remote sensing

MDPI

Review

A Review of Deep-Learning Methods for Change Detection in Multispectral Remote Sensing Images

Eleonora Jonasova Parelius

Norwegian Defence Research Establishment (FFI), NO-2007 Kjeller, Norway; eleonora-jonasova.parelius@ffi.no

Abstract: Remote sensing is a tool of interest for a large variety of applications. It is becoming increasingly more useful with the growing amount of available remote sensing data. However, the large amount of data also leads to a need for improved automated analysis. Deep learning is a natural candidate for solving this need. Change detection in remote sensing is a rapidly evolving area of interest that is relevant for a number of fields. Recent years have seen a large number of publications and progress, even though the challenge is far from solved. This review focuses on deep learning applied to the task of change detection in multispectral remote-sensing images. It provides an overview of open datasets designed for change detection as well as a discussion of selected models developed for this task—including supervised, semi-supervised and unsupervised. Furthermore, the challenges and trends in the field are reviewed, and possible future developments are considered.

Keywords: change detection; remote sensing; optical imaging; multispectral imaging; deep learning

1. Introduction

Remote sensing (RS) denotes the acquisition of information about an object from a distance. Often, as will also be the case here, the term is used more specifically to refer to the imaging of the Earth's surface from above, such as from a satellite or an aircraft. Technological development has led to an unprecedented amount of RS imagery being available today. The information in these images is of interest for a number of fields and applications, such as cartography [1], agriculture [2,3], nature conservation [4,5], climate [6,7] and disaster monitoring [8,9], archaeology [10,11], law enforcement [12] and urban planning [13].

The amount of information provided by RS poses a challenge in filtering out the relevant data. Manual exploration of the images is slow and laborious, and most applications, thus, require methods for the efficient processing of RS imagery. One of the tasks common to practically all fields where RS is used is change detection (CD)—or, more accurately stated: relevant change detection.

1.1. Detection of Relevant Changes in Remote Sensing Images

Change detection in the context of remote sensing refers to the process of identifying differences in the structure and/or properties of objects and phenomena on Earth by analysing two or more images taken at different times [14]. Change detection can serve as a basis for understanding the development of various natural or human-related phenomena through time and the interactions between them.

The goal of change detection is usually to identify the pixels within the (two or more) images that correspond to changed objects on the ground. It is, however, also possible to work at the level of a scene, i.e., to identify whether the classification of the depicted scene has changed (for example, a field turning into a residential area). Some methods also seek not only to identify changed pixels (or scenes) but also to classify the type of change that has occurred, referred to as semantic change detection.

check for updates

Citation: Parelius, E.J. A Review of Deep-Learning Methods for Change Detection in Multispectral Remote Sensing Images. *Remote Sens.* **2023**, *15*, 2092. https://doi.org/10.3390/rs15082092

Academic Editor: Adrian Stern

Received: 8 March 2023
Revised: 8 April 2023
Accepted: 12 April 2023
Published: 16 April 2023

The main challenge of change detection lies in the identification of changes that are relevant for the given task. Observed changes can be divided into three categories—apparent, irrelevant and relevant changes. The first category—apparent changes—comprises the changes seen in the images that do not result from actual changes happening to the depicted objects. Instead, changes resulting from variations in imaging equipment and circumstances, such as light and atmospheric conditions, belong in this category.

The other two categories—relevant and irrelevant changes—include all the real changes that are happening to the observed objects. The boundary between relevant and irrelevant changes is entirely dependent on the application. For instance, the relevance of seasonal changes, such as snow cover or the state of the foliage on the vegetation, is determined by the specifics of the task.

While snow cover is unimportant in urban planning and might be considered an irrelevant change, it is highly relevant for assessing the state of glaciers. Similarly, the level of vegetation hydration is of no interest in cartography but is essential in the monitoring of draughts or in crop assessment. The relevance of human-made changes is also determined by the specifics of the task, e.g., the presence or absence of vehicles is irrelevant in archaeology or cartography but is important for activity monitoring.

1.2. Multispectral Imagery

RS can be divided into active and passive. Active RS methods, such as radar, use a signal generated by the system, the echo of which is then detected by its sensors. Passive RS relies solely on naturally occurring radiation reflected and/or emitted from the surface. This review focuses on the latter, more specifically on remote sensing using optical multispectral images.

Optical images can be divided into hyperspectral, multispectral and panchromatic. Multispectral images have several (usually 3–15) spectral bands, while hyperspectral images sample the spectrum much more finely with tens to hundreds of narrow bands. Panchromatic images gather all the visible light into one single band. Multispectral sensors typically work in the visual, near infrared and short-wave infrared range of the spectrum. Red, green and blue (RGB) images are an often-used and well-known subset of multispectral imagery.

The wavelength distribution of the light reflected or emitted from an object contains valuable information about the object's nature, reflecting its material composition and properties, which can, in turn, be used in RS for various tasks, such as classification or change detection. Hyperspectral images have the highest spectral resolution, while panchromatic have the lowest.

Hyperspectral images with their high spectral resolution can then take full advantage of the information provided by fine sampling of the objects' spectral signature. The large number of bands makes hyperspectral imaging well suited for anomaly detection [15] but also allows for applications in change detection [16].

However, in order to maintain a high signal-to-noise ratio, the dense spectral sampling of hyperspectral sensors comes at the expense of lower spatial resolution and/or lower sensor coverage, as well as additional challenges due to the greatly increased complexity and very high dimensionality of the data [17]. In addition, full spectral sampling is not always necessary. Depending on the application, a number of spectral bands exists that will provide the most utility, and adding more bands will often lead to diminishing returns [18,19].

Multispectral sensors lie between the two extremes of panchromatic and hyperspectral images both in terms of spectral and spatial resolution. While the number of bands is considerably lower than that of hyperspectral images, they have moderate-to-high spatial resolution while still containing important information regarding the colours, textures and material properties of the imaged objects without excessively increasing the data dimensionality.

In addition, multispectral images (and RGB colour images especially) are very common—nearly ubiquitous, simple and affordable to obtain and, thus, an important resource in RS. They are also easy to interpret (even for the general public) and, in that regard, easier to manually annotate.

Multispectral images inherit some challenges both from panchromatic and hyperspectral images. Due to their higher number of spectral bands, there is a need to identify and utilize this additional information efficiently. On the other hand, while their high spatial resolution provides valuable information that can be used for change detection, it also captures more variability, details and higher frequency noise, which, in turn, makes automated change detection more challenging, especially for data with a limited number of spectral bands. With more detailed and challenging terrain, the need for precise co-registration increases [20], and the accurate detection of complex boundaries between objects becomes more important [21].

1.3. Change-Detection Methods

The task of identifying changes in images of the Earth's surface is not a recent problem [14]. The techniques used to accomplish this task can be divided into four groups: algebra-based, statistics-based, transformation-based and deep-learning-based.

1.3.1. Algebra-Based Methods

Algebraic methods usually consist of two steps. In the first step, a difference image is constructed, for example, by taking a difference or a ratio of the pixel values of the two images, and, in the second step, a form of thresholding is used to create a change map. Some examples of algebra-based change-detection methods include image differencing, image regression, image ratioing and change vector analysis (CVA). These methods can detect changes that are greater than a chosen threshold. The choice of threshold is crucial for the performance of these algorithms as it determines their specificity and sensitivity, and it has to be adjusted for each unique dataset and each goal [22].

1.3.2. Statistics-Based Methods

Another group of methods are based on the statistical properties of either the whole image or parts of it. The distribution of pixels and their properties are then used to detect anomalous or different objects between the images. Statistical techniques are most often used for hyperspectral [16] and synthetic aperture radar (SAR) images [23,24], as these tend to exhibit fewer apparent changes than high-resolution multispectral images.

1.3.3. Transformation-Based Methods

Transformation-based methods, such as principal component analysis (PCA), multivariate alteration detection (MAD), Gramm–Schmidt and tasseled cap, rely on first transforming the images in a way that enhances the changes and suppresses the apparent differences of the unchanged areas. Some of these methods, such as PCA, can also be categorized as classical machine learning. These methods are able to emphasise the information of interest. However, they still rely on the appropriate selection of a threshold for detecting changes, and the interpretation of the changed areas can be more difficult when using the transformed images [22].

Pixel-wise classification of images preceding change detection can also be included in the category of transformation-based methods. In effect, it transforms an image into a new simplified version of it—a class map—that makes it easier to identify changes. The objects in the images are first classified into their categories through semantic segmentation, and then the classification maps from different times are simply compared in order to uncover changes. This process, however, is not sensitive to changes that happen without a change of category, such as changes to a single building.

Methods used for classification often fall into some of the other categories. For example, a model by Zhang et al. [25] first performs a deep-learning-based classification before the classes of each pixel pair are compared and evaluated for class change.

1.3.4. Deep-Learning-Based Methods

The last group of methods is deep-learning-based change detection. Deep learning is a subset of machine learning. It uses neural networks consisting of a large number of layers (hence, the name deep) to learn and represent data.

With the development and growing popularity of deep-learning methods within computer vision, it is natural to also apply them to the problem of CD in remote sensing. Deep-learning models are able to represent complex and hierarchical features within the data, which makes them good candidates for RS CD.

One of the biggest challenges of change detection is the presence of apparent changes. Every pixel from the image taken at an earlier time can be changed in the image taken at a later time, without there being any changes to the objects depicted in them. This poses a challenge for the more classical algorithms (algebraic, statistic and transformation-based), which are more rigid and not able to represent complex features, thus, leading to many false positives and a need to fine-tune the detection threshold for each application. Due to the flexibility, scalability and hierarchical structure of deep-learning models, they have the potential to learn to represent data that are too complex to be described by simpler models. Each pixel can thus be considered within the context of its surrounding pixels, as higher-level features are used for decision making.

Classical approaches to multispectral image exploitation often rely on spectral indices [26], which combine various bands in order to emphasize the desired properties. Choosing the right index or combination of bands is a task requiring expert knowledge, and, in the end, only a portion of the available information (the information contained in the selected bands) is used. Deep learning allows the use of all the available bands without the need for expert-led pre-selection. The importance and contribution of each band is simply learned by the model in the training process.

Deep-learning change-detection models can be broadly divided into two categories: fully supervised models and models that are not fully supervised, i.e., semi-supervised and unsupervised. Fully supervised methods almost always require a large amount of labelled data in order to train the network, while semi- and unsupervised methods reduce or eliminate the need for ground-truth-labelled data.

1.3.5. About This Work

In recent years, a growing number of articles has been published on this topic as well as several reviews. Four reviews specifically on change detection using deep learning have been published recently by Shi et al. [27] and Khelifi et al. [28] in 2020; and Shafique et al. [29] and Jiang et al. [30] at the beginning of 2022. In addition, there are a number of reviews dealing with classical RS CD [14,16,22,31] or with deep learning in RS [32–36] in general. Due to the rapidly growing interest in this topic and a large number of new models introduced in recent years, our review reports on methods not previously included in review articles and focuses solely on multispectral change detection and its specific challenges in greater depth. We also provide an in-depth overview and comparison of some of the most relevant CD models.

The rest of this review is organized as follows: first, an overview and discussion of openly available multispectral datasets for CD is provided. The need for (annotated) datasets is an important requirement for most deep-learning models and is a common bottleneck in the process of developing reliable CD methods. Then, a variety of deep-learning models for change detection are presented, divided into supervised methods and un-/semi-supervised methods. The most common model structures are identified, and published CD networks are categorized, described and compared. Finally, the challenges and outlooks of deep-learning-based CD are discussed.

2. Data Sets

As with most applications of deep learning, the development of reliable models heavily depends on the availability of large annotated datasets. This issue can be circumvented by various semi-supervised or fully unsupervised methods; however, even in these cases, there can be a need for at least some (although often fewer) annotated examples.

