remote sensing

MDPI

Review

A Survey on Deep Learning-Based Change Detection from High-Resolution Remote Sensing Images

Huiwei Jiang [1,2], Min Peng [3], Yuanjun Zhong [4,*], Haofeng Xie [2], Zemin Hao [3], Jingming Lin [2], Xiaoli Ma [4] and Xiangyun Hu [2,5]

[1] National Geomatics Center of China, Beijing 100830, China; huiwei_jiang@whu.edu.cn
[2] School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; xiehaofeng@whu.edu.cn (H.X.); linjingming@whu.edu.cn (J.L.); huxy@whu.edu.cn (X.H.)
[3] Geotechnical Investigation & Surveying Research Institute Co., Ltd., Shenyang 110004, China; minpeng@whu.edu.cn (M.P.); haozemin@whu.edu.cn (Z.H.)
[4] Guangdong Surveying and Mapping Institute of Lands and Resource Department, No. 13 Guangpu Middle Road, Huangpu District, Guangzhou 510663, China; xiaoli.ma1010@gmail.com
[5] Institute of Artificial Intelligence in Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
* Correspondence: yuanjun.zhong.gd@gmail.com; Tel.: +86-20-8770-2345; Fax: +86-20-8770-3807

Abstract: Change detection based on remote sensing images plays an important role in the field of remote sensing analysis, and it has been widely used in many areas, such as resources monitoring, urban planning, disaster assessment, etc. In recent years, it has aroused widespread interest due to the explosive development of artificial intelligence (AI) technology, and change detection algorithms based on deep learning frameworks have made it possible to detect more delicate changes (such as the alteration of small buildings) with the help of huge amounts of remote sensing data, especially high-resolution (HR) data. Although there are many methods, we still lack a deep review of the recent progress concerning the latest deep learning methods in change detection. To this end, the main purpose of this paper is to provide a review of the available deep learning-based change detection algorithms using HR remote sensing images. The paper first describes the change detection framework and classifies the methods from the perspective of the deep network architectures adopted. Then, we review the latest progress in the application of deep learning in various granularity structures for change detection. Further, the paper provides a summary of HR datasets derived from different sensors, along with information related to change detection, for the potential use of researchers. Simultaneously, representative evaluation metrics for this task are investigated. Finally, a conclusion of the challenges for change detection using HR remote sensing images, which must be dealt with in order to improve the model's performance, is presented. In addition, we put forward promising directions for future research in this area.

Keywords: deep learning; change detection; high-resolution; remote sensing images

Citation: 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. https://doi.org/10.3390/rs14071552

Academic Editor: Filiberto Pla

Received: 27 February 2022
Accepted: 21 March 2022
Published: 23 March 2022

1. Introduction

Change detection based on remote sensing (RS) technology is used to discover and identify differences in ground objects using two or more images in the same geographical location [1]. The development of remote sensing technology has attracted the interest of many researchers, and it has been applied in many fields, such as disaster monitoring [2,3], resource surveys [4–6], and urban planning [7,8].

In the last few decades, change detection via multitemporal images has received intensive scrutiny depending on the requirements and conditions, and enormous efforts have been made towards developing various change detection methods, including traditional methods [9,10] and deep learning-based methods [11]. Before the prevalence of deep learning, the change detection problem was mainly solved by handcrafted features

derived from complicated feature extractors. With the increasing application of automatic change detection with remote sensing images, researchers have made enormous efforts in developing various change detection methods. However, it is challenging to select the most suitable method in practice, and the limitations of the traditional methods are becoming more and more obvious. On the one hand, remote sensing data sources are increasingly diversifying, image spatial resolution is gradually improving, and image details are being progressively enriched. The poor expressiveness of features extracted by traditional methods features reduces change detection accuracy significantly and is susceptible to the influences of factors such as seasonal variation, illumination conditions, satellite sensors, and solar altitude angle. On the other hand, although some methods can reduce false changes by combining shape and textural features, this is time-consuming and tedious. In addition, handcrafted features rely heavily on specific domain knowledge, which seriously reduces the automation capability of change detection technology. In general, traditional methods that require expert knowledge are typically suboptimal, and the empirical feature is weak in representing images.

Recently, deep learning-based (DL-based) methods have attracted much attention. They have achieved success over traditional methods in various areas, such as image classification [12], semantic segmentation [13], object tracking [14], and natural language processing [15]. Compared to conventional methods based on manual techniques, DL-based methods can learn features from given data and significantly decrease the demand for expert domain knowledge. Moreover, DL-based methods offer greater understanding of complex scenes due to their nonlinear characterization and excellent feature extraction capabilities [16,17], and they achieve a performance far beyond that of traditional methods.

Thanks to these advantages, DL-based methods have exponentially grown in use for solving remote sensing problems, for instance, in image classification [18,19], object detection [20,21], pose estimation [22,23], scene upstanding [24], and image segmentation [25]. Meanwhile, the highly discriminative features of DL methods can be used to solve change detection problems. Numerous studies have been conducted on solving change detection problems using DL-based technology by learning more helpful information about changes via well-designed loss functions. Figure 1 illustrates the increasing number of papers that focus on remote sensing, published in the last three decades. As can be seen from the figure, most researchers have been concerned with change detection, and the number of publications in this area has increased annually. It is worth mentioning that the publications on change detection involving deep learning have increased significantly since 2016. This figure shows that change detection methods based on DL rapidly overtake remote sensing change detection applications.

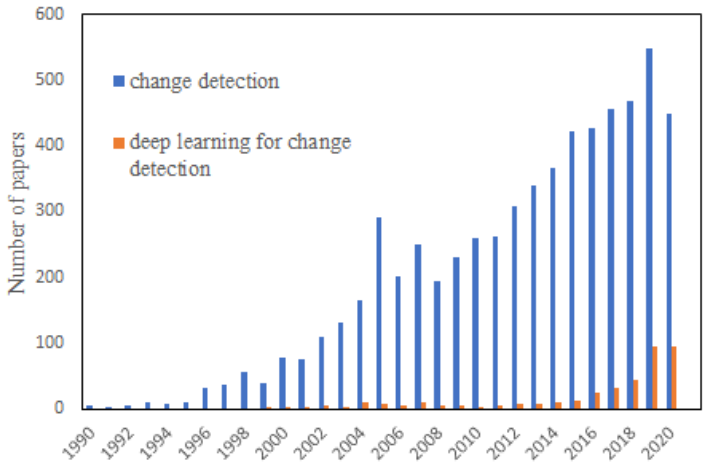


Figure 1. The statistics for published literature from 1990 to 2020. Data were collected by advanced searching on Web of Science (allintitle1: ((TS = ("deep learning")) OR TS = ("neural network")) AND TI = ("change detection"), allintitle2: TI = ("change detection")).

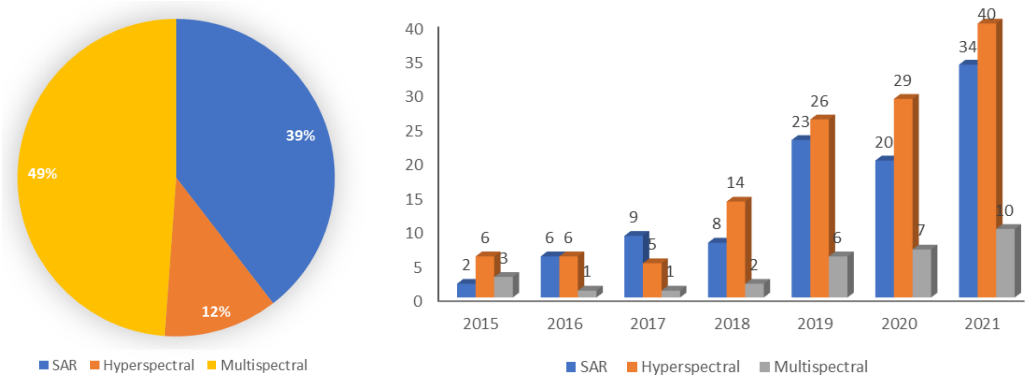
In general, the most popular solution for change detection is to use bi-temporal remote sensing images from the same sensors. Bi-temporal data mainly include synthetic aperture radar (SAR) [26,27] and optical images. Of the two types of data mentioned above, optical images are most commonly used in change detection because they can provide rich spectral and spatial information. According to the spectral resolution, there are three types of optical images: hyperspectral, multispectral, and panchromatic images. However, in order to preserve an image's spectral information while improving the spatial resolution as much as possible, multispectral and panchromatic images are usually fused together in practical applications. To illustrate the process using different data sources for change detection, we conducted a simple statistical analysis. In Figure 2, the statistics display the distribution of relevant change detection data sources based on deep learning from 2015 to April 2020, containing SAR, multispectral, and hyperspectral images. Apparently, SAR and multispectral images have always been the most used data types. However, speckle noise brings more challenges to the research of SAR change detection than optical ones. In addition, the side-looking geometry and different microwave scattering contribution patterns also introduce typical problems such as layover and shadowing. Especially in dense urban areas, buildings are often partially occluded, and signals of different buildings are often mixed, which make them difficult to interpret [26,28]. In fact, the proportions of multispectral images should be larger because the default data source in much of the literature is high-resolution satellite images, which belong to the multispectral image. However, this time, we only searched according to the keyword of "aerial" and "satellite". Meanwhile, hyperspectral images are receiving more and more attention with the rapid development of remote sensing information technology, which not only possess high spectral resolution, but also have continuously improving spatial resolution. Compared with hyperspectral images, multispectral data can be obtained in a more economical and stable way, with a relatively higher temporal and broader spatial resolution. By improving the spatial resolution, the Earth can be observed at a finer scale [29]. Delicate structural information of ground objects, even at small sizes, can be reflected in high-resolution (HR) or very high-resolution (VHR) images [30]; here, the resolution mainly refers to the spatial resolution. Moreover, HR or VHR images are extraordinarily rich in color, texture, and other features. Consequently, multispectral images have been the primary data source for many remote sensing applications, especially change detection. The main aim of this paper is to provide a summary of change detection methods based on deep learning, especially using HR remote sensing images. In particular, we focus on Earth observation satellites, such as Gaofen series [31,32], Worldview series [33], ZY series [31,33], Quickbird [33], and VHR aerial imagery, which are the most frequently used data sources in DL-based change detection methods, due to their accessibility.

At present, while many change detection reviews exist, most of them have mainly focused on traditional change detection approaches [10,34–36]. Recently, a few researchers reviewed deep learning change detection [37–39]. The objective of this survey is to summarize and classify the deep learning-based methods for change detection using HR images to assist related research. The proposed survey is different from published surveys. We mainly concentrate on HR remote sensing datasets, rather than mixing all available datasets, which makes it difficult for beginners to find a straightforward way to begin their research. Further, our review outlines the state-of-the-art change detection methods using HR remote sensing images, including a detailed analysis of each component of the complete change detection workflow.

The main contributions of this paper are fourfold.

(1) We provide a systematic review of change detection based on deep learning using HR remote sensing images, which covers the most popular feature extraction network and the construct mechanisms of every part of this framework.
(2) We analyze the granularity of change detection algorithms according to the detection unit, which allows us to apply a reasonable method independently of the particular applications.

(3) We present the popular dataset used for change detection from HR remote sensing images in detail, and the representative evaluation metrics for this task are investigated.
(4) We present several suggestions for future research on change detection using remote sensing data.



(**a**) (**b**)

Figure 2. Distribution of different image sources from 2015 to 2020. Data were collected by advanced searching on Web of Science. (**a**) Proportional diagram of the total number of papers on optical images from various sources. (**b**) Annual statistical bar chart of the number of papers related to multisource images. (allintitle for example: ((((TS = ("deep learning")) OR TS = ("neural network")) AND TI = ("change detection"))) AND (TS = ("synthetic aperture radar") OR TS = (SAR))).

The contents of this paper are organized as follows. Firstly, the framework of change detection based on DL and the most commonly used strategies for each part are presented in Section 2. Then, in Section 3, we review the change detection methods according to the detection unit. Subsequently, we list the openly available datasets and the standard evaluation metrics used for change detection in Section 4. Finally, in Section 5, we outline the conclusions of the review and provide some promising directions for further studies.

2. Deep Learning-Based Change Detection Algorithms

The continuous development and improvement of neural networks (NNs) has opened up new opportunities for change detection tasks in the last decade. In recent years, enormous efforts have been made to use DL methods for change detection in remote sensing images. A wide variety of research has indicated the superiority of these methods over the conventional approach, as they learn representative and discriminative features from a vast array of samples.

2.1. Change Detection Framework

Since the input for change detection is multitemporal data, the DL-based change detection method starts with data collection, taken at the same location in two or more periods.

Therefore, the input images should be registered first to ensure consistency in the geographical information. It should be noted that despite the better registration algorithms, errors can still occur. Then, representative features can be determined by a feature extractor that takes into account color, texture, and gradient, as well as the spatial geometric relationship between the images. Afterwards, the change features are generated from the distinguishing features and used to locate and determine the intensity of change information, which is called feature fusion. The above two steps can be defined together as change information extraction. Finally, the optimization process is performed with the change evaluation criteria to obtain the final change map. As for the implementation of change detection, the current change detection methods based on DL are designed to solve several or all sub-problems, i.e., feature extraction, feature fusion, and optimization. The overall

framework of the change detection technique can be summarized as change information extraction, network optimization, and accuracy evaluation, as depicted in Figure 3.

