost, deep learning neural network, and artificial neural network. *Geocarto Int.* **2022**, *37*, 8770–8791. [CrossRef]
79. Lemenkova, P. Deep Learning Methods of Satellite Image Processing for Monitoring of Flood Dynamics in the Ganges Delta, Bangladesh. *Water* **2024**, *16*, 1141. [CrossRef]
80. Daudt, R.C.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067.
81. Yang, Y.; Zhu, D.; Qu, T.; Wang, Q.; Ren, F.; Cheng, C. Single-stream CNN with learnable architecture for multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
82. Chen, H.; Wu, C.; Du, B.; Zhang, L. Deep siamese multi-scale convolutional network for change detection in multi-temporal VHR images. In Proceedings of the 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 5–7 August 2019; pp. 1–4.
83. Zhang, M.; Shi, W. A feature difference convolutional neural network-based change detection method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232–7246. [CrossRef]
84. Iftene, M.; Larabi, M.E.A.; Karoui, M.S. End-to-end change detection in satellite remote sensing imagery. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 4356–4359.
85. Zhang, H.; Lin, M.; Yang, G.; Zhang, L. ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 28–42. [CrossRef]
86. Chen, P.; Li, C.; Zhang, B.; Chen, Z.; Yang, X.; Lu, K.; Zhuang, L. A region-based feature fusion network for VHR image change detection. *Remote Sens.* **2022**, *14*, 5577. [CrossRef]
87. Zhang, X.; He, L.; Qin, K.; Dang, Q.; Si, H.; Tang, X.; Jiao, L. SMD-Net: Siamese multi-scale difference-enhancement network for change detection in remote sensing. *Remote Sens.* **2022**, *14*, 1580. [CrossRef]
88. Wang, Q.; Li, M.; Li, G.; Zhang, J.; Yan, S.; Chen, Z.; Zhang, X.; Chen, G. High-resolution remote sensing image change detection method based on improved siamese U-Net. *Remote Sens.* **2023**, *15*, 3517. [CrossRef]
89. Wang, J.; Liu, F.; Jiao, L.; Wang, H.; Yang, H.; Liu, X.; Li, L.; Chen, P. SSCFNet: A spatial-spectral cross fusion network for remote sensing change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4000–4012. [CrossRef]
90. Zhang, W.; Zhang, Y.; Su, L.; Mei, C.; Lu, X. Difference-enhancement triplet network for change detection in multispectral images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
91. Yu, X.; Fan, J.; Chen, J.; Zhang, P.; Zhou, Y.; Han, L. NestNet: A multiscale convolutional neural network for remote sensing image change detection. *Int. J. Remote Sens.* **2021**, *42*, 4898–4921. [CrossRef]
92. Zhang, X.; Yue, Y.; Gao, W.; Yun, S.; Su, Q.; Yin, H.; Zhang, Y. DifUnet++: A satellite images change detection network based on UNet++ and differential pyramid. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]

93. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A densely connected siamese network for change detection of VHR images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
94. Qian, J.; Xia, M.; Zhang, Y.; Liu, J.; Xu, Y. TCDNet: Trilateral change detection network for Google Earth image. *Remote Sens.* **2020**, *12*, 2669. [CrossRef]
95. Zhang, W.; Lu, X. The spectral-spatial joint learning for change detection in multispectral imagery. *Remote Sens.* **2019**, *11*, 240. [CrossRef]
96. Ye, Y.; Zhou, L.; Zhu, B.; Yang, C.; Sun, M.; Fan, J.; Fu, Z. Feature decomposition-optimization-reorganization network for building change detection in remote sensing images. *Remote Sens.* **2022**, *14*, 722. [CrossRef]
97. Lei, J.; Gu, Y.; Xie, W.; Li, Y.; Du, Q. Boundary extraction constrained siamese network for remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
98. Ding, L.; Zhu, K.; Peng, D.; Tang, H.; Yang, K.; Bruzzone, L. Adapting segment anything model for change detection in VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–11. [CrossRef]
99. Zhao, X.; Ding, W.; An, Y.; Du, Y.; Yu, T.; Li, M.; Tang, M.; Wang, J. Fast segment anything. *arXiv* **2023**, arXiv:2306.12156.
100. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
101. Jiang, M.; Zhang, X.; Sun, Y.; Feng, W.; Gan, Q.; Ruan, Y. AFSNet: Attention-guided full-scale feature aggregation network for high-resolution remote sensing image change detection. *Giscience Remote Sens.* **2022**, *59*, 1882–1900. [CrossRef]
102. Adriano, B.; Yokoya, N.; Xia, J.; Miura, H.; Liu, W.; Matsuoka, M.; Koshimura, S. Learning from multimodal and multitemporal earth observation data for building damage mapping. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 132–143. [CrossRef]
103. Li, H.; Wang, L.; Cheng, S. HARNU-Net: Hierarchical attention residual nested U-Net for change detection in remote sensing images. *Sensors* **2022**, *22*, 4626. [CrossRef]
104. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 1194–1206. [CrossRef]
105. Lu, D.; Wang, L.; Cheng, S.; Li, Y.; Du, A. CANet: A combined attention network for remote sensing image change detection. *Information* **2021**, *12*, 364. [CrossRef]
106. Li, X.; Lei, L.; Sun, Y.; Li, M.; Kuang, G. Multimodal bilinear fusion network with second-order attention-based channel selection for land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1011–1026. [CrossRef]
107. Ma, J.; Shi, G.; Li, Y.; Zhao, Z. MAFF-Net: Multi-attention guided feature fusion network for change detection in remote sensing images. *Sensors* **2022**, *22*, 888. [CrossRef] [PubMed]
108. Chen, J.; Fan, J.; Zhang, M.; Zhou, Y.; Shen, C. MSF-Net: A multiscale supervised fusion network for building change detection in high-resolution remote sensing images. *IEEE Access* **2022**, *10*, 30925–30938. [CrossRef]
109. Xu, X.; Zhou, Y.; Lu, X.; Chen, Z. FERA-Net: A building change detection method for high-resolution remote sensing imagery based on residual attention and high-frequency features. *Remote Sens.* **2023**, *15*, 395. [CrossRef]
110. Zhong, H.; Wu, C. T-UNet: Triplet UNet for change detection in high-resolution remote sensing images. *arXiv* **2023**, arXiv:2308.02356. [CrossRef]
111. Sivasankari, A.; Jayalakshmi, S. Land cover clustering for change detection using deep belief network. In Proceedings of the 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 16–18 March 2022; pp. 815–822.
112. Jia, M.; Zhao, Z. Change detection in synthetic aperture radar images based on a generalized gamma deep belief networks. *Sensors* **2021**, *21*, 8290. [CrossRef]
113. Samadi, F.; Akbarizadeh, G.; Kaabi, H. Change detection in SAR images using deep belief network: A new training approach based on morphological images. *IET Image Process.* **2019**, *13*, 2255–2264. [CrossRef]
114. Mou, L.; Zhu, X.X. A recurrent convolutional neural network for land cover change detection in multispectral images. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4363–4366.
115. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 924–935. [CrossRef]
116. Lyu, H.; Lu, H.; Mou, L.; Li, W.; Wright, J.; Li, X.; Li, X.; Zhu, X.X.; Wang, J.; Yu, L.; et al. Long-term annual mapping of four cities on different continents by applying a deep information learning method to landsat data. *Remote Sens.* **2018**, *10*, 471. [CrossRef]
117. Sun, S.; Mu, L.; Wang, L.; Liu, P. L-UNet: An LSTM network for remote sensing image change detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
118. Zhao, Y.; Chen, P.; Chen, Z.; Bai, Y.; Zhao, Z.; Yang, X. A triple-stream network with cross-stage feature fusion for high-resolution image change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–17. [CrossRef]
119. Zhu, Y.; Lv, K.; Yu, Y.; Xu, W. Edge-guided parallel network for VHR remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 7791–7803. [CrossRef]
120. Sefrin, O.; Riese, F.M.; Keller, S. Deep learning for land cover change detection. *Remote Sens.* **2020**, *13*, 78. [CrossRef]
121. Jing, R.; Liu, S.; Gong, Z.; Wang, Z.; Guan, H.; Gautam, A.; Zhao, W. Object-Based change detection for VHR remote sensing images based on a trisiamese-LSTM. *Int. J. Remote Sens.* **2020**, *41*, 6209–6231. [CrossRef]

122. Bandara, W.G.C.; Patel, V.M. A transformer-based siamese network for change detection. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 207–210.
123. Yuan, P.; Zhao, Q.; Zhao, X.; Wang, X.; Long, X.; Zheng, Y. A transformer-based siamese network and an open optical dataset for semantic change detection of remote sensing images. *Int. J. Digit. Earth* **2022**, *15*, 1506–1525. [CrossRef]
124. Yan, T.; Wan, Z.; Zhang, P. Fully transformer network for change detection of remote sensing images. In Proceedings of the Asian Conference on Computer Vision, Macao, China, 4–8 December 2022; pp. 1691–1708.
125. Zhang, C.; Wang, L.; Cheng, S.; Li, Y. SwinSUNet: Pure transformer network for remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
126. Pan, J.; Bai, Y.; Shu, Q.; Zhang, Z.; Hu, J.; Wang, M. M-Swin: Transformer-based Multi-scale Feature Fusion Change Detection Network within Cropland for Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–16. [CrossRef]
127. Song, L.; Xia, M.; Xu, Y.; Weng, L.; Hu, K.; Lin, H.; Qian, M. Multi-granularity siamese transformer-based change detection in remote sensing imagery. *Eng. Appl. Artif. Intell.* **2024**, *136*, 108960. [CrossRef]
128. Xu, X.; Li, J.; Chen, Z. TCIANet: Transformer-based context information aggregation network for remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 1951–1971. [CrossRef]
129. Ma, J.; Duan, J.; Tang, X.; Zhang, X.; Jiao, L. Eatder: Edge-assisted adaptive transformer detector for remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 1–15. [CrossRef]
130. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
131. Song, X.; Hua, Z.; Li, J. PSTNet: Progressive sampling transformer network for remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8442–8455. [CrossRef]
132. Zhang, K.; Zhao, X.; Zhang, F.; Ding, L.; Sun, J.; Bruzzone, L. Relation changes matter: Cross-temporal difference transformer for change detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–5. [CrossRef]
133. Ding, L.; Zhang, J.; Guo, H.; Zhang, K.; Liu, B.; Bruzzone, L. Joint spatio-temporal modeling for semantic change detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [CrossRef]
134. Zhou, Y.; Huo, C.; Zhu, J.; Huo, L.; Pan, C. DCAT: Dual cross-attention-based transformer for change detection. *Remote Sens.* **2023**, *15*, 2395. [CrossRef]
135. Noman, M.; Fiaz, M.; Cholakkal, H.; Narayan, S.; Anwer, R.M.; Khan, S.; Khan, F.S. Remote sensing change detection with transformers trained from scratch. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4704214. [CrossRef]
136. Yuan, J.; Wang, L.; Cheng, S. STransUNet: A siamese transUNet-based remote sensing image change detection network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9241–9253. [CrossRef]