In this section, we list and describe some of the most used freely available datasets for change detection in multispectral RS images. The open datasets and some of their key characteristics are listed in Table 1. Links to all of the datasets are in the Appendix A.

Table 1. Openly available bitemporal multispectral remote sensing datasets annotated for change detection.

Data Set	Number of Image Pairs	Image Size	Number of Pixels	Resolution (m)	Number of Bands	Year
SZTAKI [37]	13	952 × 640	8×10^6	1.5	3	2008
AICD [38]	1000	800 × 600	4.8×10^8	0.5	3	2011
OSCD [39]	24	600 × 600	8.6×10^6	10, 20, 60	13	2018
CDD [40]	16,000	256 × 256	1×10^9	0.03–1	3	2018
WHU Building CD [41]	1	32,507 × 15,345	5×10^8	0.075	3	2018
HRSCD [42]	291	10,000 × 10,000	3×10^{10}	0.5	3	2019
LEVIR-CD [43]	637	1024 × 1024	6.7×10^8	0.5	3	2020
DSIFN [44]	394	512 × 512	1×10^8	2	3	2020
MtS-WH [45,46]	1	7200 × 6000	4.3×10^7	1	4	2020
Google Data Set [47]	1067	256 × 256	7×10^7	0.55	3	2020
SYSU-CD [48]	20,000	256 × 256	1.3×10^9	0.5	3	2021
SECOND [49]	4662	512 × 512	1×10^9		3	2021
3DCD [50]	472	400 × 400	7.6×10^7	0.5	3	2022
Hi-UCD [51]	40,800	512 × 512	1×10^{10}	0.1	3	2022
Landsat-SCD [52]	8468	416 × 416	1.5×10^9	30	3	2022

Dataset	Origin	Type of changes
SZTAKI	Aerial, Hungary	Buildings, building sites, groundwork, ploughed land, large groups of trees
AICD	Synthetic	Buildings
OSCD	Sentinel-2, World	Buildings and roads
CDD	Aerial (Google Earth)	Buildings, roads, vehicles, not seasonal changes
WHU Building CD	Aerial, Christchurch, New Zealand	Buildings
HRSCD	Aerial, France	Semantic, artificial surfaces, agricultural areas, forests, wetlands, water
LEVIR-CD	Aerial (Google Earth), Texas	Buildings
DSIFN	Aerial, China	Buildings, roads
MtS-WH	IKONOS, Wuhan	Scene classification, parking, water, sparse/dense houses, residential, idle, vegetation, industrial
Google Data Set	Aerial (Google Earth)	Buildings
SYSU-CD	Aerial, Hong Kong	Buildings, groundwork, change of vegetation, roads, sea constructions
SECOND	Aerial, China	Semantic, non-vegetated ground surface, trees, low vegetation, water, buildings, playgrounds
3DCD	Aerial, Valladolid, Spain	Based on changes to elevation—mostly focused on buildings.
Hi-UCD	Aerial, Tallinn, Estonia	Semantic, 9 types of land cover, 48 types of semantic change. Water, grass, building, greenhouse, road, bridge, bare land, woodland, other
Landsat-SCD	Landsat series, China	Semantic, time series with 10 land cover change types

One of the datasets is composed of Sentinel-2 images, namely, the Onera Satellite Change Detection dataset (OSCD). This dataset has the largest number of bands (13) and contains images from a wide range of areas in the world. Another dataset composed of satellite images is the Multi-temporal Scene Wuhan (MtS-WH) dataset, acquired using IKONOS and focusing on the region of Wuhan city. The MtS-WH dataset provides pairs of images of the same area from different times, thus, allowing for change detection. However, the annotation does not focus on pixel-level changes, as is the case for the rest of the datasets mentioned here. Instead, it is based on scene classification, where frames of 150 × 150 pixels are classified into various categories. Images containing a mix of categories are not classified.

The remaining datasets are aerial datasets, often acquired via Google Earth and focusing on smaller regions, such as one city, area, state or country. They are composed of high-resolution RGB images that are often with a spatial resolution below 1 m. The list also contains one synthetic dataset—the Aerial Imagery Change Detection (AICD) dataset.

Change detection can be described as a weakly-defined problem because the nature of the changes to be detected depends, to a great extent, on the goal of the investigation. This is reflected in the variety of changes that the datasets focus on. Several focus entirely on changes to buildings, such as the AICD, the WHU Building CD (which is a subset of a larger WHU Building dataset made for building segmentation), the three-dimensional change detection (3DCD) dataset and the Google Data Set. Other datasets include additional types of changes. Changes to roads, groundwork and vegetation are commonly included.

Among these, the SZTAKI dataset (composed of two datasets, named Szada and Tiszadob, which are often treated separately), the OSCD, the Change Detection (CDD) dataset, the Sun Yat-Sen University Change Detection (SYSU-CD) dataset and the Deeply Supervised Image Fusion Network (DSIFN) dataset. The rules guiding the annotation of ground truth changes in these datasets are not identical. While, for example, changes to vegetation are marked as a change in the SYSU-CD dataset, the CDD includes a large number of seasonal variations and does not mark most changes to vegetation and snow cover as a change. Examples from several of the datasets are shown in Figure 1.

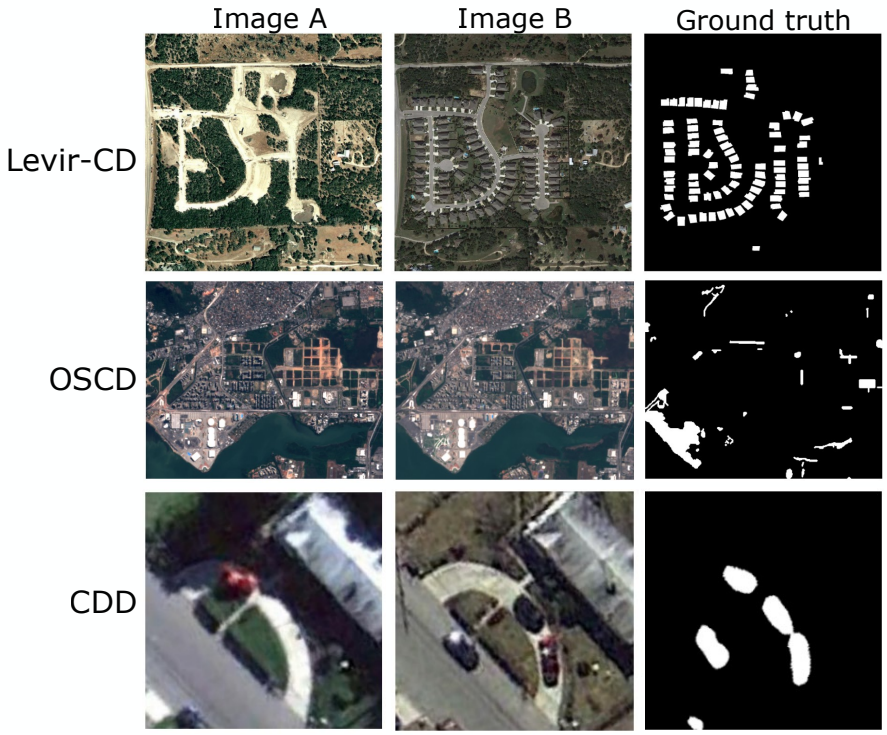

Figure 1. Examples of images and ground truth from various datasets, namely, Levir-CD [43] (with a focus on building changes), OSCD [39] and CDD [40] (various changes to infrastructure annotated).

Lastly, several of the datasets provide semantic information on the type of the changes taking place between the two imaged time points. The High-Resolution Semantic Change Detection (HRSCD) dataset (shown in Figure 2) is a large dataset containing over 10-fold more pixels compared with the next largest one. It contains aerial images from the BD ORTHO database of Institut géographique national (IGN), depicting two areas of France, where land cover maps from Urban Atlas were used for ground truth annotation. This dataset, thus, contains ground truth specifying the land cover class as well as the change to the land cover class.

The classes, however, are fairly broadly defined, such as artificial surfaces or agricultural areas. Hence, this dataset is not suitable for the detection of smaller changes, such as changes to individual buildings, the presence/absence of vehicles or the state of vegetation. The relatively small number of broadly-defined classes also means that changes are rarer, i.e., 99.232% of pixels are labelled as no change, which is a larger proportion than for most other datasets.

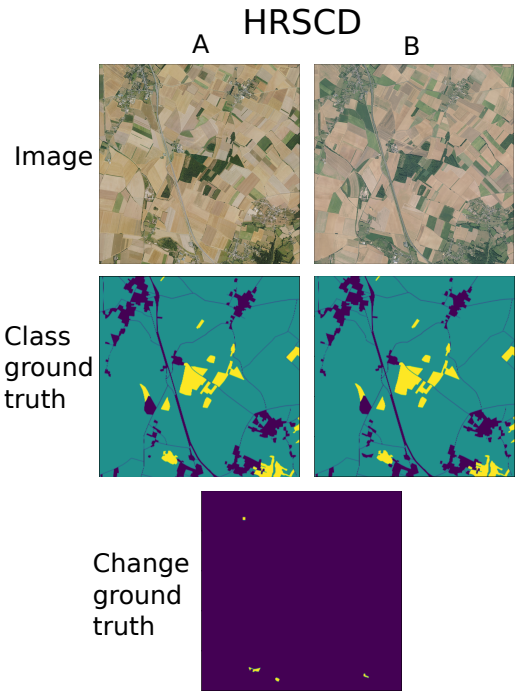

Figure 2. Examples of images, ground truth classes and change ground truth for the semantic change-detection dataset HRSCD [42].

The second semantic dataset is the SEmantic Change detectiON Dataset (SECOND). This dataset features aerial images from several cities in China, and the ground truth provides land-cover segmentation into six different categories—non-vegetated ground surface, trees, low vegetation, water, buildings and playgrounds. The change ground truth is obtained indirectly by comparing the annotated land-cover categories.

The Ultra-High Urban Change Detection (Hi-UCD) dataset is a large and detailed land cover and change dataset of optical images with a very high spatial resolution of 0.1 m. It provides both land cover labels, as well as 48 types of semantic change. This dataset has not yet been openly published as of December 2022, but the authors are planning to do so.

Another very recently published semantic dataset is the Landsat Semantic Change Detection (Landsat-SCD) dataset. Unlike the previous ones, it does not provide land cover types, only land cover changes and it is the only listed dataset that has more than two images of a given location, namely, it includes time series of 28 images spanning a total time span of 1990–2020 with individual images being 3–23 years apart. The datasets vary in size considerably, with the total number of pixels in the largest dataset (HRSCD) being 3×10^{10}, while some of the smaller datasets contain several million pixels.

3. Supervised Deep-Learning Models for Multispectral Change Detection

Supervised change-detection methods require annotated data in order to train the network. Unlike many of the staple tasks in computer vision (such as classification or semantic segmentation), change detection receives as input two (or more) images, rather than a single one, along with a single ground truth image. The two input images can be processed in various ways, which can be roughly divided into single- and double-stream structures.

In the single-stream structure, also referred to as early fusion, the two images are joined before they are fed to the network, either by simple concatenation or other procedures, such as differencing.

The alternative—double-stream structure—is based on processing each of the images on its own before they are again joined together and compared. Double-stream architectures (depicted in Figure 3c) feature two identical subnetworks that run parallel to each other, with each one taking one of the images as input. This type of structure is usually referred to as Siamese, if the subnetworks share weights, and pseudo-Siamese, if they do not share weights.

An overview of selected supervised change-detection networks is shown in Table 2 along with a categorization of their network structure, the dataset they were applied to by their authors and notes detailing their features.