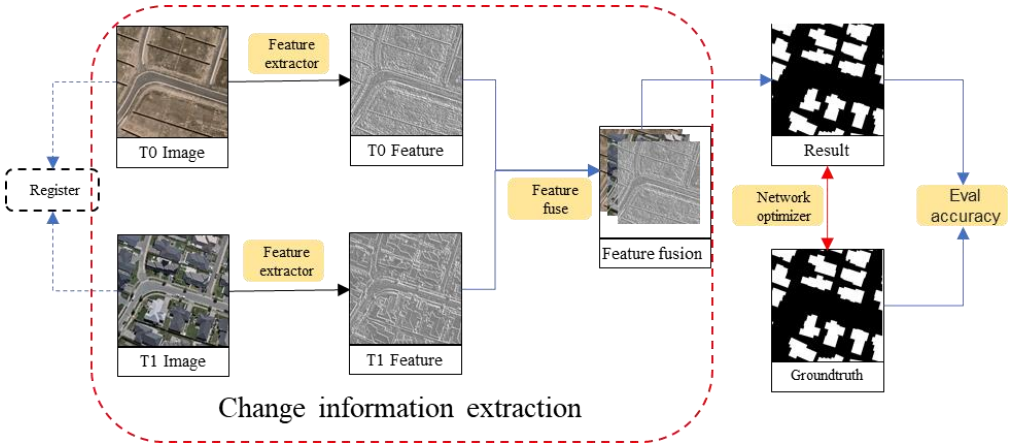


Figure 3. The general framework of change detection, in which the yellow-filled, rounded rectangles illustrate the scope of this paper.

2.2. Mainstrem Feature Extraction Network

Deep neural networks (DNNs) [40] with multiple hidden layers are designed to simulate a biological neural system. DNNs have been introduced into the change detection task for their discriminative and robust feature representation abilities.

Generally speaking, DNNs using a multilayer perceptron (MLP) [40] convert the input feature space into change feature space to obtain the classification. For example, a restricted Boltzmann machine (RBM) [41] with a visible layer and a hidden layer can be used to extract effective information from complex scenes.

In terms of the deep architecture used to extract the discriminative feature (such as edge, texture, or color), there are five major types: autoencoder (AE)-based, recurrent neural network (RNN)-based, convolution neural network (CNN)-based, generative adversarial network (GAN)-based, and transformers-based methods. Figure 4 shows the literature distribution of common network architectures used for change detection.

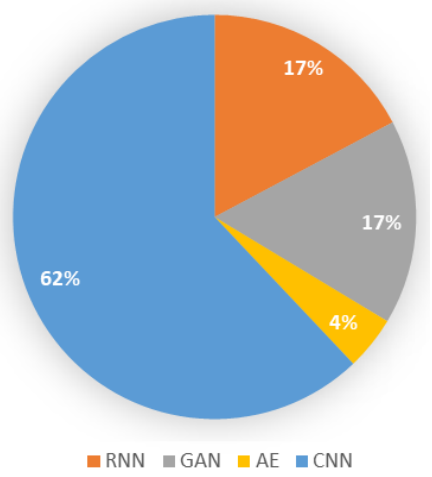


Figure 4. The percentage distribution for network architectures of commonly used for change detection. Data were collected by advanced searching on Web of Science, with a total of 116.

2.2.1. Encoder–Decoder and Autoencoder (AE) Models

As a typical and popular network structure, encoder–decoder models consist of a family of models, e.g., AE models, which learn feature mapping from an input space to an

output space. The model consists of an input layer, a hidden layer, and an output layer, as shown in Figure 5a. The encoder compresses the input vector x into a hidden layer, defined as below:

$$h(x) = f(W \cdot x + b), \tag{1}$$

where $h \in R^n$ denotes the output of hidden layers, $f(\cdot)$ is a nonlinear function, such as the logistic sigmoid function $(1 + \exp(-x))$, $W \in R^{n \times m}$ represents the weight matrix of the encoder, and $x \in R^m$ represents the input.

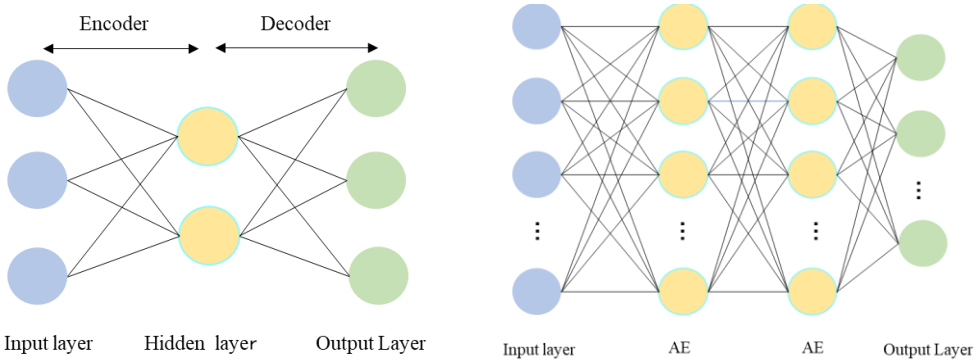


Figure 5. The architectures of (**a**) autoencoders and (**b**) stacked autoencoders.

Decoding reconstructs the hidden layer into an output layer, and the formula is as follows:

$$\widetilde{x} = f\left(W' \cdot h + b'\right), \tag{2}$$

where $\widetilde{x} \in R^m$ represents the reconstructed output, $W' \in R^{m \times n}$ denotes the weight matrix of the decoder, and b' is the bias of the hidden and output layers.

Autoencoders have the same input and output as a particular encoder–decoder model. Theoretically, an AE can reduce the dimensionality of features. Compared with a principal component analysis (PCA) algorithm, its performance is better, mainly due to the strong ability of neural networks to learn features. AE models are typically trained by minimizing the reconstruction loss function $L(x, \widetilde{x})$ using a distance metric, such as Euclidean distance, which aims to close the input and output units. These models are popular for use in image-to-image translation issues [42] and the problem of sequence-to-sequence models in natural language processing (NLP) [43]. Given its strong learning abilities, AE and its improvements (e.g., stacked AEs [44], stacked denoising AEs [45], stacked fisher AEs [46], sparse AEs [47], denoising AEs [48], fuzzy AEs [49], and contractive AEs) often serve as components in different networks. A stacked autoencoder (SAE) is a neural network that stacks multiple autoencoders, with the output of each hidden layer serving as the input for the next hidden layer. Figure 5b gives a simple representation of an SAE. For change detection tasks employing convolutional networks, AE models are widely used as the feature extractor in unsupervised tasks, because they can avoid the problem of needing extensive manual annotations [50]. Most of the above-mentioned autoencoder-based methods can obtain various features, but in a mixed and complex scene, they are unable to extract the best identification features and then distinguish the scene without thoroughly mining the scene class information.

2.2.2. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

RNNs are effective in capturing sequential relationships and have been widely used in fields associated with sequential data, such as speech [51], text [52], videos [53], and time-series [54]. As shown in Figure 6a, RNNs can be unfolded into a chain of series-connected units (i.e., RNN cells).

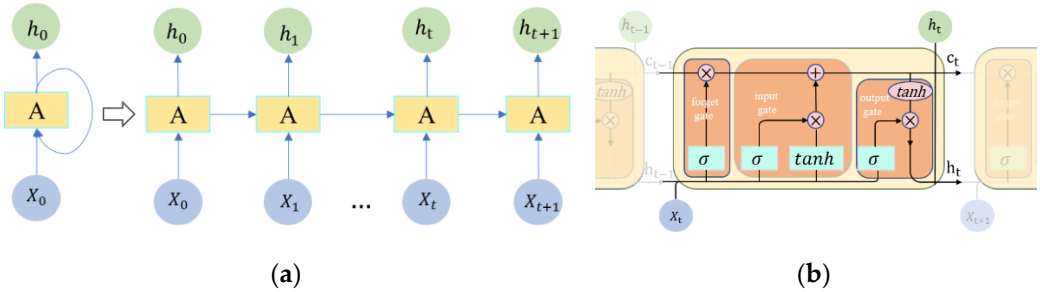


Figure 6. The architecture of (**a**) RNN; (**b**) LSTM.

A change detection task usually involves obtaining change information from two or more images [38,55]. From this perspective, RNNs can learn essential information and effectively establish the change relationship between multiperiod sequential remote sensing images to detect changes [56–59].

However, with the increasing length of the sequence/time, the problem of the vanishing gradient means RNNs are difficult to train. In order to alleviate the problem, long short-term memory (LSTM) [60] is introduced to the hidden layer unit in a classical RNN [46,57,58,61] via a gate mechanism, with an input gate, output gate, and forget gate, as shown in Figure 6b. The gate unit plays a role similar to the "switch" in the circuit, which is used to control the on and off of information transmission. The core of the LSTM is the cell state, which is controlled by the gates. The main function of the gates is to decide what to retain and what to omit from the memory. Therefore, LSTM networks have been applied to obtain change information from HR remote sensing data. In the change detection task, the cell state contains the change information of the multitemporal images learned by the three gates [58]. To their advantage, the models have sufficient generalization capacity, which can be exploited when training from different data domains [62]. Moreover, inspired by the idea of knowledge migration, the authors in [52] offered an RNN-based framework to monitor four cities' dynamic changes between years. Overall, the RNN-based methods have certain advantages when faced with the problems of temporal spectral variance and insufficient sampling and have shown superiority in long-term urban change detection [56].

2.2.3. Convolution Neural Networks (CNNs)

A CNN is a classic feed-forward neural network composed of convolutional layers, nonlinear layers, and pooling layers [63]. From Figure 4, we can see that CNNs are the most popular network architecture. As shown in Figure 7, the network is designed to imitate biological vision, so that it can effectively obtain the characteristics of the high-dimensional input via its spatial receptive field [64].

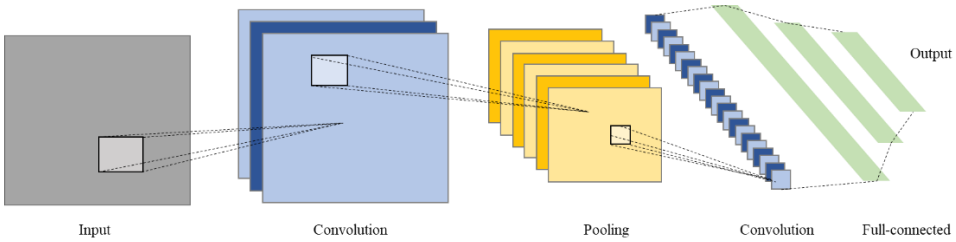


Figure 7. Architecture of simple convolutional neural network.

Compared with a fully connected neural network, the main advantage of the fully convolutional networks (FCNs) is that they can avoid the problem of expanding high-dimensional features into vectors, resulting in a loss of spatial information while significantly reducing the required training parameters through weight sharing and sparseness [65]. Therefore, since CNNs can process high-dimensional data [66–68], the data sources can be hyperspectral images, SAR images, street view images, or others.

The CNN's strong learning ability makes it more able to obtain accurate and rich features in the study of high-dimensional information than traditional methods, when the training samples are sufficient. CNNs and their improvements have made a significant contribution to remote sensing image applications, such as scene classification [69], object detection [70], and change detection [71]. The classical CNNs and their extension architectures include AlexNet [12], VGGNet [72], ResNet [73], UNet [74] and DenseNet [75], and HRNet [23], which can be used as feature extractors.

Knowledge transfer via the VGG network can transfer different domain datasets (e.g., CV-domain images to RS-domain images) [76]. By using CNNs, HR remote sensing image classification is realized, while the accuracy of high-resolution image segmentation is optimized by improving UNet [77]. This model can achieve better classification results for LiDAR using ResNet [78]. By allowing the CNN to derive high-dimensional features from the spectral information of hyperspectral imagery, the accuracy of remote sensing change detection is improved [79].

In general, most end-to-end CNN networks are only trained at the output layer to supervise learning. However, deep convolutional networks cannot effectively learn useful features because of the invisible hidden layer and the lack of supervision, meaning the network cannot efficiently learn the most useful features, especially in a deep network, which leads to a vanishing gradient problem that affects the subsequent prediction. Instead of only relying on the gradually traced gradients from the output layer, deep supervision can improve the learning ability of the feature extractor and thus help derive more effective information [17,80,81].

2.2.4. Generative Adversarial Networks (GANs)

GANs are a new family of generative models, which were introduced in 2014 [82]. As shown in Figure 8, the models consist of two separate models (a generator and a discriminator), which compete against each other in a game to achieve a dynamic balance. The generator's learning produces fake data similar to "real" samples, the purpose of which is to generate functions mapped with noise z to a target distribution, while the discriminator learns to distinguish between false and real data. A basic GAN works via minimax recursive, whereby the generator and the discriminator act as opponents, with the former trying to minimize the following functions, and the latter the other. Through continuous adversarial learning, the discrimination ability of the discriminator grows stronger, while the data generated by the generator will become more similar to the real data.