137. Deng, Y.; Meng, Y.; Chen, J.; Yue, A.; Liu, D.; Chen, J. TChange: A hybrid transformer-CNN change detection network. *Remote Sens.* **2023**, *15*, 1219. [CrossRef]
138. Wang, G.; Li, B.; Zhang, T.; Zhang, S. A network combining a transformer and a convolutional neural network for remote sensing image change detection. *Remote Sens.* **2022**, *14*, 2228. [CrossRef]
139. Li, Q.; Zhong, R.; Du, X.; Du, Y. TransUNetCD: A hybrid transformer network for change detection in optical remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–19. [CrossRef]
140. Liu, M.; Chai, Z.; Deng, H.; Liu, R. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4297–4306. [CrossRef]
141. Yin, M.; Chen, Z.; Zhang, C. A CNN-transformer network combining CBAM for change detection in high-resolution remote sensing images. *Remote Sens.* **2023**, *15*, 2406. [CrossRef]
142. Wang, W.; Tan, X.; Zhang, P.; Wang, X. A CBAM based multiscale transformer fusion approach for remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 6817–6825. [CrossRef]
143. Song, X.; Hua, Z.; Li, J. LHDACT: Lightweight hybrid dual attention CNN and transformer network for remote sensing image change detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
144. Jiang, M.; Chen, Y.; Dong, Z.; Liu, X.; Zhang, X.; Zhang, H. Multiscale fusion CNN-transformer network for high-resolution remote sensing image change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5280–5293. [CrossRef]
145. Tang, W.; Wu, K.; Zhang, Y.; Zhan, Y. A siamese network based on multiple attention and multilayer transformers for change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5219015. [CrossRef]
146. Niu, Y.; Guo, H.; Lu, J.; Ding, L.; Yu, D. SMNet: Symmetric multi-task network for semantic change detection in remote sensing images based on CNN and transformer. *Remote Sens.* **2023**, *15*, 949. [CrossRef]
147. Li, W.; Xue, L.; Wang, X.; Li, G. Mctnet: A multi-scale cnn-transformer network for change detection in optical remote sensing images. In Proceedings of the 2023 26th International Conference on Information Fusion (FUSION), Charleston, SC, USA, 27–30 July 2023; pp. 1–5.
148. Tang, X.; Zhang, T.; Ma, J.; Zhang, X.; Liu, F.; Jiao, L. Wnet: W-shaped hierarchical network for remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5615814. [CrossRef]
149. Zhang, X.; Cheng, S.; Wang, L.; Li, H. Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [CrossRef]

150. Feng, Y.; Xu, H.; Jiang, J.; Liu, H.; Zheng, J. ICIF-Net: Intra-scale cross-interaction and inter-scale feature fusion network for bitemporal remote sensing images change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
151. Fu, Z.; Li, J.; Ren, L.; Chen, Z. Slddnet: Stage-wise short and long distance dependency network for remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–19. [CrossRef]
152. Zhang, C.; Wang, L.; Cheng, S. HCGNet: A Hybrid Change Detection Network Based on CNN and GNN. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–12. [CrossRef]
153. Zhu, Y.; Li, Q.; Lv, Z.; Falco, N. Novel land cover change detection deep learning framework with very small initial samples using heterogeneous remote sensing images. *Remote Sens.* **2023**, *15*, 4609. [CrossRef]
154. Liu, M.; Shi, Q.; Marinoni, A.; He, D.; Liu, X.; Zhang, L. Super-resolution-based change detection network with stacked attention module for images with different resolutions. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–18. [CrossRef]
155. Tian, J.; Peng, D.; Guan, H.; Ding, H. RACDNet: Resolution-and alignment-aware change detection network for optical remote sensing imagery. *Remote Sens.* **2022**, *14*, 4527. [CrossRef]
156. Liu, M.; Shi, Q.; Liu, P.; Wan, C. Siamese generative adversarial network for change detection under different scales. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2543–2546.
157. Prexl, J.; Saha, S.; Zhu, X.X. Mitigating spatial and spectral differences for change detection using super-resolution and unsupervised learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3113–3116.