Table 2. Overview of selected supervised change-detection models published between 2018 and 2022, listed chronologically.

Network Name		Network Structure	Data Set	Note	Year
TransUNetCD [53]	Transformer + UNet CD	Double-Stream UNet + Transformer	WHU, CDD, LEVIR, DSIFN	UNet + transformer	2022
UVACD [54]		Double-Stream CNN + Transformer	LEVIR, WHU	transformer	2022
ChangeFormer [55]	Change Transformer	Double-Stream Transformer	LEVIR, DSIFN	transformer, 4 feature difference modules, simple decoder	2022
Pyramid-SCDFormer [52]	Pyramid, semantic CD Transformer	Double-Stream Transformer	WHU, LEVIR	transformer encoders, MLP decoder, conv units with different kernels	2022
MAEANet [56]	Multi-scale Attention and Edge-Aware Net	Double-Stream UNet, Attention	WHU, LEVIR	spatial & contour attention, UNet for feature extraction + feature fusion	2022
FTN [57]	Fully Transformer Net	Double-Stream Transformer, Attention	WHU, LEVIR, SYSU, Google	Swin transformers, attention, multiple loss functions	2022
MCTNet [58]	Multi-Scale CNN Transformer Net	Double-Stream UNet/Transformer hybrid	LEVIR, CDD	hybrid ConvTrans blocks	2022
MFATNet [59]	Multi-Scale Feature Aggregation via Transformer	Double-Stream Transformer, Attention	WHU, LEVIR, DSIFN	feature extracion by ResNet, input to transformer, channel attention	2022
RFNet [60]	Region-Based Feature Fusion Net	Double-Stream CNN	WHO, SECOND	CNN, multi-level feature fusion, region similarity module	2022
AFSNet [61]	Attention-Guided Siamese Full-Scale Feature Aggregation Net	Double-Stream UNet-like, Attention	LEVIR, CDD	full-scale skip connections, spatial and channel attention	2022
IRA-MRSNet [62]	Multi-Scale Residual Siamese Network fusing Integrated Residual Attention	Double Stream UNet-like, Attention	CDD, WHU, LEVIR, SYSU	MultiRes blocks (fusion of different size kernels) instead of traditional convolutions, channel attention	2022
Recurrent CNN [63]		Double-Stream + LSTM	Taizhou	LSTM	2018
FC-EF [64]	Fully Conv. Early Fusion	Single-Stream UNet	SZTAKI, OSCD	early fusion	2018
FC-Siam-conc [64]	Fully Conv. Siamese Concatenation	Double-Stream UNet	SZTAKI, OSCD	Siamese concatenation	2018
FC-Siam-diff [64]	Fully Conv. Siamese Difference	Double-Stream UNet	SZTAKI, OSCD	Siamese difference	2018
SSJLN [65]	Spectral-spatial joint learning	Double-Stream	other	new loss	2019
DLSF [66]	Dual-learning Siamese	Double-Stream+ GAN	SZTAKI, other	GAN-domain transfer	2019
CD-UNet++ [67]	Change Detection UNet++	Single-Stream UNet	CDD	UNet++	2019
DSMS-FCN [68]	Deep Siamese Multi-scale FCN	Double-Stream UNet	other	conv units with different kernels	2019
FC-EF-Res [42]	Fully Conv. Early Fusion Residual	Single-Stream UNet	HRSCD, OSCD	landcover mapping + CD in one	2019

Table 2. *Cont.*

Network Name		Network Structure	Data Set	Note	Year
UNetLSTM [69]		Double-Stream UNet + LSTM	OSCD	LSTM	2019
SiamCRNN [70]	Siamese Conv. RNN	Double-Stream + LSTM	other	LSTM	2019
STANet [43]	Spatial-Temporal Attention Net	Double-Stream, Attention	LEVIR, SZTAKI	ResNet, attention	2020
DSIFN [44]	Deeply Supervised Image Fusion	Double-Stream UNet	CDD, DSIFN		2020
TCDNet [71]	Trilateral CD Net	3× Double-Stream	other	parallel CNNs, dilated conv	2020
DASNet [72]	Dual Attentive Siamese Net	Double-Stream, Attention	CDD	VGG16, attention	2020
AG-GAAN [73]	Attention Gates Generative Adversarial Adaptation Net	GAN, Attention	CDD	attention, new loss, GAN	2020
SNUNet-CD [74]	Siamese Network UNet	Double-Stream UNet	CDD	Nested UNet, attention	2021
CLNet [75]	Cross-Layer CNN	Single-Stream UNet, Attention	CDD, LEVIR, WHU	conv with different strides	2021
SRCDNet [76]	Super-Resolution CD Net	GAN + Double-Stream	CDD, Google	ResNet	2021
ESCNet [77]	End-to-end Superpixel	(FE) Superpixel segm + Double-Stream UNet	CDD SZTAKI	superpixel segmentation	2021
CapsNet [78]	Capsule Net	Double-Stream	SZTAKI, other	capsule network	2021
BIT_CD [79]	Bitemporal Image Transformer CD	Double-Stream Transformer	LEVIR, WHU, DSIFN	ResNet18, then transformer	2021
CEECNet [80]	Compress–Expand/Expand–Compress Net	Double-Stream, Attention	LEVIR, WHU	attention, CEEC unit, new loss	2021
FDORNet [81]	Feature Decomposition–Optimization–Reorganization Net	Double-Stream	LEVIR	boundary extraction, strided conv	2022
MLDANets [82]	Multilevel Deformable Attention-Aggregated Networks	Double-Stream UNet, Attention	LEVIR, SECOND	attention module with deformable sampling	2022
Siamese_AUNet [83]	Siamese attention + UNet	Double-Stream UNet, Attention	LEVIR, WHU, SZTAKI	attention, atrous spatial pyramid pooling	2022
DARNet [84]	Densely Attentive Refinement Nets	Double-Stream UNet, Attention	CDD, SYSU, LEVIR	attention and refinement module	2022
SwinSUNet	Swin Transformer Siamese U-shaped Net	Double-Stream Transformer	CDD, WHU, OSCD, HRSCD	Swin transformer	2022
UCDNet [85]	Urban Change Detection Net	Double-Stream UNet	OSCD	residual connections, new spatial pyramid pooling, new loss	2022
BESNet [21]	Boundary Extraction Constrained Siamese Net	Double-Stream	CDD, DSIFN, LEVIR	boundary extraction	2022
HFA-Net [86]	High Frequency Attention Net	Double-Stream UNet, Attention	WHU, LEVIR, Google	attention, boundary	2022
ISNet [87]	Improved Separability Net	Double-Stream, Attention	LEVIR, SYSU, CDD	attention, margin maximization	2022

Among the reviewed models, networks based on the UNet architecture [88] are the most common. In fact, all of the featured single-stream networks are UNet-based. The general structure of a UNet-based single-stream network is shown in Figure 3a. A UNet can also be employed in a double-stream (Siamese) manner (Figure 3b), which can be seen as a subcategory of double-stream structures. In this case, however, the feature extraction and fusion processes are intertwined, as the fusion is happening at several feature extraction stages rather than at the end of it.

In the following, the use of UNets in CD will first be discussed before describing other types of networks.

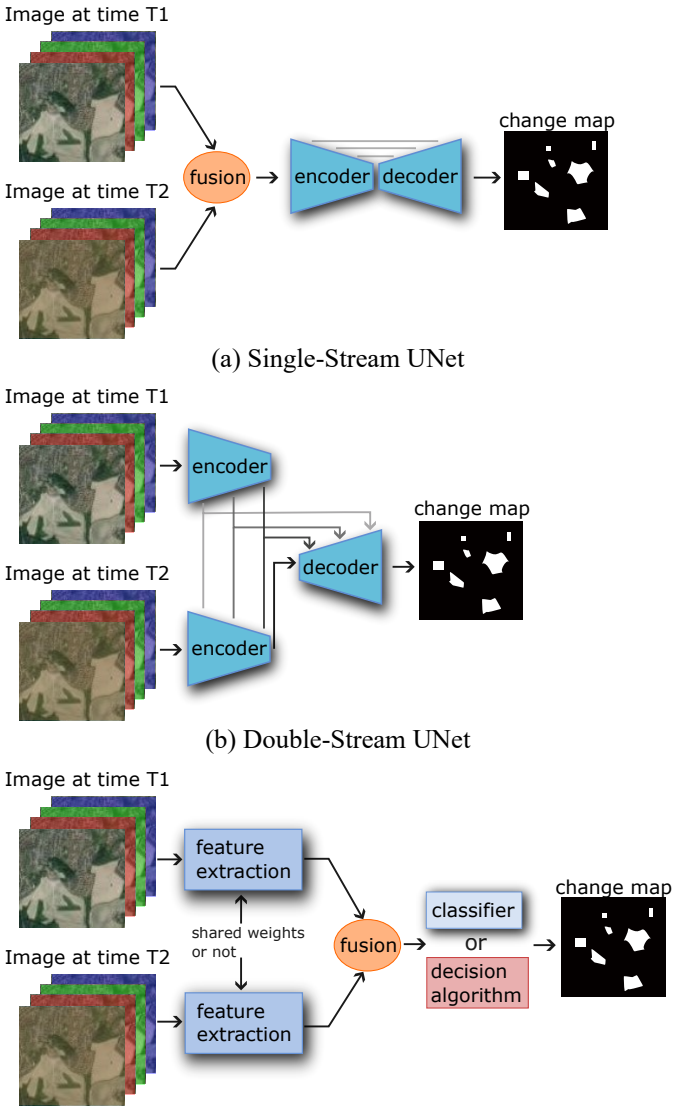

Figure 3. Structures of supervised CD models. (**a**) Single-Stream UNet-like network. The image data from two time points is first fused together, usually by concatenation, and then input into a UNet-like network featuring an encoder, a decoder and skip connections for semantic segmentation. The output of the model is a change map. (**b**) Double-Stream UNet. Each images is input separately into the encoder. The output of the two encoders is fused on multiple levels and fed to a single decoder, which produces a change map. (**c**) General Siamese Feature Extraction-based network structure. The individual images are first input into two identical Siamese (shared-weights) or pseudo-Siamese (different weights) subnetworks for feature extractions. The extracted features are then fused together, and either an automatic decision model or a machine-learning-based classifier is then used to produce a change map.

3.1. UNet in Change Detection

UNets and various modified versions of UNets are often used for the task of change detection. This fully convolutional neural network was developed for semantic segmentation [88] and is, thus, well suited for the task of pixel-wise change detection as well. In order to provide a class prediction for every pixel in the image, a UNet consists of an encoder and a decoder part. A contracting feature extractor, the encoder, is followed by an expanding decoder, which uses upsampling instead of pooling in order to increase the resolution of the output.

The contracting and expanding parts of the network are symmetrical and joined by so-called skip connections where outputs of the encoder are concatenated with the inputs to the decoder at the same level of contraction. Single-Stream UNet-like networks have a structure as depicted in Figure 3a. The images are fused in the first step, usually by simple concatenation. They are then input into a UNet-like network where features are first extracted by the encoder and then up-sampled by the decoder.

UNets can also be employed in the form of a Siamese architecture (as shown in Figure 3b). In this case, each of the images is progressing through the encoder part of the model separately, while the fusion takes place at the level of the skip connections with the decoder.

One of the earliest and most cited neural networks for supervised change detection is indeed a version of a UNet introduced by Daudt et al. in 2018 [64]. This work reports on three variations of a UNet-inspired fully convolutional neural network. All three networks consist of an encoder and a decoder with skip connections between them linking the layers with the same subsampling scales. The three networks are depicted in Figure 4, reproduced from the original article introducing these networks [64].