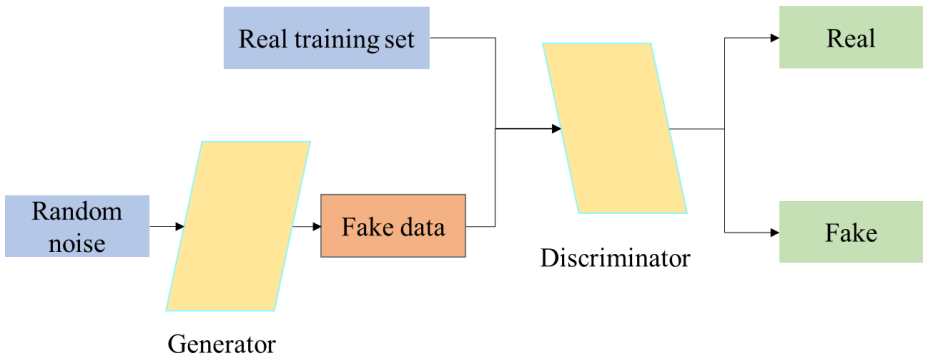


Figure 8. Architecture of GANs.

The minimax loss is given as follows:

$$min_G max_D f(D,G) = E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))], \quad (3)$$

where x is the given training data, $G(z)$ is a mapping function from a given random noise vector z, $D(x)$ represents the probability that x from the training data is real or fake, $D(G(z))$ is the discriminator's estimate of probability of the fake generated sample being real, G and

D are trained simultaneously, E_x is the expected value over all real data samples, and E_z is the expected value over all random inputs to the generator.

GANs can be trained with a small amount of labeled data (i.e., real data) and can offer an effective discrimination model to detect changes. With two-part confrontations, the problems of vanishing gradient and mode collapse become more obvious, which can make their application in remote sensing images not as effective as traditional methods [83]. Therefore, since the invention of GANs, researchers have made efforts to improve/modify them in various ways, such as DCGAN [84], InfoGAN [85], CycleGAN [86], WGAN [87], Self-Attention GAN [88], and BigGAN [89]. More works with GANs are listed in the GitHub resource [90].

As mentioned above, change detection is the process of capturing the distribution of the change or different information (DI) map via the multitemporal data's joint distribution, which compliments the GAN's function. Therefore, GANs can be utilized to explore the relationship between the desired latent distribution (change map or DI) and the multitemporal data's joint distribution.

In change detection, all generated images should show differences in the expected change area. Meanwhile, areas that have not changed should not show differences.

GANs have two main advantages in change detection. On the one hand, though data limitations are the main problem in deep learning, GANs can learn to generate vast pseudo-data, thus strengthening the generalization ability of the network. Specifically, GANs can extract useful and discriminative features from a large amount of unlabeled data with limited labeled samples, which reduces costs and ensures good performance in semi-supervised change detection [91]. On the other hand, inspired by domain adaptation and transfer learning technology, GANs can map an image in a source domain to a virtual image in a target domain.

When applying GANs in the field of object detection, image segmentation, and change detection, it was found that they are more suitable for change detection, mainly because similar scenarios reduce the risk of the GAN generating unwanted false results [83]. In change detection, the weather, season, and other factors can lead to pseudo-change, which can be excluded by GANs [92,93].

Compared to other generative models, for example, the Variational Autoencoder (VAE), GANs have no requirements as regards the reconstruction constraints, thus precluding a fuzzy change or DI map. Moreover, compared with CNNs that perform binary classification for each pixel, the GAN is insensitive to noise in pixels because it restores the full distribution of differential information.

2.2.5. Transformer-Based Networks

Transformers, first introduced in 2017, were originally used for sequence-to-sequence learning [15]. They can provide long-range dependency modeling with ease, and thus have been widely used for natural language processing (NLP) and have started to show promise in computer vision [94,95]. A transformer-based network can efficiently model the context information by leveraging the strengths of convolutions and transformers. They can also be seen as an effective attention-based tool to increase the reception field of the model, which improves the ability to represent change detection features [96]. The architecture of transformers is constructed following the encoder–decoder architecture, as shown in Figure 9. After obtaining token sets (feature sequences) for the input image, we then model the context between these tokens with a transformer encoder. With the context-rich tokens for the input image, a transformer decoder is adapted to project the representation of concepts back to pixel-space to obtain pixel-level features. Although the transformer structure has been widely used in NLP, its application in the field of vision (Vision Transformer, ViT) is still limited.

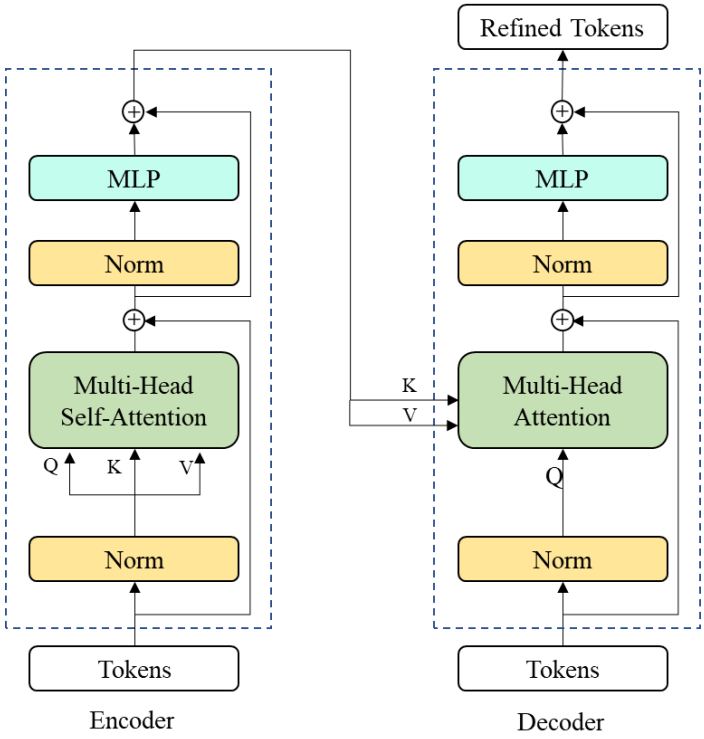


Figure 9. Architecture of transformers.

2.3. Change Information Extraction

In early change detection applications, the most traditional and classical methods are algebraic analysis methods (such as difference [97], regression [98], etc.). Following the basic idea of regarding changes as dissimilarities, the change map is generated by measuring pixel-level similarity between pairs of images. The advantage of this method is that the calculation requirement is low, while the disadvantage is that the handcrafted features cannot effectively distinguish changes. They identify significantly different pixels by the change features and then obtain a binary mask via thresholding. Change features measure the change probability at each detection unit (e.g., pixel, patch, or object). The change information extracted from multitemporal images lays the foundation for the subsequent generation of a DI map and the final change map. There is no doubt that a satisfactory change detection method depends closely on the development of appropriate distance metrics to measure the change information. In general, the change information function contains two parts: feature extraction and feature similarity.

At present, many researchers are focused on improving the robustness of change detection by ameliorating the strategies of feature extraction and feature similarity. The performance of change detection mainly depends on the recognition of change information.

2.3.1. Feature Extraction Strategy

The DC-CNN [99] was the first example of using DCNNs for change detection. Since then, DCNNs have been extensively applied. In these applications, there are two main types of feature extraction strategy: single-branch structures and dual-branch structures, as shown in Figure 10.

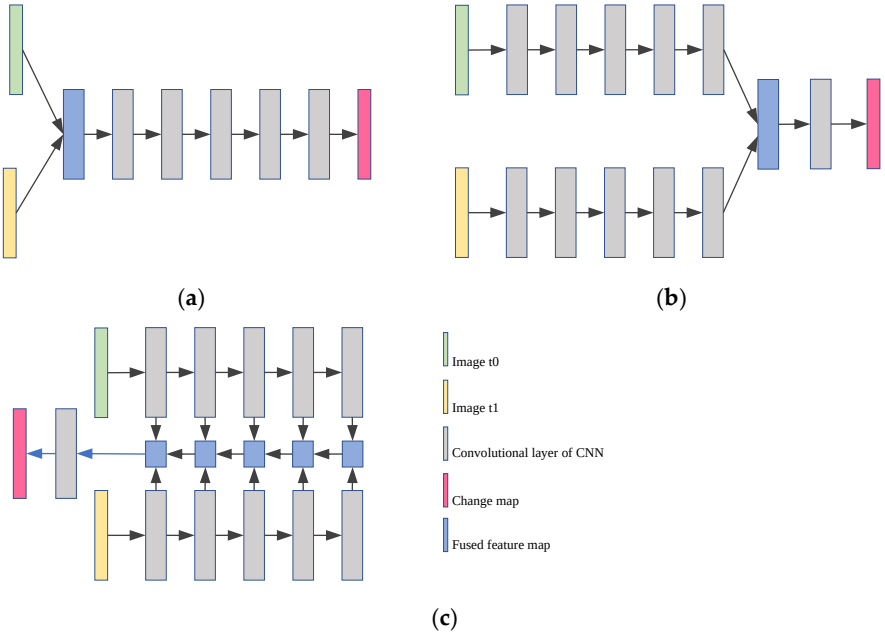


Figure 10. Feature extraction strategy. (**a**) Single-feature structure; (**b**) dual-feature structure (single-scale); (**c**) dual-feature structure (multiscale).

As an early fusion (EF) strategy, the single-branch structure fuses the two inputs before feeding them into the network to extract the features, adopting strategies such as difference or concatenation, as shown in Figure 10a. The concatenation operation treats the two inputs as different color channels. This feature extraction strategy is based on image-level feature fusion. To some extent, it is illogical to treat change detection issues as exclusively semantic segmentation problems [81].

Different from the single-branch structure, the dual-feature structure is a late fusion strategy, in which the resulting features are extracted from the fused results of two independent branches, as shown in Figure 10b,c. Essentially, it is a Siamese neural network. The Siamese network applies better to the change detection task with separate feature extraction. Firstly, the two branches extract features from both inputs separately, with the same structure and shared/no shared parameters. Then, the two branches are merged only after the convolutional layers of the network are completed. The dual-stream structure uses different network branches to fulfill differentiated feature extraction in order to obtain more targeted features. Much evidence shows that the late fusion strategy achieves a better performance than the previous image-level feature fusion approach.

Generally speaking, the two branches of the Siamese network share weights. That is, the network extracts feature from the two inputs using the same approach, as shown in Figure 11a. Sharing parameters may prevent each branch from reaching its respective optimum weight, and so other groups recommend using pseudo-Siamese as a substitute, as shown in Figure 11b. A Siamese network with two parallel encoding streams that share weights has fewer parameters and converges faster [81]. Chen [100] verified that the efficiency of the Siamese network is close to that of the pseudo-Siamese network but has lower computation costs.

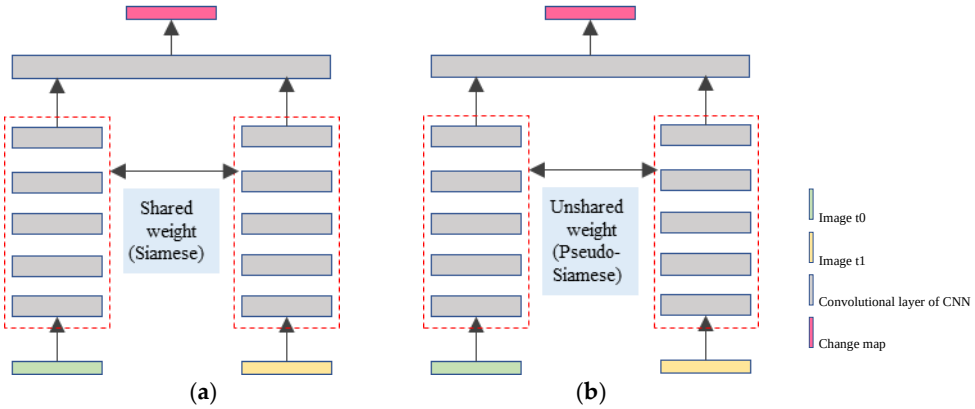


Figure 11. Dual-feature structure with different weights. (**a**) Shared weight (Siamese network); (**b**) unshared weight (pseudo-Siamese network).

2.3.2. Feature Similarity Strategy

The change detection regards changes via similarities, directly measuring them with distance metrics and aiming to learn discriminative features that group unchanged features and separate changed parts. Previously, change detection has been greatly limited by the performance of its hand-crafted features. Inspired by the success of the CNN model in image classification, we produced a highly abstract and discriminating feature map via learned filters. The most popular ways to measure the distance between the features are Euclidean distance [101–105] and cosine similarity [106,107]. The suitability of the distance metric will seriously affect the performance of the model, and this depends on the corresponding task. For instance, Euclidean distance and cosine similarity are suitable for face recognition and text processing, respectively. The experiment shows that Euclidean distance, also known as mean square error (MSE) distance, performs better than cosine similarity when measuring changes [108]. Other distances such as Manhattan distance [109], Minkowski distances [92], and Mahalanobis distance [92] are also used. To reuse a pre-trained deep CNN, a pre-trained weight was introduced to solve the change detection problem [103]. The simplest and easiest way to do this is to directly measure changes in the classic feature extraction network with a distance metric [103,110], such as AlexNet [12], or the VGG model [72], which is pre-trained on a large auxiliary dataset [111]. A thresholding step is performed to design the final change map, such as OSTU [112]. However, models without finetuned learning are not sufficiently discriminative to describe changes.