158. Li, S.; Wang, Y.; Cai, H.; Lin, Y.; Wang, M.; Teng, F. MF-SRCDNet: Multi-feature fusion super-resolution building change detection framework for multi-sensor high-resolution remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *119*, 103303. [CrossRef]
159. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
160. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; pp. 38–45.
161. Liu, M.; Shi, Q.; Li, J.; Chai, Z. Learning token-aligned representations with multimodel transformers for different-resolution change detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
162. Sun, B.; Liu, Q.; Yuan, N.; Tan, J.; Gao, X.; Yu, T. Spectral token guidance transformer for multisource images change detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2559–2572. [CrossRef]
163. Chen, H.; Zhang, H.; Chen, K.; Zhou, C.; Chen, S.; Zou, Z.; Shi, Z. Continuous cross-resolution remote sensing image change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5623320. [CrossRef]
164. Chen, H.; Wu, C.; Du, B.; Zhang, L.; Wang, L. Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2848–2864. [CrossRef]
165. Benedetti, P.; Ienco, D.; Gaetano, R.; Ose, K.; Pensa, R.G.; Dupuy, S. M3Fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4939–4949. [CrossRef]
166. Ebel, P.; Saha, S.; Zhu, X.X. Fusing multi-modal data for supervised change detection. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 243–249. [CrossRef]
167. Hafner, S.; Nascetti, A.; Azizpour, H.; Ban, Y. Sentinel-1 and Sentinel-2 data fusion for urban change detection using a dual stream u-net. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
168. He, X.; Zhang, S.; Xue, B.; Zhao, T.; Wu, T. Cross-modal change detection flood extraction based on convolutional neural network. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103197. [CrossRef]
169. Li, H.; Zhu, F.; Zheng, X.; Liu, M.; Chen, G. MSCDUNet: A deep learning framework for built-Up area change detection integrating multispectral, SAR, and VHR data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5163–5176. [CrossRef]
170. Chen, H.; Wu, C.; Du, B.; Zhang, L. DSDANet: Deep siamese domain adaptation convolutional neural network for cross-domain change detection. *arXiv* **2020**, arXiv:2006.09225.
171. Zhang, C.; Feng, Y.; Hu, L.; Tapete, D.; Pan, L.; Liang, Z.; Cigna, F.; Yue, P. A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *109*, 102769. [CrossRef]
172. Luppino, L.T.; Hansen, M.A.; Kampffmeyer, M.; Bianchi, F.M.; Moser, G.; Jenssen, R.; Anfinsen, S.N. Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *5*, 60–72. [CrossRef]
173. Wu, Y.; Li, J.; Yuan, Y.; Qin, A.; Miao, Q.G.; Gong, M.G. Commonality autoencoder: Learning common features for change detection from heterogeneous images. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 4257–4270. [CrossRef]
174. Farahani, M.; Mohammadzadeh, A. Domain adaptation for unsupervised change detection of multisensor multitemporal remote-sensing images. *Int. J. Remote Sens.* **2020**, *41*, 3902–3923. [CrossRef]
175. Jiang, X.; Li, G.; Liu, Y.; Zhang, X.P.; He, Y. Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1551–1566. [CrossRef]
176. Touati, R.; Mignotte, M.; Dahmane, M. Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 588–600. [CrossRef]

177. Zheng, X.; Chen, X.; Lu, X.; Sun, B. Unsupervised change detection by cross-resolution difference learning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]
178. Wei, L.; Chen, G.; Zhou, Q.; Liu, C.; Cai, C. Cross-mapping net: Unsupervised change detection from heterogeneous remote sensing images using a transformer network. In Proceedings of the 2023 8th International Conference on Computer and Communication Systems (ICCCS), Guangzhou, China, 21–24 April 2023; pp. 1021–1026.
179. Lu, T.; Zhong, X.; Zhong, L. mSwinUNet: A multi-modal U-shaped swin transformer for supervised change detection. *J. Intell. Fuzzy Syst.* **2024**. *Preprint*.