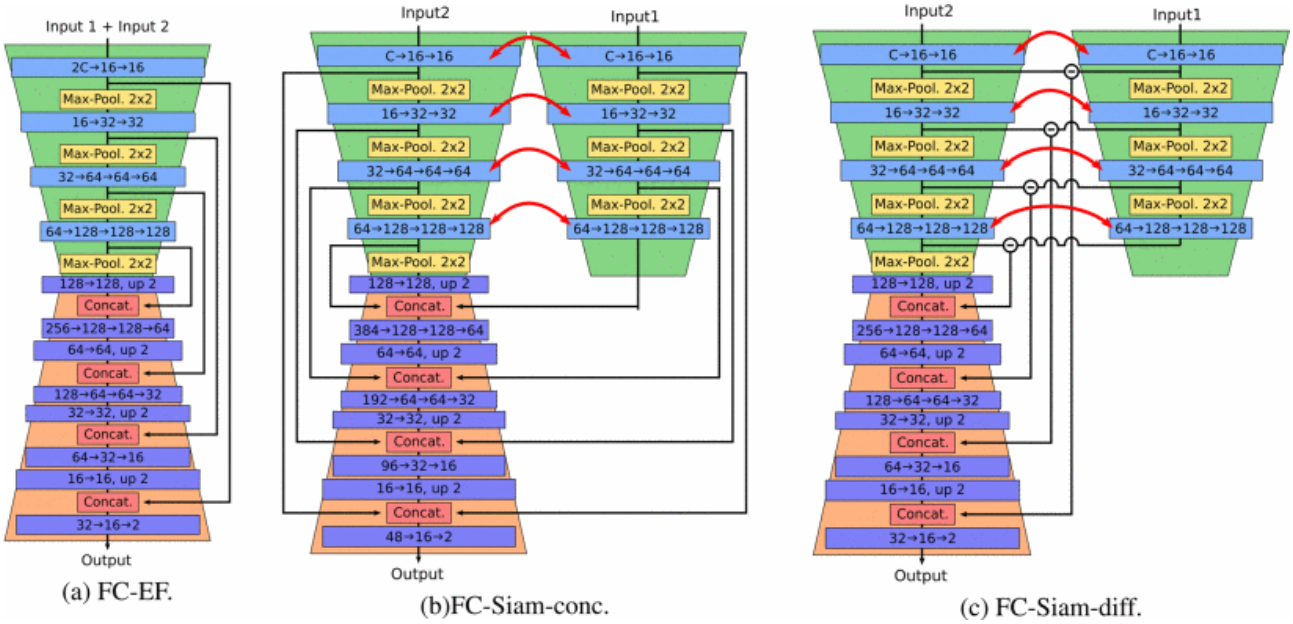

Figure 4. Schematics of the three architectures proposed by Caye Daudt et al. [64] for change detection. (**a**) Fully Connvolutional Early Fusion (FC-EF) network, which is a single-stream UNet-like network. (**b**) Fully Convolutional Siamese—Concatenation (FC-Siam-conc) and (**c**) Fully Convolutional Siamese—Difference (FC-Siam-diff) networks which are double-stream UNets. Block colour legend: blue is convolution, yellow is max pooling, red is concatenation, and purple is transpose convolution. Red arrows illustrate shared weights.

The first network, Fully Convolutional Early Fusion (FC-EF), is single-stream. It concatenates the two input images before feeding them to the network, thus, treating them simply as different bands. The other two networks have a Siamese architecture where the encoder part of the network is run in parallel for each of the images, and they differ in how the skip connections are performed. For FC-Siam-conc (Fully Convolutional Siamese/hl—Concatenation), the two skip connections coming from each encoding stream are simply concatenated before joining the decoder, while in the second network FC-Siam-diff (Fully Convolutional Siamese—Difference) the skip connections from each stream are first integrated together by taking the absolute value of their difference. All three network variations achieved good results surpassing classical approaches, with FC-Siam-diff being the best, closely followed by FC-EF.

While the early fusion type network did not perform badly in comparison with the Siamese structure in the aforementioned study, Siamese architectures have been favoured by most authors in the subsequent years.

In any case, the UNet has featured heavily in the field of change detection, either as the original, or as a large number of variations, all having the same key components: a contracting encoder, an expanding decoder and skip connections connecting them at multiple levels.

3.2. Single-Stream Networks—UNets

Apart from the work of Daudt et al. [64] introduced in the previous Section 3.1, only three more of the reviewed articles featured a single-stream architecture, and they are all variations of a UNet.

The first one is a Fully Convolutional Early Fusion Residual (FC-EF-Res) network [42], which is a modified FC-EF UNet with added residual blocks. It combines several UNets to perform both land cover mapping and change detection simultaneously.

The second one, CD-UNet++ [67], is a UNet that fuses information from various levels through upsampling and dense skip connections. The output from the various levels can be combined for change prediction.

The third one is Cross-Layer Network (CLNet) [75], which is a modified UNet with novel Cross Layer Blocks (CLB). A CLB splits the input temporarily into two parallel but asymmetrical branches, which use different convolution strides in order to extract multi-scale features.

3.3. Double-Stream Networks

Double-stream networks are based on initial feature extraction (FE) from each of the images and subsequent comparison of the features leading to identification of changes. The images are first processed in parallel for FE, and then the outputs are joined to be fed to the decision making part of the model as seen in Figure 3c. The parallel feature extraction subnetworks can share weights (Siamese networks), which is more common, or not share weights (pseudo-Siamese). The extracted features are then fused together, often employing some form of attention, and subsequently input into an algorithm that compares them and outputs a change map. The algorithm responsible for comparing the feature maps can either be a more traditional, automatic decision algorithm, such as a Gaussian distance, or it can be a neural network-based classifier, which needs to be trained with the rest of the network.

3.3.1. UNets

Double-stream UNets, similar to the aforementioned (Section 3.1) FC-Siam-conc and FC-Siam-diff, can be viewed as a subcategory of general double-stream networks. The UNetLSTM [69] combines a UNet with a long short-term memory (LSTM) [89] block, which is a type of a recurrent neural network (RNN). Its UNet-based encoder–decoder architecture has a convolutional LSTM block added at each encoder level. This LSTM block is inserted at the level where the comparison of the two images takes place.

The Siamese Network UNet for Change Detection (SNUNet-CD) [74] is, similarly to the aforementioned CD-UNet++, based on a UNet++ (also referred to as a nested UNet) [90], which uses a more compact information transition between the encoder and the decoder by adding upsampled modules and skip connections between corresponding semantic levels. However, while the CD-UNet++ has a single stream structure, the SNUNet encodes each image separately and employs their differences as well.

The Deeply Supervised Image Fusion Network (DSIFN) [44] also has a structure similar to a double-stream UNet with two parallel contracting streams, which extract features from each of the images. The features from each image are then concatenated and upsampled through an expanding stream. Lower level features from the contracting streams are concatenated to the corresponding expanding levels through skip connections.

The authors also employ a difference discrimination network for deep supervision, which combines outputs from the various layers.

Similarly, the End-to-end Superpixel Change Network (ESCNet) [77] features a UNet subnetwork for change detection. The main contribution of this network, however, is in the preprocessing of the images before they are fed to the UNet. Two Siamese superpixel sampling networks are used to extract features and perform superpixel segmentation on the input images. The features are fed to a UNet-like network, and the information about the superpixels is used to reduce the noise and improve edge identification.

The Deep Siamese Multi-scale Fully Convolutional Network (DSMS-FCN) [68] is a fully convolutional network with encoder and decoder parts similar to those of a Siamese UNet, including difference-based skip connections. It introduces a multi-scale feature convolution unit, which is a modified convolution unit that splits the input into several branches and performs convolution with different kernel sizes before joining them again. This unit allows for the extraction of multi-scale features in the same layer. The network also employs change vector analysis to refine the resulting change maps.

The Region-Based Feature Fusion Net (RFNet) [60] aims to reduce the impact of spatially offset bitemporal images, such as due to imperfect co-registration or differences resulting from differing viewing angles. It is a fully convolutional double-stream network with feature interaction modules and region-based feature-fusion modules between the encoders and the decoder. The feature interaction modules fuse features from different scales, similar to dense skip connections in other UNet-like networks, and the region-based feature-fusion modules compare the features with those in their neighbourhoods in order to account for possible spatial offset between the images.

3.3.2. UNets with Attention

Attention blocks [91] have been used increasingly often in the last year, as adding an attention mechanism often leads to considerable improvement in performance. Many of the UNet-based networks employ this strategy. The attention mechanism can be employed at various stages in the model.

As in the case of the aforementioned SNUNet, an attention module can be used as a last step in order to best combine the multiple outputs of the decoder. The Attention-Guided Full-Scale Feature Aggregation Network (AFSNet) [61] resembles a double-stream UNet, with the encoder based on VGG16. Similarly to other UNet variations, its main feature is the enhanced full-scale skip connection to combine features from different scales. The multiple side-outputs, again at various scales, are refined by an attention module combining spatial and channel attention and then fused for the final CD map. However, the attention is most commonly inserted between the encoder and the decoder of the network.

The Multilevel Deformable Attention Aggregated Networks (MLDANets) [82] is a double-stream UNet that uses a single attention module, which receives features from all levels of the Siamese VGG16-based encoder and outputs new improved features into multiple levels of the decoder. The attention module uses deformable sampling with a learnable shape.

Another recent Siamese UNet with attention mechanism is the Siamese Attention + UNet (Siamese_AUNet) [83]. Unlike the MLDANets, which feed features from multiple levels of the encoder into the attention module, this model inserts an attention block as the last step at every level of the encoder. The attention comprises non-local attention (addressing relationships between pixels regardless of their position) as well as channel and local spatial attention. The last step of the two Siamese encoders is an atrous spatial pyramid pooling (ASPP) [92], which uses atrous sampling with various rates to improve the learning of multi-scale features.

The Urban Change Detection Network (UCDNet) [85] also uses a version of spatial pyramid pooling (SPP) [93], namely, new spatial pyramid pooling (NSPP) between the double-stream encoders and the decoder. The model also features modified residual

connections in the encoder. These introduce additional maps of feature differences between the streams at each level of the encoders in order to improve change localization.

The Densely Attentive Refinement Network (DARNet) [84] is another double-stream UNet with attention modules. The DARNet has dense skip connections, where features from various levels are being combined. The hybrid attention module (combining temporal, spatial and channel attention) is inserted at the level of the skip connections. Its output is then fed to the decoder, again, at various levels. The decoder also features a recurrent refinement module with deep supervision inserted at the end of each decoder layer.

The Multi-Scale Residual Siamese Network Fusing Integrated Residual Attention (IRA-MRSNet) [62] resembles a double-stream UNet in its overall structure; however, instead of the typical convolutions, it introduces multi-resolution blocks in order to enhance feature extraction at multiple scales. In essence, these blocks combine convolutions with kernels of different sizes. It also employs an attention unit between the encoder and the decoder.

3.3.3. UNets with Enhanced Boundary Detection

Boundaries and edges of depicted objects are areas that can be difficult to identify and correctly assess, since they tend to be represented by high-frequency features, especially in high-resolution multispectral images. Some networks seek to improve CD performance by addressing edges specifically. Among these are the High Frequency Attention Net (HFA-Net) [86], which is a Siamese UNet; the Multi-scale Attention and Edge-Aware Siamese Network (MAEANet) [56]; and several double-stream networks that are discussed later.

The HFA-Net employs both attention and boundary detection for enhanced change detection. This is achieved by adding a High-Frequency Attention Block (HFAB) at each level of the encoders and the decoder. It consists of a spatial attention part and high-frequency enhancement part. High-frequency enhancement is based on the application of a classical method, namely, the Sobel operator [94].