Deep metric learning methods have been successfully used for learning more discriminative features, such as learning a parameterized embedding space using deep Siamese fully convolutional networks [13]. To suppress the features of the changed regions and retain the unchanged ones as effectively as possible, various loss functions have been investigated for use in change detection, such as contrastive loss [104,113] and triplet loss [114]. Triplet loss obtains better results than contrastive loss because triplet loss takes advantage of the greater spatial relationships among pixels [115]. To avoid having to set parameters in the contrast loss function, the feature fusion [116] method is introduced to learn image similarity based on dual branches with shared weights, wherein the feature correlations between two images can be fully utilized. Using a decision network can eliminate the thresholding procedure in the change map. Compared with the representations developed using a learned distance metric, such as that with fully connected layers [79], the decision network can embrace more complex similarity functions beyond distance metrics. However, metric-based methods process multitemporal images separately, ignoring the correlations within the temporal information [115].

2.3.3. Feature Fusion Strategy

The Siamese network structure and feature fusion operation are combined to derive a difference map, and then a threshold algorithm is used to produce the change map.

Feature fusion has the ability to combine meaningful information from the two branches to generate a single fused feature, which is more beneficial for subsequent applications because it contains richer information. In other words, feature fusion is introduced to learn image similarity on the basis of dual branches, eliminating the need to apply an extra thresholding procedure to the change map. At present, most of the studies in this field of feature fusing mainly focus on two operations: feature aggregation and the attention mechanism.

Feature Aggregation. The single-branch structure directly fuses features at the image level by difference, summation, or concatenation. The feature fusion approach of the dual-branch structure can be roughly classified as single-scale fusion [59,104,117–120] or multiscale [32,33,121] fusion, as shown in Figure 10b,c, respectively.

Single-scale fusion is only used to fuse the top level of the two branch features. Multiscale fusion performs dense feature fusion in a deep-to-shadow manner, and the feature transformation method is the same at each scale, employing difference, summation, or concatenation, similarly to single-scale fusion.

Much previous work has revealed that the shallow features of deep networks contain more detailed information—for example, image texture and object boundary—but lack semantic cues. With the increase in layers, the hierarchical features will become more abstractive (due to sequences of pooling, etc.) [103]. Generally speaking, the deeper layers have more semantic information, but it is less detailed. The multiscale fusion strategy can bridge the gap in different-level feature maps by combing the hierarchical feature maps with a broader context and a finer spatial detail.

The multiscale feature fusion structure and the skip connect structure have been proven as feasible schemes to map shallow spatial information to deep semantic features. They can all be adapted to both single-branch [122] and dual-branch [123] feature extraction networks. Changed objects are usually irregular and of diverse scales in the remote sensing images. The multiscale feature fusion structure is an efficient technique for solving the problem and is employed to derive features at various scales. The basic operations of concatenation [2,32,120,122], summation [123], and difference [33,80,121,123,124] can be used to obtain global information, unchanged information, and changed information, respectively. Skip connections [80] bridge high-level features and low-level features to obtain a better result in image segmentation networks [74,125]. It should be noted that the results of mathematical transformation, such as difference, summation, stacking, concentration, or other logical operations, can be regarded as the end element of skip connections [31,123].

Attention Mechanism. Due to the semantic discrepancy, feature aggregation by simple logical operations often leads to feature confusion. In the past decade, the attention mechanism has played an increasingly important role and has benefited various computer vision applications, such as image classification, object extraction, and semantic segmentation. Through the attention mechanism, a network can emphasize important information related to images/features in channels or positions automatically, which significantly improves the efficiency and accuracy of the network. In change detection, the attention mechanism can enhance the representation of change features and suppress the features of irrelevant changes in bi-temporal images. The attention-based fusion strategy builds an information bridge between two features learned from the Siamese structure to realize information fusion. Therefore, the final feature is not obtained from two feature branches in isolation, but instead from their interaction [100].

In fact, the two feature fusion methods are usually used simultaneously and complement each other. Multiscale feature representations of the images could be extracted via the multiscale feature fusion structure. This is particularly effective for use with ground objects of various sizes. The attention mechanism makes the network pay more attention to changed areas and improves the efficiency of feature extraction.

It has always been difficult in the field of deep learning to understand how to obtain a larger receptive field without pooling, which gives rise to a partial loss of image information. Besides the two strategies, some methods that concentrate on global or local features

are available and are critical for extracting better change information [114,120], such as the pyramid pooling module (PPM) [126], atrous spatial pyramid pooling (ASPP) [127], dilated convolution [128], and inception architecture [129], which can capture multiscale information via different receptive fields.

2.4. Optimization Strategy

Currently, binary cross-entropy loss is most prevalent in binary classification based on DL. Since it calculates the similarity between two distributions, it is more suitable for binary change detection.

However, class imbalance [130] is a common issue when using the deep learning method for classification, where the number of some classes are far more numerous than others, which is similar to the case of remote sensing image change detection. Training with very few changed pixels causes the network to tend towards a particular class. The weighted cross-entropy loss is introduced as an improvement to balance the proportional relationship of the different class samples. Based on the ratio of the two unbalanced samples, a distribution weight [17] or weight matrix [33] can be constructed to control the cross-entropy loss. Besides this, focal loss and dice loss [17] are also proposed to resolve the imbalance problem. Generally, the two losses usually occur simultaneously to realize complementary advantages, because focal loss is pixel-wise loss without spatial relations, while dice loss is region-wise loss that includes spatial dependence. Xiang [120] proposed a separable loss function to optimize the network. The loss is calculated for the changed pixels and the unchanged pixels, respectively, to weaken the influence of the imbalance label. The leading theory of separable loss is that features of changed regions should be pushed away, and unchanged areas should be kept as close as possible. Inspired by this, the contrastive loss function [131] was employed to evaluate the similarity between two images by enlarging the measure between changed pairs and simultaneously reducing the measure between unchanged ones. Similarly to the improvement of binary cross entropy, Zhan optimized weight contrastive loss [104].

In addition, data-based oversampling approaches, such as data augmentation [132–134], offer another effective way to address the problem of imbalanced data and have been widely used in DL to improve the robustness of deep models.

Furthermore, some researchers used post-processing to obtain a better change map by updating the initial result via k-nearest neighbor [104], fuzzy c-means clustering [121], or multiscale segmentation [31,110].

Unsupervised learning. Currently, very few open annotated datasets are available for remote sensing change detection, which severely limits the practical applicability of DL models, especially FCNs. It is worth mentioning that random noise generated from a simple distribution can be transformed into new image–label pairs using generative adversarial networks (GANs). This promising unsupervised learning method has achieved great success in many applications. However, although they have an excellent ability to generate data, it is difficult for GANs to generate densely annotated images to train FCN models, and very few people pay attention to this research. Furthermore, Kullback–Leibler (KL) divergence loss [135] makes sense for use in unsupervised learning.

Weakly supervised learning. In contrast to supervised learning, which requires many annotated datasets, and unsupervised learning, which leads to poor model performance, weakly supervised learning (SSL) provides an alternative way to improve the robustness and generalization performance of the DL-based model. It can mine discriminable features using incomplete, inaccurate, or imprecise labels, instead of a large amount of accurately labeled information [136]. Semi-supervised learning is typically weakly supervised learning that learns from incomplete samples, consisting of labeled and unlabeled datasets, to train a model. Many researchers have adopted this approach in vision tasks [124,137,138], including change detection [29]. When it comes to imprecise supervision, the labels are coarse-grained, such as image-level labels [139–141]. The inaccuracy of the supervised approach is related to the fact that given samples do not always match the ground truth,

the process often referred to as learning with noisy labels, which is closer to the context of deep learning in the industry. The current research on noise learning mainly pays attention to classification problems [142] and rarely focuses on complex change detection problems.

3. The Analysis of Granularity for Change Detection

With the improved spatial resolution of remote sensing images in recent decades, many deep learning methods have been proposed for aerial and satellite image change detection. Depending on the granularity of the detection unit, we can roughly classify these methods into two main categories: scene-level methods (SLCDs) and region-level methods (RLCDs). These two categories are not necessarily independent of each other, and sometimes, the same change detection process may be present in different methods simultaneously. Figure 12 shows a taxonomy of methods used for change detection that employ HR remote sensing images.

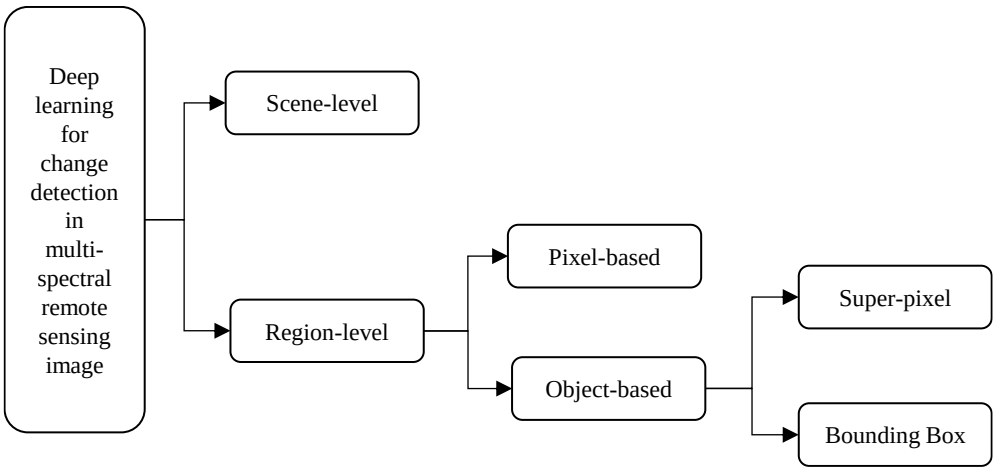


Figure 12. Taxonomy of methods for change detection in high-resolution remote sensing images.

3.1. Scene-Level Change Detection

With the increasing availability of high-resolution images covering the Earth's surface, the contextual and textural information of landscapes is becoming more abundant and detailed. It is now possible to achieve land use analysis at the scene level, such as scene classification [143–145], scene segmentation [138,146,147] and scene change detection [148–150].

Although single-image classification has been extensively explored due to its wide applicability, especially in natural images, few studies have focused on scene change detection in multitemporal images. For multitemporal images, most work focuses on change detection at the pixel and object level, or on the further identification of change types.

Nevertheless, pixel-level or object-level change detection methods are not appropriate for land use variation analysis. The main reason for this could be that the objects in the scene, such as vegetation growth and the demolition/construction of individual buildings, do not directly affect the land use category, i.e., their changes within the scene do not change the land-use category, for example, from a residential area to an industrial area. Therefore, it is crucial to improve change detection methods at the scene scale. Detecting scene changes with multitemporal images and identifying land-use transitions ("from–to") at the scene scale is a new area of interest for urban development analysis and monitoring [151]. For example, the appearance of residential and commercial areas can indicate the development of a city [150].

Scene-level change detection (SLCD) by remote sensing, i.e., scene change detection in remote sensing images, seeks to analyze and identify land use changes in any given multitemporal remote sensing image of the same area from a semantic point of view. Here, "scene" refers to an image cropped from a large-scale remote sensing image that includes unique land-cover information [145].

In the SLCD method, the features of the two input images are used to generate a DI map and then to classify the input patch into two classes (change and no change [152]) using a decision method, such as threshold segmentation or decision network. The decision method carries out change detection as a binary classification task with two outputs: change or no change. Thus, the two challenges of SLCD are finding an effective method to extract distinguishing features of the image and seeking an optimal transformation feature space to explore temporal correlations.

As with most computer vision tasks, extracting the discriminative features of a multitemporal image is an important and challenging step. Before deep learning was developed, significant efforts were made to derive discriminative visual features, including handcrafted local features, e.g., scale-invariant feature transformation (SIFT) [153] and encoding local features using bag-of-visual-words (BoVW) [154]. Considering the weakness of handcrafted features, several researchers turned to unsupervised learning techniques, e.g., sparse coding [155]. The features automatically produced from unlabeled data were successfully applied in scene classification and were then introduced to change detection. However, representative features of image scenes based on unsupervised learning have not been exploited adequately, limiting their ability to discriminate between different scene classes in remote sensing images. As regards the decision method for scene classification from the two features, the support vector machine (SVM) [156] is the most common and effective classifier [157].

With the collection of many annotated samples, the development in machine learning theory, and the enhancement of computation ability, deep learning models (e.g., autoencoders, CNNs, and GANs) have demonstrated their power in learning productive features [145]. This decisive advantage has spread to scene classification and change detection from remote sensing images [108].