180. Hu, X.; Zhang, P.; Ban, Y.; Rahnemoonfar, M. GAN-based SAR and optical image translation for wildfire impact assessment using multi-source remote sensing data. *Remote Sens. Environ.* **2023**, *289*, 113522. [CrossRef]
181. Zhao, T.; Wang, L.; Zhao, C.; Liu, T.; Ohtsuki, T. Heterogeneous image change detection based on deep image translation and feature refinement-aggregation. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 1705–1709.
182. Manocha, A.; Afaq, Y. Optical and SAR images-based image translation for change detection using generative adversarial network (GAN). *Multimed. Tools Appl.* **2023**, *82*, 26289–26315. [CrossRef]
183. Du, Z.; Li, X.; Miao, J.; Huang, Y.; Shen, H.; Zhang, L. Concatenated deep learning framework for multi-task change detection of optical and SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 719–731. [CrossRef]
184. Wang, M.; Huang, L.; Tang, B.H.; Le, W.; Tian, Q. TDSCCNet: Twin-depthwise separable convolution connect network for change detection with heterogeneous images. *Geocarto Int.* **2024**, *39*, 2329673. [CrossRef]
185. Su, Z.; Wan, G.; Zhang, W.; Wei, Z.; Wu, Y.; Liu, J.; Jia, Y.; Cong, D.; Yuan, L. Edge-bound change detection in multisource remote sensing images. *Electronics* **2024**, *13*, 867. [CrossRef]
186. Xu, J.; Luo, C.; Chen, X.; Wei, S.; Luo, Y. Remote sensing change detection based on multidirectional adaptive feature fusion and perceptual similarity. *Remote Sens.* **2021**, *13*, 3053. [CrossRef]
187. Peng, X.; Zhong, R.; Li, Z.; Li, Q. Optical remote sensing image change detection based on attention mechanism and image difference. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 7296–7307. [CrossRef]
188. Ienco, D.; Interdonato, R.; Gaetano, R.; Minh, D.H.T. Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 11–22. [CrossRef]
189. Wang, L.; Wang, L.; Wang, H.; Wang, X.; Bruzzone, L. SPCNet: A subpixel convolution-based change detection network for hyperspectral images with different spatial resolutions. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [CrossRef]
190. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource remote sensing data classification based on convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 937–949. [CrossRef]
191. Chen, Y.; Li, C.; Ghamisi, P.; Jia, X.; Gu, Y. Deep fusion of remote sensing data for accurate classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1253–1257. [CrossRef]
192. Feng, Q.; Zhu, D.; Yang, J.; Li, B. Multisource hyperspectral and LiDAR data fusion for urban land-use mapping based on a modified two-branch convolutional neural network. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 28. [CrossRef]
193. Mohla, S.; Pande, S.; Banerjee, B.; Chaudhuri, S. Fusatnet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and lidar classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 92–93.
194. Ma, W.; Karakuş, O.; Rosin, P.L. AMM-FuseNet: Attention-based multi-modal image fusion network for land cover mapping. *Remote Sens.* **2022**, *14*, 4458. [CrossRef]
195. Liu, J.; Gong, M.; Qin, K.; Zhang, P. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *29*, 545–559. [CrossRef]
196. Liu, Z.; Li, G.; Mercier, G.; He, Y.; Pan, Q. Change detection in heterogenous remote sensing images via homogeneous pixel transformation. *IEEE Trans. Image Process.* **2017**, *27*, 1822–1834. [CrossRef]
197. Roy, S.K.; Deria, A.; Hong, D.; Rasti, B.; Plaza, A.; Chanussot, J. Multimodal fusion transformer for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–20. [CrossRef]
198. Luppino, L.T.; Bianchi, F.M.; Moser, G.; Anfinsen, S.N. Unsupervised image regression for heterogeneous change detection. *arXiv* **2019**, arXiv:1909.05948. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.