The MAEANet similarly uses attention and edge detection, namely, it employs Siamese UNets with full encoder–decoder structures for multi-scale feature extraction, to then fuse the features and apply an attention module with spatial and contour attention, followed by an edge-aware module for enhanced edge detection. The MAEANet does not follow the typical structure of double-stream UNets as depicted in Figure 3b. This is because the full UNet is used for feature extraction, rather than only the encoder part. In this case, both images go separately through the whole UNet, encoder and decoder alike, before feature fusion takes place. In this regard, this model can be categorized as a more general double-stream network featuring a UNet, rather than a double-stream UNet.

3.3.4. Non-UNet Double-Stream Models

The general structure of double-stream networks (Figure 3c) allows for a variety of different models. At its base is a Siamese feature extractor followed by feature fusion and a decision-making module. Among these networks, we can list the Spectral-Spatial Joint Learning Network (SSJLN) [65], which uses a CNN to extract features from the images in parallel. The features are then fused, and a fully connected subnetwork is used for comparison and change detection.

There are several networks implementing recurrent neural networks (RNN), usually in the form of long short-term memory (LSTM) [89], such as the Recurrent CNN [63], the SiamCRNN [70] (and the UNetLSTM, which was discussed in the previous section). The idea behind including a RNN in change detection is its ability to handle related sequences of data. In the case of change detection, the sequence usually consists of images from two different time points. The Recurrent CNN and the SiamCRNN resemble each other in overall architecture. They both first use two Siamese CNNs for feature extraction from the bitemporal images. The features are then fed to a recurrent sub-network, which adds a temporal component to the extracted features, and, finally, a decision on binary or multi-class change is made by applying fully-connected layers. The SiamCRNN is also designed to handle bitemporal images from heterogenous sources.

A form of attention was used by a number of UNet-based networks and is also one of the main features of several double-stream networks. Among them, the Dual Attentive Siamese Net (DASNet) [72], which is a fully convolutional network consisting of two Siamese contracting streams (namely, VGG16) followed by a dual attention module, which uses both channel and spatial attention.

Similarly, the Spatial-Temporal Attention Net (STANet) [43] employs Siamese feature-extracting CNNs (in this case, ResNets), followed by an attention module. The Compress–Expand/Expand–Compress Network (CEECNet) [80] introduces not only a new attention module but also a number of modifications that distinguish this network from most others, such as a new loss function, new feature extraction building blocks and a new backbone architecture.

The Super-Resolution-Based Change Detection Network (SRCDNet) [76] is designed to perform change detection on images with different resolutions. Using a generative adversarial network (GAN) super-resolution module, a higher resolution version of the low-resolution image is generated in order to obtain images with the same resolution to be used for change detection. The two images with the same resolution are then fed to ResNet-based feature extractors in parallel. The process of feature extraction is aided by a stacked attention module, which enhances useful information from multiple layers. The final change map is generated by calculating a distance map between the features.

The Capsule Network (CapsNet) [78] was designed to better deal with different viewpoints in bitemporal images and perform well with less training data. Vector-based features are extracted by two pseudo-Siamese capsule networks. The features corresponding to unchanged regions are kept more consistent by an unchanged region reconstruction module, and, finally, a change map is generated by analysing the vector cosine and the vector difference of the image features.

The Dual Learning-Based Siamese Framework for Change Detection (DLSF) [66] is a network that performs two tasks simultaneously. First, it uses a GAN to translate each of the bitemporal images into the domain of the other image in order to suppress irrelevant changes in the images. Each original image and the corresponding translated version of the other image are then input into a Siamese CNN feature extractor. The features are then compared using a pixel-wise Euclidean distance in order to generate the change predictions.

The Trilateral Change Detection Network (TCDNet) [71] is composed of three different CNNs acting as feature extractors in parallel. The main module is based on a ResNet and is an early fusion module into which the images are input together. There are two auxiliary modules that are both Siamese: the difference module and the assimilation module, which focus on changed and unchanged areas, respectively. The network uses dilated convolutions instead of pooling in order to increase the receptive fields without losing information.

The Attention Gates Generative Adversarial Adaptation Network (AG-GAAN) [73] is a GAN-based network with an attention mechanism. The image pair is input into the generator, which generates a change map attempting to deceive a discriminator. The discriminator seeks to distinguish the real change map from the generated one. Attention gates are added to the generator to improve its performance.

Similarly to the HFA-Net described in the section about UNet-based models, there are several more networks employing edge detection to improve their performance For example, the Feature Decomposition–Optimization–Reorganization Network (FDORNet) [81] seeks to improve the detection performance of the edges by decomposing the images into edges and main object bodies. In order to achieve this, the standard initial feature extraction by a Siamese network (here, a ResNet) is followed by a module designed to separate the main body and edge pixels by leveraging convolutions with larger strides to identify low-frequency features corresponding to the main bodies of objects. The features are then optimized using the ground truths and eventually reorganized in order to uncover changes.

Similarly, the Boundary Extraction Constrained Siamese Net (BESNet) [21] focuses on the extraction of edges. The model is composed of a double-stream feature extractor (VGG-based), which runs parallel to a multi-scale boundary extraction stream that uses CNNs with a Sobel operator (as does the HFA-Net) in order to learn gradients and extract object-boundary positions. The outputs from both streams are then fused in order to explore the relationships between various feature maps and the information about boundary positions is used to refine the features extracted by the Siamese encoders. The model also introduces a loss function that takes the gradient into account.

In order to improve edge detection, the Improved Separability Network (ISNet) [87] uses an attention mechanism and a margin maximization strategy to maximize the feature difference between changed and unchanged regions. The model is composed of three parts: a feature extractor, a feature-refinement module and a classification module. First, the images are input into a Siamese feature extractor based on a ResNet backbone with channel-attention modules added at each level. Then, the features from several levels are combined in a feature-refinement module consisting of margin maximization and spatial attention. Margin maximization uses, among other methods, also a deformable convolution. Finally, the last step consists of classification performed on the refined features.

3.3.5. Transformers

Following the success of transformers in the field of natural language processing [91,95], they have been appearing increasingly often within computer vision topics [96], and change detection is not an exception. A number of models based on transformers have been published in the last year (2022). Transformer models are predominantly double-stream, and most of them use transformers together with other elements, such as a UNet, other CNNs, additional attention blocks and more.

The Bitemporal Image Transformer Change Detection (BIT_CD) [79] incorporates a bitemporal image transformer (BIT) after an initial feature extraction. Siamese CNNs (ResNet18s) are first used to extract the features of each image, and the features are then converted into a set of semantic tokens using spatial attention. A transformer is applied to the tokens to model contexts in space and time. The resulting tokens are projected back to pixel space, and a feature difference image is computed from the two context-rich feature maps. Finally, a CNN is used to make the change predictions.

Similarly, the UNet-like Visual Transformer for CD (UVACD) [54] also uses a double-stream CNN feature extraction, which is then followed by a transformer for feature enhancement, and, finally, a decoder. In the case of the Transformer + UNet CD (TransUNetCD) [53], the transformer module is placed between the Siamese encoders of a double-stream UNet and its decoder.

The Swin Transformer Siamese U-shaped Net (SwinSUNet) [97] and Fully Transformer Net (FTN) [57] both use Swin transformers [98], and both can be described as pure transformer models. SwinSUNet uses Swin transformer blocks to encode, fuse and decode features of the bitemporal images. FTN similarly uses Swin transformers for feature extraction at various levels, followed by feature enhancement and change prediction. An attention module is added, and, finally, multiple loss functions are combined for improved results.

The ChangeFormer [55] consists of a double-stream transformer-based encoder, four feature difference modules and a lightweight fully connected CNN decoder.

The Pyramid Semantic CD Transformer (Pyramid-SCDFormer) [52] is designed for semantic CD. It features a double-stream pyramid transformer encoder based on shunted self-attention designed to more efficiently capture multi-level features, followed by a lightweight decoder.

Lastly, the Multi-Scale CNN Transformer Net (MCTNet) [84] uses a combination of transformers and CNN within a single block. The overall structure of this model resembles a double-stream UNet with its Siamese encoders connected by skip connections to a decoder. However, each level of the MCTNet features a ConvTrans block—a block that combines transformer modules and CNN layers.

4. Semi-Supervised and Unsupervised Deep-Learning Models for Multispectral Change Detection

One of the biggest challenges within deep learning for change detection is the availability of large annotated datasets. A possible way to deal with this issue lies in reducing or eliminating the need for the annotations, since unannotated data are abundant. Semi-supervised networks aim to reduce the amount of annotations needed, while unsupervised networks do not require any annotated data at all.

It can often be easier to achieve good results on end-to-end trained supervised networks compared to semi- and unsupervised networks. However, the availability, or rather unavailability, of annotated training data makes semi- and unsupervised networks very attractive. Table 3 provides an overview of selected unsupervised and semi-supervised change-detection networks.

Table 3. Overview of selected unsupervised and semi-supervised change-detection models published between 2017 and 2021.

Unsupervised				
Network Name		**Type**		**Year**
VGG16_LR [99]	VGG16 Low Rank	pretraining	superpixels, VGG16 on scene class., low rank decomp	2017
GDCN [100]	Generative Discriminatory Classified Network	generates training set	GAN, CVA for training set	2019
DCVA [101]	Deep Change Vector Analysis	pretraining	CNN on scene class., deep CVA	2019
DSMS-CN [68]	Deep Siamese Multi-Scale	generates training set	CVA for training set, Double-Stream UNet	2019
S^2-cGAN [102]	Self Supervised Conditional GAN	generates training set	trained on no change, GAN, Generator (UNet)	2020
KPCAMNet [103]	Kernel Principal Component Analysis Network	unsupervised	layerwise training of KPCA modules	2021
Semi-supervised				
FDCNN [104]	Feature Difference CNN	pretraining	VGG16 pretrained on RS scene	2020
Self-supervised Pre-training [105]		pretraining	pretrained on a pretext task	2021
SemiCDNet [47]	Semi-Supervised Change Detection Network	semi-supervised	GAN, Generator (UNet) + 2x Discriminator	2020
IAug_CDNet [106]	Instance-Level Augmentation CD Net	semi-supervised, augmentation	GAN	2021
GCN [107]	Graph Convolutional Network	semi-supervised	graph conv net	2021

4.1. Unsupervised

Most unsupervised change-detection networks can be divided into two categories based on their structure. The first type of structure is depicted in Figure 5a and relies, in essence, on transfer learning. A double-stream architecture provides a natural way of using transfer learning. The parallel subnetworks that fulfill the role of feature extraction can be pretrained on tasks other than change detection. The feature extraction step is then followed by an automated algorithm that uses the features to make a decision about changed areas in the images without the need for additional training. The task used for pretraining is often related to the end goal of the method, such as using RS scene classification to pretrain the feature extractors.

This approach is employed both by the VGG16_LR [99] and by the Deep Change Vector Analysis (DCVA) Network [101]. VGG16_LR first uses Simple Linear Iterative Clustering (SLIC) [108]—a superpixel model—to segment the image into meaningful superpixels. The feature vectors for these segments are then extracted via Siamese VGG16s finetuned on aerial classification images, and a feature difference vector is calculated as an absolute difference of the segment feature vectors of each image. A low-rank decomposition and a simple thresholding are then used to create a binary change map. While the feature extractors require pretraining on labelled images, the rest of the network is unsupervised.

This method, thus, allows the use of data annotated for classification (which is an easier task than annotating pixel-wise change detection) to be used for pixel-based change detection without any need for pixel-annotated change data.