The simplest method of SLCD, i.e., the post-classification method, treats the scene change detection task as an independent classification that ignores temporal correlation information and thus suffers from error accumulation. In other words, it barely considers the temporal correlation of the multitemporal images. Some researchers have begun to consider the temporal correlation between multitemporal image scenes, developing Deep Canonical Correlation Analysis (DCCA) Regularization [149] and an improved method called Soft DCCA [158]. However, these only focus on learning the correlated features from two inputs and cannot be optimized to improve feature representation capabilities. A learned fully connected layer can be used to model the similarity between bi-temporal scenes and improve the reliability of feature representation [159].

3.2. Region-Level Change Detection

In the change detection task, the pixels and objects in images are the two main categories of analysis. Region-level—including pixel-based and object-based—methods have been studied, wherein change detection can be regarded as a dense binary segmentation task [17]. Following the preparation of the images, a fully convolutional network (FCN) can classify the segmentation result into change or no change for each region (pixel or object). A general approach is to assign a change score to each region, whereby the changed region has a higher score than the unchanged ones. To some extent, this type of method allows for end-to-end change detection and avoids the accumulation of error. Moreover, it offers a tremendous benefit in its detection speed, which is helpful for large-scale data processing. Since most change detection applications involve identifying changes in specific regions or targets among multitemporal remote sensing images, region-level change detection methods are more popular than scene-level change detection methods.

3.2.1. Patch/Super-Pixel-Based Change Detection

Patches and super-pixels are the most common detection unit in remote sensing image processing applications. Patches or super-pixels are first constructed, then the DI map is generated by voting for use as a pseudo-training set, which can be used to learn the

change type of the center pixel [152,160,161]. A patch is a regular image grid cell, while a super-pixel is an irregular adjacent pixel cluster.

Patch/super-pixel-based change detection (PBCD) is only performed when the input pairs change globally. After concatenating the network channels' features in a multilevel Siamese network, the generated vector is applied to train a decision network with two layers. This network handles change detection as a binary classification task with two outputs: 1/0 for change and no change, respectively [105]. Each patch or super-pixel passes through a convolutional network to generate a fixed-dimensional representation. The features of the super-pixels should be transformed into 1-D features, due to their irregular shape, resulting in the loss of spatial information. Besides this, excessive interference information in the rectangular box also seriously influences the classification result.

To tackle the problem, the patch-based deep learning framework is used, which is an algorithm that trains pixels and their neighbors to form patches with a fixed size of 3×3, 5×5, 7×7, etc. The method pulls the patch size into a vector as a model input to predict the change category of the patch's center pixel based on its neighborhood's values, according to the principle of spatial proximity. For instance, the deep belief network (DBN) [160] and the multilayer perceptron (MLP) are, relatively speaking, the simplest methods that use 1-D neural network models in patch-based change detection. In these methods, the patch is flattened to a 1-D vector as input; then, the weights are initialized using greedy layer-wise pre-training and finetuned with labeled patch samples. However, to their detriment, the two architectures with fully connected layers suffer from a large number of learnable parameters, with only a limited number of annotated training examples for change detection, leading to overfitting and an increased computational cost. Furthermore, another drawback of the before-mentioned networks is that they squeeze spatial features into 1-D vectors, resulting in the 2-D spatial properties of the imagery being neglected.

Another factor that can affect the model's performance is the size of the patch, which can affect the size of the receptive field. It is usually challenging to find the appropriate size for the best performance. If the patch size is too small, the small receptive field with insufficient contextual information may limit the performance of change detection. The network cannot learn the change information and the enclosing fields thoroughly, thus failing to detect changes correctly. In addition, the method may reduce the computational efficiency and increase memory consumption due to the significant overlap between the neighboring fields.

Moreover, without losing spatial information [116], the patch-based attention mechanism can effectively overcome the uncertainty of predicted categories for PBCD methods. However, the images in the remote sensing application are significantly larger than natural images, but there are smaller objects for each class in remote sensing images. Almost every image contains various object categories, and it is not easy to classify the scene information at the global level. In other words, typical attention-based procedures are not appropriate for large-scale remote sensing images' semantic learning for the patches descriptors providing limited essential information of the local contexts.

Thus, the main limitations of PBCD methods can be listed as follows: first, it is difficult to find an appropriate patch size, which significantly affects DNN performance; second, redundant information in pixel patches leads to overfitting and increases the computational cost.

3.2.2. Pixel-Based Change Detection

Generally, pixel-based change detection methods extract features from individual pixels and those surrounding them and predict binary masks, classified pixel by pixel, as changed or unchanged. It is noteworthy that an encoder–decoder architecture is becoming increasingly popular in pixel-based change detection due to its high flexibility and superiority. As we mentioned earlier, spectral–spatial information is important for change detection. However, most algorithms compare the spectral or textual values of a

single pixel without considering the relationship between neighboring pixels and ignore the spatial environment.

FCNs [13] use a fully convolutional (FC) layer instead of the fully connected layer in CNNs to produce a pixel-wise prediction. Thereafter, FCNs and their variants provide an effective method for fine-grained change detection, such as FC-EF [162], FC-Siam-conc [162] and FC-Siam-diff [162], and W-Net [163]. The most used encoder–decoder CNN, SegNet, which was improved via VGG16, is often used for the semantic segmentation of images. However, when directly applied to change detection, it will achieve low accuracy without skip connections. Although a simple connection can help recover the loss of spatial information, it remains challenging to fulfill the needs of change detection tasks, especially for objects of various sizes. Therefore, UNet++ [17] employs a series of nested and dense skip connections to achieve multiscale feature extraction and reduce the pseudo-changes induced by scale variance. It is a promising avenue to exploit the potential of UNet++ for the pixel-level segmentation of remote sensing images, which has the advantages of capturing fine-grained details. To fully exploit the spatial–temporal dependence between multitemporal remote sensing images, BiDateNet [115] was proposed to better distinguish the spatial and temporal features. In BiDateNet, LSTM convolutional blocks are added to the skip connection to detect temporal patterns between bi-temporal remote sensing images using a U-Net architecture. In addition, some studies [114,164] employed ASPP [46] to extract multiscale features, which would improve change detection.

Moreover, the attention mechanism improves average or maximum pooling used in CNN models and enables the models to evaluate the influences of features at different locations and ranges. Attention mechanisms have been used in computer vision research for years, so it is no wonder that numerous publications apply this mechanism in change detection [165]. The convolutional block attention module (CBAM) [166] is used to make features from different phases more recognizable in the channel and spatial aspects [167]. Self-attention [52] is a mechanism that links different locations within a sequence to estimate the feature of each location in that sequence. It can model long-range correlations between bi-temporal remote sensing data. Non-local neural networks [35] have developed self-attention in various tasks, such as video classification and object recognition.

3.2.3. Object-Based Change Detection

Object-based methods take objects instead of pixels as the analysis unit. An object is a group of local pixel clusters, wherein all the pixels are assigned the same classification label. An object-based method effectively exploits the homogeneous information in images and eliminates the effects of image noise, boundaries [168], and misalignments. Because of the possible benefits of object-based methods, they are prevalent in land-cover mapping. In various publications, they have achieved better performances than pixel-based methods. This success has led to their general use in object-level investigations, e.g., object detection and instance separation. In recent years, object-based change detection techniques have also been developed for the detection of changed objects. Theoretically, this method can reduce the number of falsely detected changes that often appear in the predictions of pixel-based methods. This approach generates object-level predictions, e.g., the masks or bounding boxes of various changed objects. The methods fall broadly into two categories. The first performs super-pixel-based change detection and outputs masks. The second group of change detection methods is based on an object detection framework for finding a changed object in the form of a bounding box. The two categories can use the post-classification comparison method, which considers change detection as classifying pairs of images/boxes. In this task, land-cover classes comparison is carried out between two classified images/boxes, of which different classes are changed.

Super-pixel-based. This method works with homogenous pixel groups acquired by image segmentation, utilizing spectral texture and geometric features, e.g., pattern and area. Some of the "salt-and-pepper" noise in the change detection results is eliminated with the use of a super-pixel object. Sometimes, the super-pixels generated by multiresolution

segmentation are used to refine the results to the object level [31]. Nevertheless, whatever the super-pixel formation, inappropriate scale setting performed by hand will add extra errors. For instance, the cleanliness of the objects decreases as the segmentation scale expands. The computational effort and small observation field are the two main limiting factors in extreme segmentation (approximating the pixel-based methods). Therefore, the focus of object-level change detection is to break through the constraints of prior parameters and collect adaptive objects. Not every object produced in this way is the same size; consequently, over-segmentation and under-segmentation lead to worse change detection results [34].

Bounding box candidates. In this method, change objects are taken as targets for object detection (OD). The usual OD methods, such as SSD [169], Faster R-CNN [170], and YOLO1-5 [171–175], have the potential for use in change detection.

This approach considers the "changed area" in remote sensing images as the detection target, while the "unchanged area" is the background. The OD methods are applied in high-resolution remote sensing image change detection [176]. The detection results in a group of square areas and then intersecting areas with a specific change type are mixed. The feature extraction network can be a single-branch or dual-branch network. For a single-branch network, the multitemporal images are merged or subtracted first, and the result is then fed into the OD network to determine the change [176]. The dual-branch network generates the basic and representative features of each image, respectively, and then fuses the features [177] or proposal regions [178] of each branch to predict the class scores and the confidence of difference. In addition, object-based instance segmentation, such as that using Mask R-CNN, can be used as a basis for detecting changes, which produces the initialized object instance [179]. In fact, acquiring the object's location is the first step in determining the location of a changed object.

4. Related Datasets and Evaluation Metrics

Numerous methods have been proposed for detecting changes in recent decades, but choosing the right one is difficult. Therefore, it is essential to adopt appropriate metrics for use within the same data. Here, we list some publicly available benchmarking datasets and give some standard evaluation metrics used to compare the performances of the algorithms for application in further research.

4.1. Popular Datasets for Change Detection

In the following, we summarize some popular and public high-resolution optical remote sensing datasets for change detection. The aim is to arouse the reader's interest in change detection using high-resolution optical remote sensing images and facilitate preliminary experiments on change detection. We group the datasets into the following three categories by considering the necessary change detection tasks.

4.1.1. Binary Change Detection Dataset

DSIFN-CD [80] consists of 3940 pairs bi-temporal images, of which 394 pairs were cropped from manually labeled images, and 3940 pairs were generated by data augmentation. The dataset covers the changes in different land-cover classes, obtained from six pairs of high-resolution (2 m) satellite images, and each pair covers one city in China.

SZTAKI AirChange Benchmark Set (SZTAKI) [180] has a spatial resolution of 1.5 m, with 13 aerial image pairs of 952 × 640 pixels. The dataset is made up of three regions (e.g., Szada, Tiszadob, and Archive), containing 7, 5, and 1 image pairs, respectively. In early CD studies, this was the most commonly used dataset with various change types, including new built-up regions, building operations, planting forests, fresh ploughlands, and groundwork before building completion.

The Onera Satellite Change Detection (OSCD) [152] contains a collection of 24 pairs of satellite images acquired between 2015 and 2018, covering the world's urban areas (including Asia, Brazil, Europe, the Middle East, and the USA) with 10 m resolution. Each

of these is approximately 600 × 600 pixels. This dataset is mostly focused on urban changes caused by human activities, such as urban growth and urban decline.

The Aerial Imagery Change Detection (AICD) [181] is a synthetic dataset for binary CD with 500 pairs of aerial images, the pairs of which were generated from 100 original images with five different viewpoints. Moreover, some kind of man-made change object (e.g., buildings, trees, or relief) was added to each image to generate the image pair.

Season-varying CDD [182] consists of 16,000 pairs of Google Earth (GE) images, each with 256 × 256 pixels and a spatial resolution of 0.03–1 m/pixel, collected from seven pairs of 4725 × 2700 seasonal variation remote sensing images. Compared to the dataset AICD with synthetic images, CDD contains much change information that reflects real changes. In other words, this dataset pays more attention to changes related to the appearance and disappearance of objects, rather than to difference due to natural factors, such as seasonal differences, brightness, and so on.

SYSU-CD dataset [167] is a large compilation of existing change detection datasets with varying spatial resolutions, change types and data volumes. The dataset contains 20,000 pairs of orthographic aerial images with a size of 256 × 256 pixels and a spatial resolution of 0.5 m, taken between 2007 and 2014. In addition, it contains dense high-rise buildings.

Satellite-UAV heterogeneous images Change Detection (HTCD) [183] is a combined dataset shot by satellite in 2008 and UAV in 2020. The original satellite images are from Google Earth with a spatial resolution of 0.5971 m, and UAV images with a spatial resolution of 7.465 cm are from Open Aerial Map. Their size is 11 K × 15 K pixels and 1.38 M × 1.04 M pixels (divided into 15 blocks), respectively. The dataset mainly focuses on the changes in urban man-made objects (such as buildings, roads, and other artificial facilities) covering Chisinau and its surrounding area, and the area is approximately 36 square kilometers.

4.1.2. Semantic Change Detection Dataset

High-Resolution Semantic Change Detection (HRSCD) [184] consists of 291 pairs of 10,000 × 10,000-pixel aerial images with a resolution of 0.5 m per pixel, acquired in 2005/2006 and 2012. The dataset provides both binary change maps and land-cover maps (e.g., artificial surface, agricultural area, forest, wetland, and water).