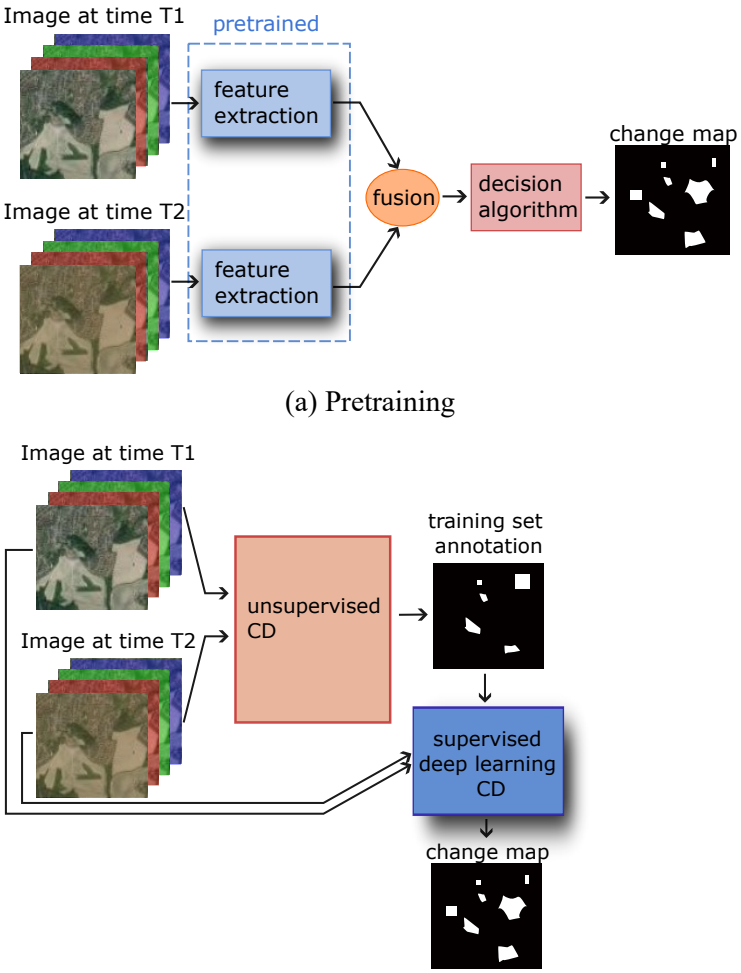

Figure 5. Two possible unsupervised network structures. (**a**) Network with pretrained feature extractors. Siamese feature extractors that have been pretrained on a related task are used to automatically extract image features. The features are then input into an automatic decision algorithm that compares them and outputs a change map. (**b**) Using an unsupervised classical change-detection method to create a labelled training dataset for a supervised deep-learning-based CD model.

The DCVA Network [101] has a similar overall structure, in that it is composed of a pretrained CNN for feature extraction and a more classical decision algorithm. It consists of a feature extraction using a CNN pretrained on classification of aerial images, followed by an automatic feature selection. The feature vectors are then compared in order to obtain a change vector, which can be analysed using deep change vector analysis—a procedure similar to traditional change vector analysis. DCVA is used to identify the pixels with markedly changed features and produce a change map.

Another method often used to achieve unsupervised change detection is automated training-set generation (Figure 5b). In this case, a method that does not require annotations is used to (often partially) annotate the data. The CVA is a commonly used algorithm for this purpose, which provides initial classifications for image pixels, dividing them into changed, unchanged and undecided categories. This auto-generated partially annotated training set is then used to train a deep-learning-based supervised classifier, which outputs a change map. In many cases, GANs are used as part of the method to generate annotations.

The Deep Siamese Multi-Scale Convolutional Net (DSMS-CN) [68] is an unsupervised version of DSMS-FCN discussed previously. As with DSMS-FCN, it uses the multi-scale feature convolution unit that splits the input and performs convolution with varying kernel sizes. It is a fully unsupervised network with no pretraining, which uses classical change-detection methods to generate annotations and, thus, create a training set. Suitable training samples are generated by an automatic preclassification algorithm consisting of a change vector analysis (CVA) and a fuzzy c-means clustering (FCM). Using preclassification, the pixels are divided into three groups: nearly certainly changed w_c, nearly certainly unchanged w_u and undetermined w_{tbd}. Then, small patches around w_c and w_u pixels are used as training samples for a UNet-like fully convolutional neural network with a Siamese encoder, which then learns to predict the change probability of the central pixel in the patch.

The Generative Discriminatory Classified Network (GDCN) [100] is an unsupervised model taking advantage of generative adversarial networks. The model consists of two subnetworks: a generator and a discriminatory classified network (DCN). Unlabelled input data pixels are first classified into three categories: changed, unchanged and no-label. This process is performed using automatic methods (namely, CVA followed by Otsu [109]) to label all pixels as changed and unchanged as well as the application of a neighbourhood-based criterion to select the pixels where this automatic CD is sufficiently reliable and those that should remain unlabelled. The automatically labelled data, the unlabelled data and a set of fake data created by the generator are then input into DCN, which has the task of distinguishing between the changed, unchanged and fake pixels.

The Self-Supervised Conditional Generative Adversarial Network (S^2-cGAN) [102] is composed of a generator and a discriminator. For training, it requires pairs of images with no relevant changes. The authors of the paper have, in this case, created a second image by adding noise to the first one. During training, the generator (UNet-based) is fed one of the images and trained to output a similar image with no relevant ground changes. The discriminator learns to distinguish pairs of unchanged pixels (i.e., from the real images) from pairs of changed pixels (one of them generated by the generator). Effectively, the generator learns to produce unchanged images, while the discriminator is trained to recognize changed/fake pixels. During inference, both the generator and the discriminator are used: change detection is based both on the reconstruction error of the generator and on the pixel-based out-of-distribution likelihood produced by the discriminator.

The Kernel Principal Component Analysis—Mapping Network (KPCA-MNet) [103] is an unuspervised model that introduces a Kernel Principal Component Analysis (KPCA) convolution module. A Siamese deep convolutional network utilising the KPCA convolution first extracts features from the images. The feature difference map is then mapped into a 2D polar domain, and change detection is performed by thresholding segmentation and clustering. The feature extractor network is trained layerwise in an unsupervised manner—random input patches from the previous layer are selected, and the parameters of the KPCA are calculated. KPCA fuses deep learning and classical methods by using a deep CNN architecture with a KPCA-based components, effectively avoiding the need for annotated training data.

4.2. Semi-Supervised

In the case of semi-supervised networks, the need for training data is reduced but not completely eliminated. This can be achieved through various approaches. As with fully unsupervised models, pretraining can be used, such as in the case of the Feature Difference CNN (FDCNN) [104]. This network consists of Siamese CNNs (namely, VGG16s) for feature extraction, followed by a feature difference CNN and a feature fusion CNN.

The contracting VGG16, together with the expanding feature difference CNN, resembles UNet, and all of the networks are fully convolutional. The authors pretrained the VGG16 on scene-level classification of remote-sensing images; thus, the training of the whole network for change detection only involves the last feature fusion CNN. Pretraining,

in this case, allows for fewer pixel-annotated change maps, thereby taking advantage of the easier obtainable scene classifications.

The method by Leenstra et al. [105] uses self-supervised pretraining in order to improve change detection in Sentinel-2 images, namely, the model is pretrained on a pretext task that does not require supervision but is designed to teach the model to recognize features important for change detection. One such pretext task is deciding whether two image patches from bitemporal images are spatially overlapping. The model can then be further trained on a smaller annotated dataset.

Another possibility to reduce the amount of necessary annotated data is to use a generative adversarial network. The SemiCDNet [47] is a semi-supervised GAN-based convolutional network designed for change detection. It consists of one generator G and two discriminators D_s and D_e and is trained using a combination of labelled and unlabelled data. G is a UNet++ segmentation network that outputs predicted change/no-change labels for the input images. The segmentation discriminator D_s is fed both ground truths and the predictions output by the generator and its goal is to distinguish between the two. During testing, this leads to the generated labels resembling the ground truths more. To further improve the results, the entropy discriminator D_e encourages the entropy maps of the generator's predictions from unlabelled data to resemble its predictions from labelled data. The two discriminators are trained alternately during the training process.

The Instance-Level Augmentation CD Net (IAug_CDNet) [106] also uses a GAN, in this case, however, for image augmentation, which precedes the change detection. A GAN is used to add buildings into the images and, thus, supply additional labelled data. The change detection part of IAug_CDNet is based on a UNet, although, in this case, the UNet sub-network is, strictly speaking, used to extract the features from the two images in parallel and another sub-network (here, a shallow fully convolutional neural network) then uses the difference of these features to predict the changed pixels.

A different approach can be seen in the Graph Convolutional Network (GCN) [107], which takes advantage of some labelled data within an unlabelled dataset by employing a graph convolutional network. The Graph CNN encodes multitemporal images as a graph via multi-scale parcel segmentation capturing the features of the images. The information from the labelled data is then propagated to the unlabelled data over training iterations.

5. Performance

Several performance metrics can be used to evaluate the results of a change detection process. Change detection is, in essence, a classification task, classifying each pixel into categories of changed/unchanged. If we designate change as a "positive" and no change as "negative", there are four possible outcomes: true positive (TP) being a correctly identified change/positive, true negative (TN) being a correctly identified no change, false positive (FP) being an unchanged pixel labelled as changed and false negative (FN) a changed pixel labelled as unchanged.

The simplest metric of performance is the overall accuracy (OA), which quantifies the fraction of correctly labelled pixels:

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

For the task of change detection, however, OA can be misleading due to the large class imbalance. The rarity of changes means that a model can achieve a high OA by simply labelling all pixels as unchanged.

It is, therefore, more common to employ other metrics that provide a more nuanced insight into the performance of the model. The most used are precision (or positive predictive value (PPV)) and recall (also called sensitivity or true positive rate (TPR)):

$$\mathrm{precision} = \mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$\text{recall} = \text{sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Another commonly used metric is the F1 score, defined as:

$$\text{F1} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

Table 4 shows an overview of the performance metric values reported for models featured in this review. The precision, recall, F1 and OA are listed, and the models are grouped by the datasets they were evaluated on. Many models were tested on one or several of the reported open datasets to allow for comparison.

Table 4. Performance values of a selection of models grouped by the dataset used for evaluation. Semi-supervised models are denoted with *. The **best** results for each dataset are shown in **bold**, the second best are underlined, and the (third best) results are enclosed in parentheses (). For some semi-supervised models, the percentage of the dataset used for training is also stated.