Semantic Change Detection (SECOND) [185] is a dual task-based semantic change detection dataset, including one change class and six land-cover classes (e.g., non-vegetated ground surface, tree, low vegetation, water, buildings, and playgrounds). It contains 4662 pairs of aerial images with a size of 512 × 512 pixels.

Ultra-high-resolution UCD (Hi-UCD) [186] focuses on urban change by annotating nine land-cover classes in bi-temporal aerial images. The dataset contains 359 image pairs from 2017 to 2018, 386 pairs from 2018 to 2019, and 548 pairs from 2017 to 2019. Each image pair in the dataset, including an image, a semantic map, and a change map, has a spatial resolution of 0.1 m and a size of 1024 × 1024 pixels.

4.1.3. Buildings Change Detection Dataset

WHU-CD [187] is a public building change detection dataset. It contains one pair of high-resolution (0.075 m) aerial images with a size of 32,507 × 15,354 pixels, which covers the area of Christchurch, New Zealand, for 2012 and 2016.

Learning, Vision and Remote Sensing Laboratory (LEVIR-CD) [115] is a public, large-scale building change detection dataset. It contains 637 pairs of high-resolution (0.5 m) Google Earth (GE) images of 1024 × 1024 pixels, covering 20 different regions from 2002 to 2018.

AIST Building Change Detection (ABCD) dataset [188] consists of 22,171 pairs of aerial images, of which 10,777 are fixed-scale (160 × 160 pixels) and 11,394 are resized (120 × 120 pixels). The images are cropped from several pairs of pre-tsunami and post-tsunami images, and cover 66 km^2 of tsunami-affected area. The pre-tsunami images were acquired in August 2000 at a resolution of 0.4 m per pixel, while the post-tsunami images

were acquired on 11 March 2011 at a resolution of 0.12 m per pixel and resampled to 0.4 m per pixel.

S2Looking dataset [189] comprises 5000 image patch pairs extracted from side-looking rural-area satellite images. The bi-temporal images span 1–3 years, with a resolution of 0.5–0.8 m/pixel.

xBD [190] is the first and largest building damage assessment dataset to date, containing 22,068 satellite images and 850,736 building polygons, covering 45,000 square kilometers before and after the disasters (e.g., earthquakes, floods, volcanic eruption, wildfire, and wind). Each image is 1024 × 1024 pixels in size, with a resolution of 0.3 m/pixel.

Table 1 shows the basic information of the remote sensing image dataset for change detection. There are very few open datasets for change detection, and some of them have a low spatial resolution. That is to say, the public dataset that can be used for DL-based change detection has significant limitations. This detrimentally affects DL-based change detection methods in the following two ways. First, model overfitting can easily occur when the amount of data exceeds the number of model parameters and there are insufficient annotated datasets to learn from, severely limiting the practical applicability of the DL models. Second, the imagery of some available datasets has a low spatial resolution, such as SZTAKI, DSIFN, and OSCD, thus not only excluding many small change objects, but also making the boundaries of many man-made objects, such as buildings and roads, unclear. As a result, the unclear delineation leads to ambiguities in the models.

Table 1. Summary of popular high-resolution remote sensing datasets for change detection tasks.

Datasets	Classes	Sensor	Image Pairs	Image Size	Resolution (m)	Real	Period (Years)
SZTAKI	2	aerial	13	952 × 640	1.5	√	-
CDD	2	satellite	16,000	256 × 256	0.03–1	√	-
OSCD	2	satellite	24	600 × 600	10	√	2
ABCD	2	aerial	8506/8444	160 × 160/ 120 × 120	0.4	√	11
AICD	2	aerial	500	800 × 600	0.5	×	-
LEVIR-CD	2	satellite	637	1024 × 1024	0.5	√	6
WHU-CD	2	aerial	1	32,507 × 15,354	0.075	√	4
SYSU-CD	2	aerial	20,000	256 × 256	0.5		7
HRSCD	1 + 5	aerial	291	10,000 × 10,000	0.5	√	6, 7
SECOND	1 + 6	aerial	4662	512 × 512	-	√	-
Hi-UCD	1 + 9	aerial	1293	1024 × 1024	0.1	√	1, 2
DSIFN	2	satellite	3940	512 × 512	2	√	5, 8, 10, 15, 17
S2Looking	2	satellite	5000	1024 × 1024	0.5–0.8	√	1–3
HTCD	2	satellite, aerial	1	11 K × 15 K	0.59710.007465	√	12

The total number of literature citations for each dataset is shown in Figure 13, reflecting the importance of each dataset. The statistical analysis is based on the work of Google Scholar. The figure illustrates that WHU-CD is the most popular dataset, and the application of LEVIR-CD has increased sharply.

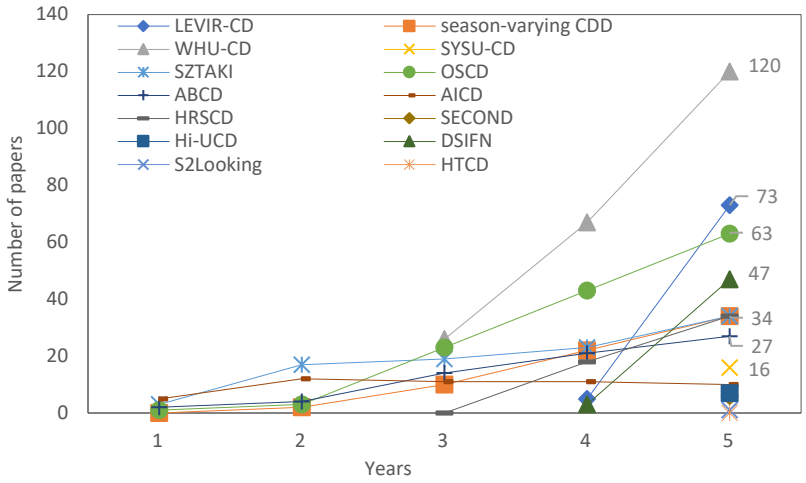


Figure 13. The total number of literature citations for each dataset.

4.2. Evaluation Metrics

It is well known that the evaluation of the performance of a change detection algorithm is a critical issue. Here, we introduced some commonly used evaluation metrics.

The confusion matrix [191,192] is commonly used for the quantitative analysis of binary classification accuracy, considering sensitivity and specificity. It is also suitable for binary change detection, via possible changed and unchanged statuses. The definition of the confusion matrix is shown in Table 2. *FP* (false positive) and *FN* (false negative) refer to the pixels that were incorrectly classified as changed and unchanged, respectively. *TP* (true positive) and *TN* (true negative) represent the changed pixels and unchanged pixels that were correctly detected, respectively.

Table 2. The confusion matrix of binary change detection.

Detected	**Reference**	
	Changed	**Not Changed**
Changed	*TP*	*FP*
Not Changed	*FN*	*TN*

The most widely used performance measures are defined as follows:

$$OA = \frac{TP + TN}{TP + FP + TN + FN}, \tag{4}$$

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{6}$$

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall}, \tag{7}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{8}$$

Overall accuracy (*OA*) is the general evaluation metric for prediction results. Precision measures the fraction of detections that were actually changed, recall measures the fraction of correctly detected changes, and F1 refers to recall and precision together. In general, higher precision indicates fewer false prediction results, and higher recall indicates that fewer changes are missed. *IoU* or the Jaccard Index is defined as the rate between the

intersection and unification of the predicted segmentation map and the ground truth. As regards the metrics, the larger their values are, the better the prediction results will be. Generally, in the binary CD task, all the metrics except *OA* are only calculated for the changed class because of the balance problem.

Besides these, there are some metrics with special applications, such as Separated Kappa (*SeK*), average precision (*AP*), and the trajectory error matrix (*TEM*).

SeK [120] is proposed to alleviate the influence of an imbalance in different change categories.

AP calculates the average value under the precision–recall curve, which is commonly used as an indicator of precision and recall at the object level.

$$AP = \int_0^1 p(r)dr, \tag{9}$$

Here, p and r refer to object-level precision and recall, respectively.

TEM enables the evaluation of results of multitemporal change detection [193]. The *TEM* consists of two indicators (e.g., binary classification and change detection), while the classification index is used to assess whether the detected change/non-change is correct. The other indicator shows the changed paths detected using the reference data. Some studies classify the possible change path into six subgroups [4,194], as listed in Table 3.

Table 3. Sub-groups in the trajectory error matrix (TEM).

Groups	Classification Result	Reference	Detected
S_1	Correct	non-changed	non-changed
S_2		changed	changed
S_3	Incorrect	non-changed	non-changed
S_4		non-changed	changed
S_5		changed	non-changed
S_6		changed	changed with incorrect trajectory

5. Conclusions

In this paper, we present a comprehensive review of change detection based on deep learning using high-spatial-resolution images, which covers the most popular feature extraction deep neural networks and the construction mechanisms. In addition, the granularity of change detection algorithms according to the detection unit is analyzed, which allows us to apply a suitable method independent of their particular applications.

The related change detection methods mentioned in this paper show that deep learning techniques have successfully contributed to the development of change detection and have made significant progress. However, there are still many challenges in change detection due to the diversity of requirements and complexity of data. In addition, the multisource data fusion and multiscale problems of remote sensing data are also issues that need to be paid attention to in remote sensing applications. It is worth mentioning that it is possible to meet different kinds of difficulties for different ground targets. For example, the displacement of high-rise buildings is one of the biggest challenges in images captured from different viewpoints [165], and heterogeneous appearance [195] is also an outstanding issue. Therefore, several suggestions for future research directions in change detection using remote sensing data are strongly recommended to focus more on these challenges. First, producing a great number of labeled samples for change detection is important when training a large model with sufficiently high generalization power to deal with varying complex scenes. Second, in order to handle the difficulties of change detection, domain knowledge (the spatial–temporal–spectral characteristics of remote sensing images, geographic information, and other geoscience-related knowledge) must be integrated into

the learning framework to enhance the reliability of the method. Third, learning with small sample sets is useful for the development of algorithms when the data on a large number of labeled samples are lacking and challenging.

Author Contributions: Conceptualization, Y.Z., H.J. and X.H.; writing—original draft preparation, H.J., M.P., H.X., Z.H., J.L. and X.M.; writing—review and editing, H.J., Y.Z., H.X. and X.H.; supervision, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Chinese National Natural Science Foundation Projects (Grant Nos. 92038301 and 41771363), and was supported by the fundings of Guangdong Surveying and Mapping Institute of Lands and Resource Department, Shenyang Geotechnical Investigation & Surveying Research Institute Co., Ltd.

Acknowledgments: The authors sincerely appreciate that academic editors and reviewers give their helpful comments and constructive suggestions.

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

References

1. Singh, A. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]
2. Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A.; Zhang, L. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. *Remote Sens. Environ.* **2021**, *265*, 112636. [CrossRef]
3. Moya, L.; Muhari, A.; Adriano, B.; Koshimura, S.; Mas, E.; Marval-Perez, L.R.; Yokoya, N. Detecting urban changes using phase correlation and ℓ1-based sparse model for early disaster response: A case study of the 2018 Sulawesi Indonesia earthquake-tsunami. *Remote Sens. Environ.* **2020**, *242*, 111743–111756. [CrossRef]
4. Liu, R.; Kuffer, M.; Persello, C. The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach. *Remote Sens.* **2019**, *11*, 2844. [CrossRef]
5. Bruzzone, L.; Serpico, S.B. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 858–867. [CrossRef]
6. De Bem, P.P.; De Carvalho Junior, O.A.; Fontes Guimarães, R.; Trancoso Gomes, R.A. Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 901. [CrossRef]
7. Zhang, Z.; Vosselman, G.; Gerke, M.; Tuia, D.; Yang, M.Y. Change Detection between Multimodal Remote Sensing Data Using Siamese CNN. *arXiv* **2018**, arXiv:1807.09562.
8. Chen, J.; Liu, H.; Hou, J.; Yang, M.; Deng, M. Improving Building Change Detection in VHR Remote Sensing Imagery by Combining Coarse Location and Co-Segmentation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 213. [CrossRef]
9. Qin, R.; Tian, J.; Reinartz, P. 3D change detection-Approaches and applications. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 41–56. [CrossRef]
10. Ban, Y.; Yousif, O. Change Detection Techniques: A Review. In *Multitemporal Remote Sensing. Remote Sensing and Digital Image Processing*; Ban, Y., Yousif, O., Eds.; Remote Sensing and Digital Image Processing; Springer: Cham, Switzerland, 2016; Volume 20, pp. 19–43.
11. Liu, T.; Yang, L.; Lunga, D. Change detection using deep learning approach with object-based image analysis. *Remote Sens. Environ.* **2021**, *256*, 112308. [CrossRef]
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [CrossRef]
13. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the Conference On Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 3431–3440.
14. Chen, L.-C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to Scale: Scale-Aware Semantic Image Segmentation. In Proceedings of the Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3640–3649.
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
16. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851.
17. Peng, D.; Zhang, M.; Wanbing, G. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [CrossRef]
18. 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.* **2018**, *56*, 937–949. [CrossRef]
19. Li, Y.; Zhang, H.; Xue, X.; Jiang, Y.; Shen, Q. Deep learning for remote sensing image classification: A survey. *Wiley Interdiscip Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1264. [CrossRef]

20. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
21. Zhang, Z.; Jiang, R.; Mei, S.; Zhang, S.; Zhang, Y. Rotation-Invariant Feature Learning for Object Detection in VHR Optical Remote Sensing Images by Double-Net. *IEEE Access* **2020**, *8*, 20818–20827. [CrossRef]
22. Newell, A.; Yang, K.; Deng, J. Stacked Hourglass Networks for Human Pose Estimation. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 483–499.
23. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. In Proceedings of the Conference on Computer Vision And Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
24. Zhu, Q.; Sun, X.; Zhong, Y.; Zhang, L. High-Resolution Remote Sensing Image Scene Understanding: A Review. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium(IGARSS), Yokohama, Japan, 28 July–2 August 2019; pp. 3061–3064.
25. Michael, K.; Arnt-Børre, S.; Robert, J. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 680–688.
26. Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 125–138. [CrossRef]
27. Gao, F.; Dong, J.; Li, B.; Xu, Q. Automatic Change Detection in Synthetic Aperture Radar Images Based on PCANet. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1792–1796. [CrossRef]
28. Yao, S.; Shahzad, M.; Zhu, X.X. Building height estimation in single SAR image using OSM building footprints. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, United Arab Emirates, 6–8 March 2017; pp. 1–4.
29. Peng, D.; Bruzzone, L.; Zhang, Y.; Guan, H.; Ding, H.; Huang, X. SemiCDNet: A Semisupervised Convolutional Neural Network for Change Detection in High Resolution Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5891–5906. [CrossRef]
30. Jacobsen, K. Characteristics of very high resolution optical satellites for Topographic mapping. In *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus GmbH: Göttingen, Germany, 2011; Volume XXXVIII-4/W19, pp. 137–142. [CrossRef]
31. 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]
32. Bao, T.; Fu, C.; Fang, T.; Huo, H. PPCNET: A Combined Patch-Level and Pixel-Level End-to-End Deep Network for High-Resolution Remote Sensing Image Change Detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1797–1801. [CrossRef]
33. Zhang, M.; Shi, W. A Feature Difference Convolutional Neural Network-Based Change Detection Method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232–7246. [CrossRef]
34. Hussain, M.; Chen, D.; Cheng, A.; Wei, H.; Stanley, D. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* **2013**, *80*, 91–106. [CrossRef]
35. Bovolo, F.; Bruzzone, L. The Time Variable in Data Fusion: A Change Detection Perspective. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 8–26. [CrossRef]
36. Tewkesbury, A.P.; Comber, A.J.; Tate, N.J.; Lamb, A.; Fisher, P.F. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens. Environ.* **2015**, *160*, 1–14. [CrossRef]
37. 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]
38. Shi, W.; Zhang, M.; Zhang, R.; Chen, S.; Zhan, Z. Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges. *Remote Sens.* **2020**, *12*, 1688. [CrossRef]
39. Khelifi, L.; Mignotte, M. Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. *IEEE Access* **2020**, *8*, 126385–126400. [CrossRef]
40. Lan, G.; Yoshua, B.; Aaron, C. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
41. Fischer, A.; Igel, C. An Introduction to Restricted Boltzmann Machines. In Proceedings of the Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications(CIARP), Buenos Aires, Argentina, 3–6 September 2012; pp. 14–36.
42. Liu, M.-Y.; Breuel, T.; Kautz, J. Unsupervised Image-to-Image Translation Networks. In Proceedings of the International Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
43. Baziotis, C.; Androutsopoulos, I.; Konstas, I.; Potamianos, A. SEQˆ3: Differentiable Sequence-to-Sequence-to-Sequence Autoencoder for Unsupervised Abstractive Sentence Compression. In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies(NAACL), Minneapolis, MN, USA, 6–7 June 2019.
44. Zabalza, J.; Ren, J.; Zheng, J.; Zhao, H.; Qing, C.; Yang, Z.; Du, P.; Marshall, S. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing* **2016**, *185*, 1–10. [CrossRef]
45. Nurmaini, S.; Darmawahyuni, A.; Sakti Mukti, A.N.; Rachmatullah, M.N.; Firdaus, F.; Tutuko, B. Deep Learning-Based Stacked Denoising and Autoencoder for ECG Heartbeat Classification. *Electronics* **2020**, *9*, 135. [CrossRef]
46. Liu, G.; Li, L.; Jiao, L.; Dong, Y.; Li, X. Stacked Fisher autoencoder for SAR change detection. *Pattern Recogn.* **2019**, *96*, 106971. [CrossRef]
47. Gong, M.; Yang, H.; Zhang, P. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *129*, 212–225. [CrossRef]

48. Ye, X.; Wang, L.; Xing, H.; Huang, L. Denoising hybrid noises in image with stacked autoencoder. In Proceedings of the 2015 IEEE International Conference on Information and Automation(ICIA), Lijiang, China, 8–10 August 2015; pp. 2720–2724.
49. Shao, Z.; Deng, J.; Wang, L.; Fan, Y.; Sumari, N.S.; Cheng, Q. Fuzzy autoencode based cloud detection for remote sensing imagery. *Remote Sens.* **2017**, *9*, 311. [CrossRef]
50. Iyer, V.; Aved, A.; Howlett, T.B.; Carlo, J.T.; Abayowa, B. Autoencoder versus pre-trained CNN networks: Deep-features applied to accelerate computationally expensive object detection in real-time video streams. In Proceedings of the Target and Background Signatures IV, Berlin, Germany, 10–11 September 2018; p. 107940Y.
51. Amberkar, A.; Awasarmol, P.; Deshmukh, G.; Dave, P. Speech Recognition using Recurrent Neural Networks. In Proceedings of the International Conference on Current Trends towards Converging Technologies(ICCTCT), Coimbatore, India, 1–3 March 2018; pp. 1–4.
52. Liu, P.; Qiu, X.; Huang, X. Recurrent neural network for text classification with multi-task learning. In Proceedings of the Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence(IJCAI), New York, NY, USA, 9–15 July 2016; pp. 2873–2879.
53. Zhong, Y.; Li, H.; Dai, Y. Open-World Stereo Video Matching with Deep RNN. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
54. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [CrossRef]
55. Shi, W.; Zhang, M.; Ke, H.; Fang, X.; Zhan, Z.; Chen, S. Landslide recognition by deep convolutional neural network and change detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4654–4672. [CrossRef]
56. Mou, L.; Bruzzone, L.; Zhu, X. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 924–935. [CrossRef]
57. Liu, R.; Cheng, Z.; Zhang, L.; Li, J. Remote Sensing Image Change Detection Based on Information Transmission and Attention Mechanism. *IEEE Access* **2019**, *7*, 156349–156359. [CrossRef]
58. Lyu, H.; Lu, H.; Mou, L. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection. *Remote Sens.* **2016**, *8*, 506. [CrossRef]
59. 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.* **2020**, *58*, 2848–2864. [CrossRef]
60. Ordóñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]
61. Song, A.; Choi, J.; Han, Y.; Kim, Y. Change Detection in Hyperspectral Images Using Recurrent 3D Fully Convolutional Networks. *Remote Sens.* **2018**, *10*, 1827. [CrossRef]
62. Lyu, H.; Lu, H. Learning a transferable change detection method by recurrent neural network. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5157–5160.
63. Sarigul, M.; Ozyildirim, B.M.; Avci, M. Differential convolutional neural network. *Neural Netw.* **2019**, *116*, 279–287. [CrossRef] [PubMed]
64. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 2021, *Early Access*. [CrossRef] [PubMed]
65. Han, Y.; Tang, B.P.; Deng, L. An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes. *Comput. Ind.* **2019**, *107*, 50–58. [CrossRef]
66. Lee, H.; Kwon, H. Contextual deep cnn based hyperspectral classification. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 3322–3325.
67. Mazzini, D.; Buzzelli, M.; Pauy, D.P.; Schettini, R. A CNN Architecture for Efficient Semantic Segmentation of Street Scenes. In Proceedings of the International Conference on Consumer Electronics(ICCE), Berlin, Germany, 2–5 September 2018.
68. Sharifzadeh, F.; Akbarizadeh, G.; Kavian, Y.S. Ship Classification in SAR Images Using a New Hybrid CNN-MLP Classifier. *J. Indian Soc. Remote Sens.* **2019**, *47*, 551–562. [CrossRef]
69. Pires De Lima, R.; Marfurt, K. Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sens.* **2020**, *12*, 86. [CrossRef]
70. Lei, J.; Luo, X.; Fang, L.; Wang, M.; Gu, Y. Region-Enhanced Convolutional Neural Network for Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5693–5702. [CrossRef]
71. Cao, C.; Dragićević, S.; Li, S. Land-Use Change Detection with Convolutional Neural Network Methods. *Environments* **2019**, *6*, 25. [CrossRef]
72. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the International Conference on Learning Representations(ICLR), San Diego, CA, USA, 7–9 May 2014.
73. He, K.; Zhang, J.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
74. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.

75. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Densely Connected Convolutional Networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
76. Liu, X.; Chi, M.; Zhang, Y.; Qin, Y. Classifying high resolution remote sensing images by fine-tuned VGG deep networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium(IGARSS), Valencia, Spain, 22–27 July 2018; pp. 7137–7140.
77. Guo, Y.; Liao, J.; Shen, G. A Deep Learning Model With Capsules Embedded for High-Resolution Image Classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2020**, *14*, 214–223. [CrossRef]
78. Wang, A.; Wang, M.; Wu, H.; Jiang, K.; Iwahori, Y. A novel LiDAR data classification algorithm combined capsnet with resnet. *Sensors* **2020**, *20*, 1151. [CrossRef]
79. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 3–13. [CrossRef]
80. 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]
81. Li, K.; Li, Z.; Fang, S. Siamese NestedUNet Networks for Change Detection of High Resolution Satellite Image. In Proceedings of the International Conference on Control, Robotics and Intelligent System(CCRIS), Xiamen, China, 27 October 2020; pp. 42–48.
82. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2014**, *63*, 139–144. [CrossRef]
83. Li, X.; Luo, M.; Ji, S.; Zhang, L.; Lu, M. Evaluating generative adversarial networks based image-level domain transfer for multi-source remote sensing image segmentation and object detection. *Int. J. Remote Sens.* **2020**, *41*, 7343–7367. [CrossRef]
84. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
85. Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; Abbeel, P. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. *arXiv* **2016**, arXiv:1606.03657.
86. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251.
87. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
88. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. *arXiv* **2018**, arXiv:1805.08318.
89. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In Proceedings of the 7th International Conference on Learning Representations(ICLR), New Orleans, LA, USA, 6–9 May 2018.
90. GAN_Zoo. Available online: https://github.com/hindupuravinash/the-gan-zoo (accessed on 20 March 2022).
91. Jiang, F.; Gong, M.; Zhan, T.; Fan, X. A semisupervised GAN-based multiple change detection framework in multi-spectral images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1223–1227. [CrossRef]
92. Zhao, W.; Mou, L.; Chen, J.; Bo, Y.; Emery, W.J. Incorporating metric learning and adversarial network for seasonal invariant change detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2720–2731. [CrossRef]
93. 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]
94. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the European Conference on Computer Vision(ECCV), Online, 23–28 August 2020; pp. 213–229.
95. Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.08005.
96. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *arXiv* **2021**, arXiv:2103.00208. [CrossRef]
97. Ke, L.; Lin, Y.; Zeng, Z.; Zhang, L.; Meng, L. Adaptive Change Detection With Significance Test. *IEEE Access* **2018**, *6*, 27442–27450. [CrossRef]
98. Ridd, M.K.; Liu, J. A Comparison of Four Algorithms for Change Detection in an Urban Environment. *Remote Sens. Environ.* **1998**, *63*, 95–100. [CrossRef]
99. Liu, T.; Li, Y.; Cao, Y.; Shen, Q. Change detection in multitemporal synthetic aperture radar images using dual-channel convolutional neural network. *J. Appl. Remote Sens.* **2017**, *11*, 042615. [CrossRef]
100. Chen, P.; Guo, L.; Zhang, X.; Qin, K.; Ma, W.; Jiao, L. Attention-Guided Siamese Fusion Network for Change Detection of Remote Sensing Images. *Remote Sens.* **2021**, *13*, 4597. [CrossRef]
101. Adam, W.H.; Konstantinos, G.D.; Iasonas, K. Segmentation-Aware Convolutional Networks Using Local Attention Masks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5048–5057.
102. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A Discriminative Feature Learning Approach for Deep Face Recognition. In Proceedings of the European Conference on Computer Vision(ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 499–515.
103. Larabi, M.; Liu, Q.; Wang, Y. Convolutional neural network features based change detection in satellite images. In Proceedings of the First International Workshop on Pattern Recognition(IWPR), Tokyo, Japan, 11–13 July 2016.
104. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change Detection Based on Deep Siamese Convolutional Network for Optical Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1845–1849. [CrossRef]

105. Faiz, R.; Bhavan, V.; Jared, V.C.; John, K.; Andreas, S. Siamese Network with Multi-Level Features for Patch-Based Change Detection in Satellite Imagery. In Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, CA, USA, 26–29 November 2018; pp. 958–962.
106. Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. In Proceedings of the 27th International Conference on Neural Information Processing Systems(NIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 1988–1996.
107. Mueller, J.; Thyagarajan, A. Siamese recurrent architectures for learning sentence similarity. In Proceedings of the Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 5 March 2016; pp. 2786–2792.
108. Guo, E.; Fu, X.; Zhu, J.; Deng, M.; Liu, Y.; Zhu, Q.; Li, H. Learning to Measure Change: Fully Convolutional Siamese Metric Networks for Scene Change Detection. *arXiv* **2018**, arXiv:1810.09111.
109. Ren, C.; Wang, X.; Gao, J.; Zhou, X.; Chen, H. Unsupervised Change Detection in Satellite Images With Generative Adversarial Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10047–10061. [CrossRef]
110. Sakurada, K.; Okatani, T. Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; pp. 1–12.
111. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
112. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
113. 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.* **2018**, *29*, 545–559. [CrossRef]
114. 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.* **2019**, *16*, 266–270. [CrossRef]
115. Chen, H.; Shi, Z. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sens.* **2020**, *12*, 1162. [CrossRef]
116. Xufeng, H.; Leung, T.; Jia, Y.; Sukthankar, R.; Berg, A.C. MatchNet: Unifying feature and metric learning for patch-based matching. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3279–3286.
117. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Lin, T.; Li, H. DASNet: Dual attentive fully convolutional siamese networks for change detection of high resolution satellite images. *arXiv* **2020**, arXiv:2003.03608. [CrossRef]
118. Mesquita, D.B.; Santos, R.F.D.; Macharet, D.G.; Campos, M.F.M.; Nascimento, E.R. Fully Convolutional Siamese Autoencoder for Change Detection in UAV Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1455–1459. [CrossRef]
119. Liu, J.; Chen, K.; Xu, G.; Sun, X.; Yan, M.; Diao, W.; Han, H. Convolutional Neural Network-Based Transfer Learning for Optical Aerial Images Change Detection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 127–131. [CrossRef]
120. Xiang, S.; Wang, M.; Jiang, X.; Xie, G.; Zhang, Z.; Tang, P. Dual-Task Semantic Change Detection for Remote Sensing Images Using the Generative Change Field Module. *Remote Sens.* **2021**, *13*, 3336. [CrossRef]
121. 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 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 5–7 August 2019.
122. 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]
123. Song, L.; Xia, M.; Jin, J.; Qian, M.; Zhang, Y. SUACDNet: Attentional change detection network based on siamese U-shaped structure. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102597. [CrossRef]
124. Guo, H.; Shi, Q.; Marinoni, A.; Du, B.; Zhang, L. Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images. *Remote Sens. Environ.* **2021**, *264*, 112589. [CrossRef]
125. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support(DLMIA), Granada, Spain, 20 September 2018; pp. 3–11.
126. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
127. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *arXiv* **2016**, arXiv:1606.00915. [CrossRef] [PubMed]
128. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the International Conference on Learning Representations(ICLR), San Juan, Puerto Rico, 2–4 May 2016; pp. 1–13.
129. Szegedy, C.; Wei, L.; Yangqing, J.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
130. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef] [PubMed]
131. Nguyen, T.L.; Han, D. Detection of Road Surface Changes from Multi-Temporal Unmanned Aerial Vehicle Images Using a Convolutional Siamese Network. *Sustainability* **2020**, *12*, 2482. [CrossRef]

132. Li, X.; Duan, H.; Hui, Z.; Wang, F.-Y. Data Augmentation Using Image Generation for Change Detection. In Proceedings of the 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), Beijing, China, 15 July–15 August 2021; pp. 188–191.
133. Chen, H.; Li, W.; Shi, Z. Adversarial Instance Augmentation for Building Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
134. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
135. Li, X.; Yuan, Z.; Wang, Q. Unsupervised Deep Noise Modeling for Hyperspectral Image Change Detection. *Remote Sens.* **2019**, *11*, 258. [CrossRef]
136. Li, Y.F.; Guo, L.Z.; Zhou, Z.H. Towards Safe Weakly Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 334–346. [CrossRef]
137. Ibrahim, M.S.; Vahdat, A.; Ranjbar, M.; Macready, W.G. Semi-Supervised Semantic Image Segmentation with Self-correcting Networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2018; pp. 12712–12722.
138. Kerdegari, H.; Razaak, M.; Argyriou, V.; Remagnino, P. Urban scene segmentation using semi-supervised GAN. In Proceedings of the Image and Signal Processing for Remote Sensing XXV, Strasbourg, France, 9–11 September 2019.
139. Jiang, X.; Tang, H. Dense High-Resolution Siamese Network for Weakly-Supervised Change Detection. In Proceedings of the 2019 6th International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2–4 November 2019; pp. 547–552.
140. Khan, S.H.; He, X.; Porikli, F.; Bennamoun, M.; Sohel, F.; Togneri, R. Learning deep structured network for weakly supervised change detection. *arXiv* **2016**, arXiv:1606.02009.
141. Andermatt, P.; Timofte, R. A Weakly Supervised Convolutional Network for Change Segmentation and Classification. In Proceedings of the Asian Conference on Computer Vision(ACCV), Kyoto, Japan, 30 November–4 December 2020; pp. 103–119.
142. Song, H.; Kim, M.; Park, D.; Shin, Y.; Lee, J.-G. Learning from Noisy Labels with Deep Neural Networks: A Survey. *arXiv* **2020**, arXiv:2007.08199. [CrossRef] [PubMed]
143. Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
144. Wang, S.; Guan, Y.; Shao, L. Multi-Granularity Canonical Appearance Pooling for Remote Sensing Scene Classification. *IEEE Trans. Image Process.* **2020**, *29*, 5396–5407. [CrossRef]
145. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2020**, *13*, 3735. [CrossRef]
146. Hazel, G.G. Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1199–1211. [CrossRef]
147. Chen, C.; Fan, L. Scene segmentation of remotely sensed images with data augmentation using U-net++. In Proceedings of the 2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), Shanghai, China, 27–29 August 2021; pp. 201–205.
148. Wu, C.; Zhang, L.; Zhang, L. A scene change detection framework for multi-temporal very high resolution remote sensing images. *Signal Processing* **2016**, *124*, 184–197. [CrossRef]
149. Wang, Y.; Du, B.; Ru, L.; Wu, C.; Luo, H. Scene Change Detection VIA Deep Convolution Canonical Correlation Analysis Neural Network. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium(IGARSS), Yokohama, Japan, 28 July–2 August 2019; pp. 198–201.
150. Wu, C.; Zhang, L.; Du, B. Kernel Slow Feature Analysis for Scene Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2367–2384. [CrossRef]
151. Huang, X.; Liu, H.; Zhang, L. Spatiotemporal Detection and Analysis of Urban Villages in Mega City Regions of China Using High-Resolution Remotely Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3639–3657. [CrossRef]
152. Daudt, R.C.; Saux, B.L.; Boulch, A.; Gousseau, Y. Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 2115–2118.
153. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
154. Yang, J.; Jiang, Y.-G.; Hauptmann, A.G.; Ngo, C.-W. Evaluating bag-of-visual-words representations in scene classification. In Proceedings of the Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval, Augsburg, Bavaria, Germany, 24–29 September 2007; pp. 197–206.
155. Bernhard, S.; John, P.; Thomas, H. Efficient sparse coding algorithms. In Proceedings of the Conference on Neural Information Processing Systems(NIPS), Vancouver, British, 4–7 December 2007; pp. 801–808.
156. Burges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [CrossRef]
157. Zhang, L.; Zhang, L.; Tao, D.; Huang, X. On Combining Multiple Features for Hyperspectral Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 879–893. [CrossRef]
158. Chang, X.; Xiang, T.; Hospedales, T.M. Scalable and Effective Deep CCA via Soft Decorrelation. *arXiv* **2017**, arXiv:1707.09669.
159. Ru, L.; Du, B.; Wu, C. Multi-Temporal Scene Classification and Scene Change Detection With Correlation Based Fusion. *IEEE Trans. Image Process.* **2021**, *30*, 1382–1394. [CrossRef] [PubMed]

160. Gong, M.; Zhan, T.; Zhang, P.; Miao, Q. Superpixel-Based Difference Representation Learning for Change Detection in Multispectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2658–2673. [CrossRef]
161. Lei, Y.; Liu, X.; Shi, J.; Lei, C.; Wang, J. Multiscale Superpixel Segmentation With Deep Features for Change Detection. *IEEE Access.* **2019**, *7*, 36600–36616. [CrossRef]
162. Daudt, R.C.; Saux, B.L.; 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.
163. Hou, B.; Liu, Q.; Wang, H.; Wang, Y. From W-Net to CDGAN: Bitemporal Change Detection via Deep Learning Techniques. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1790–1802. [CrossRef]
164. Ding, Q.; Shao, Z.; Huang, X.; Altan, O. DSA-Net: A novel deeply supervised attention-guided network for building change detection in high-resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102591. [CrossRef]
165. 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]
166. Sanghyun, W.; Jongchan, P.; Joon-Young, L.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision(ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
167. Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [CrossRef]
168. Fisher, P. The pixel: A snare and a delusion. *Int. J. Remote Sens.* **1997**, *18*, 679–685. [CrossRef]
169. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
170. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497. [CrossRef] [PubMed]
171. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
172. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *arXiv* **2016**, arXiv:1612.08242.
173. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
174. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
175. Jocher, G. Yolov5. Available online: https://github.com/ultralytics/yolov5 (accessed on 10 March 2022).
176. Wang, Q.; Zhang, X.; Chen, G.; Dai, F.; Gong, Y.; Zhu, K. Change detection based on Faster R-CNN for high-resolution remote sensing images. *Remote Sens. Lett.* **2018**, *9*, 923–932. [CrossRef]
177. Zhang, L.; Hu, X.; Zhang, M.; Shu, Z.; Zhou, H. Object-level change detection with a dual correlation attention-guided detector. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 147–160. [CrossRef]
178. Han, P.; Ma, C.; Li, Q.; Leng, P.; Bu, S.; Li, K. Aerial image change detection using dual regions of interest networks. *Neurocomputing* **2019**, *349*, 190–201. [CrossRef]
179. Ji, S.; Shen, Y.; Lu, M.; Zhang, Y. Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
180. Benedek, C.; Sziranyi, T. Change Detection in Optical Aerial Images by a Multilayer Conditional Mixed Markov Model. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3416–3430. [CrossRef]
181. Bourdis, N.; Marraud, D.; Sahbi, H. Constrained optical flow for aerial image change detection. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 4176–4179.
182. Lebedev, M.A.; Vizilter, Y.V.; Vygolov, O.V.; Knyaz, V.A.; Rubis, A.Y. Change Detection in Remote Sensing Images Using Conditional Adversarial Networks. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 565–571. [CrossRef]
183. 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]
184. Rodrigo, C.D.; Bertrand, L.S.; Alexandre, B.; Yann, G. Multitask Learning for Large-scale Semantic Change Detection. *arXiv* **2018**, arXiv:1810.08452.
185. Yang, K.; Xia, G.-S.; Liu, Z.; Du, B.; Yang, W.; Pelillo, M.; Zhang, L. Asymmetric Siamese Networks for Semantic Change Detection in Aerial Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]
186. Tian, S.; Ma, A.; Zheng, Z.; Zhong, Y. Hi-UCD: A Large-scale Dataset for Urban Semantic Change Detection in Remote Sensing Imagery. *arXiv* **2020**, arXiv:2011.03247.
187. 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.* **2019**, *57*, 574–586. [CrossRef]
188. Fujita, A.; Sakurada, K.; Imaizumi, T.; Ito, R.; Hikosaka, S.; Nakamura, R. Damage detection from aerial images via convolutional neural networks. In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May 2017; pp. 5–8.
189. Shen, L.; Lu, Y.; Hao, C.; Wei, H.; Xie, D.; Yue, J.; Chen, R.; Zhang, Y.; Zhang, A.; Lv, S.; et al. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection. *arXiv* **2021**, arXiv:2107.09244. [CrossRef]

190. Ritwik, G.; Richard, H.; Sandra, S.; Nirav, P.; Bryce, G.; Jigar, D.; Eric, H.; Howie, C.; Matthew, G. Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–21 June 2019; pp. 10–17.
191. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
192. Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131. [CrossRef]
193. Li, B.; Zhou, Q. Accuracy assessment on multi-temporal land-cover change detection using a trajectory error matrix. *Int. J. Remote Sens.* **2009**, *30*, 1283–1296. [CrossRef]
194. Pratomo, J.; Kuffer, M.; Kohli, D.; Martinez, J. Application of the trajectory error matrix for assessing the temporal transferability of OBIA for slum detection. *Eur. J. Remote Sens.* **2018**, *51*, 838–849. [CrossRef]
195. Gong, J.; Hu, X.; Pang, S.; Wei, Y. Roof-Cut Guided Localization for Building Change Detection from Imagery and Footprint Map. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 543–558. [CrossRef]