Network Name	**Precision (%)**	**Recall** (%)	**F1 (%)**	**OA (%)**
		SZTAKI–Szada		
FC-EF	43.57	62.65	51.4	(93.08)
FC-Siam-conc	40.93	(65.61)	50.41	92.46
FC-Siam-diff	41.38	**72.38**	52.66	92.4
DSMS-FCN	**52.78**	63.39	**57.72**	**94.57**
STANet	(45.5)	63.5	53.0	
ESCNet	48.89	58.21	(53.73)	94.07
CapsNet	44.4	68.9	54.0	
* FDCNN		56.05		92.86
		SZTAKI–Tiszadob		
FC-EF	(90.28)	96.74	93.4	**97.66**
FC-Siam-conc	72.07	**96.87**	82.65	93.04
FC-Siam-diff	69.51	88.29	77.78	91.37
DSMS-FCN	89.18	88.56	88.86	96.20
STANet	95.0	90.8	(93.0)	
ESCNet	76.33	72.87	74.56	(93.95)
CapsNet	**96.8**	(95.3)	**96.0**	
		OSCD		
FC-EF	64.42	50.97	56.91	96.05
FC-Siam-conc	42.39	65.15	51.36	93.68
FC-Siam-diff	57.8	57.99	(57.92)	95.68
FC-EF-Res	54.93	66.48	60.15	95.64
UNetLSTM	(63.59)	52.93	57.78	(96.00)
* FDCNN		(65.47)		91.17
UCDNet	**92.53**	**86.16**	**89.21**	**99.30**
SwinSUNet	55.0	54.0	54.5	95.3
		CDD		
CD-UNet++	89.54	87.11	87.56	96.73
DSIFN	94.96	86.08	90.30	97.71
DASNet	92.2	93.2	92.7	98.2
SNUNet-CD	96.3	96.2	96.2	
CLNet	94.7	89.7	92.1	98.1
SRCDNet	92.07	88.07	90.02	
ESCNet	90.90	(96.35)	93.54	98.47
ISNet	95.18	94.43	94.80	98.78
BESNet	95.20	92.40	93.78	98.51
DARNet	97.05	96.91	96.98	**99.29**
SwinSUNet	95.7	92.3	94.0	98.5
TransUNetCD	(96.93)	**97.42**	**97.17**	
MCTNet	96.56	95.33	95.94	(99.05)
ASFNet	**98.44**	92.85	95.56	98.94
IRA-MRSNet	96.81	96.13	(96.47)	99.14

Table 4. *Cont.*

Network Name	Precision (%)	Recall (%)	F1 (%)	OA (%)
		LEVIR		
STANet	83.8	(91.0)	87.3	
CLNet	89.8	90.3	90.0	98.9
BIT_CD	89.24	89.37	89.31	98.92
CEECNet	93.81	89.92	91.83	
* IAug_CDNet 20%	90.1	85.1	87.5	
* IAug_CDNet 100%	91.6	86.5	89	
ISNet	92.46	88.27	90.32	99.04
HFA-Net			88.32	98.90
BESNet	**94.41**	84.26	89.05	97.69
SiameseAUNet	85.82	87.02	85.57	
DARNet	92.67	**91.31**	**91.98**	97.76
MLDANets	(93.08)	90.18	(91.57)	**99.15**
FDORNet	91.29	90.42	90.85	99.07
TransUNetCD	92.43	89.82	91.11	
UVACD	91.90	90.70	91.30	99.12
ChangeFormer	92.05	88.80	90.40	99.04
Pyramid-SCDFormer-B	92.72	90.18	91.41	98.39
MAEANet	88.84	(91.00)	89.90	89.35
FTN	92.71	89.37	91.01	99.06
MCTNet	91.21	90.76	90.98	(99.08)
MFATNet	91.85	88.93	90.36	99.03
ASFNet	90.74	91.06	90.90	99.07
IRA-MRSNet	84.81	89.37	86.23	98.74
		DSIFN		
DSIFN	67.11	67.54	67.33	88.86
BIT_CD	68.36	70.18	69.26	89.41
BESNet	(83.60)	(72.17)	(77.47)	**97.98**
TransUNetCD	71.55	69.42	66.62	
ChangeFormer	88.48	84.94	86.67	(95.56)
MFATNet	**88.65**	**86.62**	**87.62**	95.84
		WHU		
CLNet	**96.9**	**95.7**	**96.3**	**99.7**
CEECNet	(95.57)	(92.04)	(93.77)	
* IAug_CDNet 20%	86.8	78.1	82.2	
* IAug_CDNet 100%	91.4	86.9	89.1	
* SemiCDNet 5%			82.90	94.34
* SemiCDNet 10%			85.28	95.17
* SemiCDNet 20%			86.57	95.59
* SemiCDNet 50%			87.74	95.95
HFA-Net			88.23	97.58
SiameseAUNet	82.02	86.33	84.47	
SwinSUNet	95.0	92.6	93.8	99.4
TransUNetCD	93.59	89.60	93.59	
UVACD	94.59	91.17	92.84	99.14
Pyramid-SCDFormer-B	92.22	86.86	89.31	96.43
MAEANet	92.82	90.38	91.56	99.36
FTN	93.09	91.24	92.16	(99.37)
MFATNet	93.18	83.93	88.31	99.01
RFNet	95.72	89.46	92.49	
IRAM-MRSNet	84.07	85.18	84.52	98.63
		SYSU		
ISNet	80.27	(76.41)	78.29	90.01
DARNet	(83.04)	**79.11**	81.03	91.26
FTN	**86.86**	76.82	**81.53**	**97.79**
IRA-MRSNet	85.39	75.20	(79.98)	(90.85)
		Google		
HFA-Net			82.77	96.47
FTN	86.99	84.21	**85.58**	**97.92**

All of the values featured in the three tables are taken from articles written by the originators of each given model. It should be noted that this means that the training

conditions for the various models can and do differ and the provided metric values can only serve as a rough overview rather than as grounds for comparison. Ideally, in order to provide fair comparison, all of the models should be reimplemented and tested under identical conditions on a range of datasets.

However, this undertaking would be very time-consuming, and its accuracy and utility would be questionable as the code has not been published for all of the models, thus, making the risk of inaccurate and less-than-optimal reimplementations high. Therefore, we chose to compile the metric values as reported by the respective originators of the models in order to provide a general overview. Although reports on reimplementations were published for a number of them, we decided to include only the performance values published by the original authors.

There can be a large difference in the performance of a given model on different datasets, as can be seen from a number of models used on both of the two SZTAKI sub-datasets. The Szada sub-dataset appears to be more challenging than the Tiszadob, and all of the models achieved significantly better results on the Tiszadob. Similar differences in performance can be seen for other models that were evaluated on several datasets, such as the DSIFN, which showed better performance on the CDD dataset compared with on the DSIFN dataset. Where two or more models were evaluated on the same two or more datasets, it is possible that one performed better on one of the datasets, while the other one performed better on the other, such as CLNet being the best on the WHU dataset, while CEECNet performed somewhat better on the LEVIR, even though both of these datasets exclusively focus on buildings.

Almost all of the models, except for DSIFN and BIT_CD on the DSIFN dataset and S^2-cGAN, had an overall accuracy above 90% percent and often above 95%. However, this is less a result of high performance and more an artefact created by the class imbalance present in the datasets. Real relevant changes are rare in remote sensing, and the number of changed pixels is relatively low, which artificially increases the overall accuracy.

The accuracy was over 90% for all of the models evaluated on the SZTAKI–Szada, even when the precision of some of these models was below 50%, meaning that only half of the pixels they labelled as changes were an actual change, and the other half were false positives. To illustrate this further, the HRSCD dataset had as many as 99.232% of pixels labelled as unchanged. Predicting "no change" for every pixel in this dataset would yield an overall accuracy of 99.232%, while both the precision and the recall would be 0.

Semi-supervised models (marked with * in the tables) can achieve good performance with only a fraction of the annotated data, but they tend to perform slightly worse than their supervised counterparts.

6. Challenges and Outlook

Change detection in remote sensing is a useful but demanding task with a unique set of challenges. Despite the large amount of available RS imagery, high-quality large annotated CD datasets are not simple to create. They often need to be annotated pixel-wise by hand, which is time-consuming and laborious. Furthermore, unlike most other applications of deep learning, they require two or more images, which increases the amount of data to acquire, makes the process of annotating more complex and introduces the additional need for coregistration.

The fact that the idea of "change" itself can be defined in more than one way means that there will be large differences between various annotated datasets. This leads to difficulties comparing and evaluating networks if they have not been tested on the same data. It also reduces the transferability of a network trained on one dataset to another dataset. One could argue that the concept of change is so broad that it should not be affected by (minor) differences between datasets; however, in reality, providing an annotation always implicitly chooses the category of changes to be considered as relevant, as well as their context. A network trained on one dataset will, thus, learn a particular type of changes and will not necessarily be able to recognise a type of change that it has not encountered in training.

An additional difficulty is posed by the fact that true changes are fairly rare in RS imagery. This means that the vast majority of pixels within any dataset are unchanged, and this large class imbalance requires a special approach, such as a well-chosen loss function.

The creation of a standard set of open test datasets would significantly facilitate the evaluation and comparison of various models. It is currently difficult to efficiently compare the published models, as comparisons of performance on different datasets are nearly meaningless. Ideally, models should be evaluated on several datasets, as they can exhibit significant differences.

From the presented overview of the reported models, several trends can be identified. Convolutional neural networks with a double-stream architecture are the most commonly used models for supervised change detection. The most recent ones generally include some form of an attention mechanism, which seems to improve the model's performance significantly. It is also apparent that the choice of a loss function is an important factor due to the class imbalance in the change-detection datasets. Among the most recent models are several transformer-based ones, which naturally follows from the success transformers have been having in the fields of natural language processing and computer vision.

In recent years, there has been more focus on unsupervised and semi-supervised networks to eliminate or at least reduce the need for annotated datasets. However, these models generally do not yet achieve the accuracy of their supervised counterparts. Using an unsupervised model also leads to a loss of some control over the type of change to be detected. The main method of teaching the model to distinguish between what we consider relevant and irrelevant, lies in the annotation of the changes as such.

Unsupervised models that rely on automatic algorithms for preliminary annotations, but also those that rely on pretraining on other types of data, give up this ability to fine-tune the types of changes to be considered relevant. This minor loss of control over the type of change is traded for the ability to train on much larger amounts of data with much less upfront effort, which is the main advantage of unsupervised models. Further development in the direction of unsupervised as well as semi-supervised models can be expected.

Deep-learning models have achieved great results in the application of change detection and, in the majority of cases, have surpassed more classical methods. This is largely due to their ability to model complex relationships between pixels within the images, thus, taking into account sufficient context to distinguish apparent and irrelevant changes from changes of interest. This is particularly important for high-resolution images and images of complex, varying landscapes.

Mountain areas, for instance, represent a large part of the Earth's surface but present unique challenges due to their high topographic variability, steep slopes, line-of-sight challenges, snow cover of varying depths and difficulty establishing ground truth [110–112]. Deep learning has the potential to address some of these challenges, as it is well suited for dealing with complex landscapes; however, in most cases, it is reliant on high-quality abundant annotations. The need to monitor areas that are difficult to access and for which it is difficult to establish ground truth once again emphasizes the utility of semi- and unsupervised deep-learning models.

The development of change-detection methods for remote sensing goes hand-in-hand with the development of the technology used for remote sensing. The amount of available RS imagery is increasing every year as new and improved airborne and space-borne platforms are being deployed. Satellites, in particular, are a source of large amounts of data, due to their frequent revisit times and increasingly higher spatial resolutions. They provide a means for monitoring hard-to-access areas as well as for observing changes regularly over long periods of time.

In fact, the increasing availability of satellite images with frequent revisit times and improved spatial resolution opens up the use of longer time-series, rather than focusing on two-image change detection [113]. Time-series, by definition, include more information and present a very interesting platform for change detection, which has been thus far comparatively little explored. The facts that true changes are rare and that the revisit times

of many satellites are short opens up possibilities for the development of self-supervised models, exploiting the readily available time-series of images.

Another possible avenue for future research is semantic change detection, where not only is the presence of a change detected but also the type of change is classified. Preparing ground truth including this information is more challenging than simply focusing on change/no-change, but some such datasets have already been published, such as the the HRSCD and the SECOND datasets mentioned in Section 2. Models focusing on semantic change detection are still rare compared to binary change detection. Change detection is a complex and multifaceted problem to the point that one could say it is many problems with some common characteristics. The choice of the right approach will heavily depend on the type of data to be analysed and on the goal of the analysis, as the performance of models will somewhat differ from dataset to dataset and task to task.

While the recent years have seen an unprecedented growth in the interest for change detection in remote sensing, the challenge is far from solved.

7. Conclusions

Change detection lies at the core of many practical applications of remote sensing. In recent years, deep learning has been increasingly used to address this task.

This review provides an overview of deep-learning methods applied to change detection specifically within multispectral remote-sensing imagery. It includes a discussion of currently available open datasets that are suited for change detection and an analysis of a selection of recent deep-learning models along with their performance. The models are categorized into supervised, unsupervised and semi-supervised, and their general structures are described. The article also provides a discussion on the unique challenges of change detection.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No original data were reported in this review. All data discussed can be found from their respective sources. Links to the discussed datasets are also provided in Appendix A.

Acknowledgments: The author would like to thank everyone who contributed by reading the manuscript and providing valuable comments and discussion, namely, Mathias Bynke, Eirik Anette Flynn Opland and especially Ingebjørg Kåsen and Simen Ellingsen Rustad, without whom the article would not have been the same.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RS	Remote sensing
CD	Change detection
CVA	Change vector analysis
RGB	Red, green, blue
SAR	Synthetic aperture radar
PCA	Principal component analysis
MAD	Mulitvariate alteration detection
LSTM	Long short-term memory
CNN	Convolutional neural network
RNN	Recurrent neural network
GAN	Generative adversarial network
TP	True positive

TN	True negative
FP	False positive
FN	False negative
OA	Overall accuracy

Appendix A

Table A1. Links to the reviewed datasets.

SZTAKI	http://mplab.sztaki.hu/remotesensing/airchange_benchmark.html (accessed on 11 April 2023)
OSCD	https://rcdaudt.github.io/oscd/ (accessed on 11 April 2023)
CDD	https://drive.google.com/file/d/1GX656JqqOyBi_Ef0w65kDGVto-nHrNs9/edit (accessed on 11 April 2023)
WHU Building CD	https://study.rsgis.whu.edu.cn/pages/download/building_dataset.html (accessed on 11 April 2023)
HRSCD	https://rcdaudt.github.io/hrscd/ (accessed on 11 April 2023)
LEVIR-CD	https://justchenhao.github.io/LEVIR/ (accessed on 11 April 2023)
DSIFN	https://github.com/GeoZcx/A-deeply-supervised-image-fusion-network-for-change-detection-in-remote-sensing-images/tree/master/dataset (accessed on 11 April 2023)
MtS-WH	https://github.com/rulixiang/MtS-WH-dataset (accessed on 11 April 2023)
Google Data Set	https://github.com/daifeng2016/Change-Detection-dataset-for-High-Resolution-Satellite-Imagery (accessed on 11 April 2023)
SYSU-CD	https://github.com/liumency/SYSU-CD (accessed on 11 April 2023)
SECOND	http://www.captain-whu.com/project/SCD/ (accessed on 11 April 2023)
3DCD	https://bit.ly/3wDdo41 (accessed on 11 April 2023)
Landsat-SCD	https://doi.org/10.6084/m9.figshare.19946135.v1 (accessed on 11 April 2023)

References

1. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-Of-The-Art Review. *Remote Sens.* **2020**, *12*, 1444. [CrossRef]
2. Hu, T.; Su, Y.; Xue, B.; Liu, J.; Zhao, X.; Fang, J.; Guo, Q. Mapping Global Forest Aboveground Biomass with Spaceborne LiDAR, Optical Imagery, and Forest Inventory Data. *Remote Sens.* **2016**, *8*, 565. [CrossRef]
3. Xie, Q.; Dash, J.; Huete, A.; Jiang, A.; Yin, G.; Ding, Y.; Peng, D.; Hall, C.C.; Brown, L.; Shi, Y.; et al. Retrieval of Crop Biophysical Parameters from Sentinel-2 Remote Sensing Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 187–195. [CrossRef]
4. Xu, W.; Vinã, A.; Kong, L.; Pimm, S.L.; Zhang, J.; Yang, W.; Xiao, Y.; Zhang, L.; Chen, X.; Liu, J.; et al. Reassessing the Conservation Status of the Giant Panda Using Remote Sensing. *Nat. Ecol. Evol.* **2017**, *1*, 1635–1638. [CrossRef] [PubMed]
5. Stapleton, S.; LaRue, M.; Lecomte, N.; Atkinson, S.; Garshelis, D.; Porter, C.; Atwood, T. Polar Bears from Space: Assessing Satellite Imagery as a Tool to Track Arctic Wildlife. *PLoS ONE* **2014**, *9*, e101513. [CrossRef] [PubMed]
6. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4909–4918. [CrossRef]
7. Zhang, E.; Liu, L.; Huang, L.; Ng, K.S. An Automated, Generalized, Deep-Learning-Based Method for Delineating the Calving Fronts of Greenland Glaciers from Multi-Sensor Remote Sensing Imagery. *Remote Sens. Environ.* **2021**, *254*, 112265. [CrossRef]
8. Browning, D.M.; Steele, C.M. Vegetation Index Differencing for Broad-Scale Assessment of Productivity Under Prolonged Drought and Sequential High Rainfall Conditions. *Remote Sens.* **2013**, *5*, 327–341. [CrossRef]
9. Ghaffarian, S.; Kerle, N.; Pasolli, E.; Arsanjani, J.J. Post-Disaster Building Database Updating Using Automated Deep Learning: An Integration of Pre-Disaster OpenStreetMap and Multi-Temporal Satellite Data. *Remote Sens.* **2019**, *11*, 2427. [CrossRef]
10. Malakhov, D.; Dyke, G.; King, C. Remote Sensing Applied to Paleontology: Exploration of Upper Cretaceous Sediments in Kazakhstan for Potential Fossil Sites. *Palaeontol. Electron.* **2009**, *12*, 1935–3952.
11. Emerson, C.; Bommersbach, B.; Nachman, B.; Anemone, R. An Object-Oriented Approach to Extracting Productive Fossil Localities from Remotely Sensed Imagery. *Remote Sens.* **2015**, *7*, 16555–16570. [CrossRef]
12. Silván-Cárdenas, J.L.; Caccavari-Garza, A.; Quinto-Sánchez, M.E.; Madrigal-Gómez, J.M.; Coronado-Juárez, E.; Quiroz-Suarez, D. Assessing Optical Remote Sensing for Grave Detection. *Forensic Sci. Int.* **2021**, *329*, 111064. [CrossRef] [PubMed]
13. Wellmann, T.; Lausch, A.; Andersson, E.; Knapp, S.; Cortinovis, C.; Jache, J.; Scheuer, S.; Kremer, P.; Mascarenhas, A.; Kraemer, R.; et al. Remote Sensing in Urban Planning: Contributions towards Ecologically Sound Policies? *Landsc. Urban Plan.* **2020**, *204*, 103921. [CrossRef]
14. Singh, A. Digital Change Detection Techniques Using Remotely-Sensed Data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]

15. Su, H.; Wu, Z.; Zhang, H.; Du, Q. Hyperspectral Anomaly Detection: A survey. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 64–90. [CrossRef]
16. Liu, S.; Marinelli, D.; Bruzzone, L.; Bovolo, F. A Review of Change Detection in Multitemporal Hyperspectral Images: Current Techniques, Applications, and Challenges. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 140–158. [CrossRef]
17. Elmaizi, A.; Sarhrouni, E.; Hammouch, A.; Chafik, N. Hyperspectral Images Classification and Dimensionality Reduction using spectral interaction and SVM classifier. *arXiv* **2022**, arXiv:2210.15546.
18. Shen, S.S.; Bassett, E.M. Information-Theory-Based Band Selection and Utility Evaluation for Reflective Spectral Systems. *SPIE* **2002**, *4725*, 18–29. [CrossRef]
19. Kåsen, I.; Rødningsby, A.; Haavardsholm, T.V.; Skauli, T. Band Selection for Hyperspectral Target Detection Based on a Multinormal Mixture Anomaly Detection Algorithm. *SPIE* **2008**, *6966*, 53–58. [CrossRef]
20. Wang, S.; Quan, D.; Liang, X.; Ning, M.; Guo, Y.; Jiao, L. A deep learning framework for remote sensing image registration. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 148–164. [CrossRef]
21. Lei, J.; Gu, Y.; Xie, W.; Li, Y.; Du, Q. Boundary Extraction Constrained Siamese Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5621613. [CrossRef]
22. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Change Detection Techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2401. [CrossRef]
23. Vu, P.X.; Duc, N.T.; Yem, V.V. Application of Statistical Models for Change Detection in SAR Imagery. In Proceedings of the 2015 International Conference on Computing, Management and Telecommunications, ComManTel 2015, Da Nang, Vietnam, 28–30 December 2015; pp. 239–244. [CrossRef]
24. Zhao, J.; Chang, Y.; Yang, J.; Niu, Y.; Lu, Z.; Li, P. A Novel Change Detection Method Based on Statistical Distribution Characteristics Using Multi-Temporal PolSAR Data. *Sensors* **2020**, *20*, 1508. [CrossRef] [PubMed]
25. Zhang, C.; Wei, S.; Ji, S.; Lu, M. Detecting Large-Scale Urban Land Cover Changes from Very High Resolution Remote Sensing Images Using CNN-Based Classification. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 189. [CrossRef]
26. Bhandari, A.; Kumar, A.; Singh, G. Feature Extraction using Normalized Difference Vegetation Index (NDVI): A Case Study of Jabalpur City. *Procedia Technol.* **2012**, *6*, 612–621. [CrossRef]
27. 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]
28. 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]
29. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep Learning-Based Change Detection in Remote Sensing Images: A Review. *Remote Sens.* **2022**, *14*, 871. [CrossRef]
30. Jiang, H.; Peng, M.; Zhong, Y.; Xie, H.; Hao, Z.; Lin, J.; Ma, X.; Hu, X. A Survey on Deep Learning-Based Change Detection from High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 1552. [CrossRef]
31. 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]
32. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
33. Zhang, L.; Zhang, L.; Du, B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
34. Zhong, Y.; Ma, A.; Ong, Y.s.; Zhu, Z.; Zhang, L. Computational Intelligence in Optical Remote Sensing Image Processing. *Appl. Soft Comput.* **2018**, *64*, 75–93. [CrossRef]
35. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
36. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools, and Challenges for the Community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
37. 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]
38. 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. [CrossRef]
39. Daudt, R.C.; Le Saux, B.; Boulch, A.; Gousseau, Y. Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2115–2118.
40. 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. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII*, 565–571. [CrossRef]
41. 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]
42. Caye Daudt, R.; Le Saux, B.; Boulch, A.; Gousseau, Y. Multitask Learning for Large-Scale Semantic Change Detection. *Comput. Vis. Image Underst.* **2019**, *187*, 102783. [CrossRef]
43. Chen, H.; Shi, Z. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection. *Remote Sens.* **2020**, *12*, 1662. [CrossRef]

44. 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]
45. Wu, C.; Zhang, L.; Zhang, L. A Scene Change Detection Framework for Multi-Temporal Very High Resolution Remote Sensing Images. *Signal Process.* **2016**, *124*, 184–197. [CrossRef]
46. Wu, C.; Zhang, L.; Du, B. Kernel Slow Feature Analysis for Scene Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2367–2384. [CrossRef]
47. 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]
48. 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.* **2021**, *60*, 5604816. [CrossRef]
49. Yang, K.; Xia, G.S.; Liu, Z.; Du, B.; Yang, W.; Pelillo, M.; Zhang, L. Semantic Change Detection with Asymmetric Siamese Networks. *arXiv* **2020**, arXiv:2010.05687. [CrossRef]
50. Coletta, V.; Marsocci, V.; Ravanelli, R. 3DCD: A New Dataset for 2D and 3D Change Detection Using Deep Learning Techniques. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *XLIII*, 1349–1354. [CrossRef]
51. Tian, S.; Zhong, Y.; Zheng, Z.; Ma, A.; Tan, X.; Zhang, L. Large-Scale Deep Learning Based Binary and Semantic Change Detection in Ultra High Resolution Remote Sensing Imagery: From Benchmark Datasets to Urban Application. *ISPRS J. Photogramm. Remote Sens.* **2022**, *193*, 164–186. [CrossRef]
52. Yuan, P.; Zhao, Q.; Zhao, X.; Wang, X.; Long, X.; Zheng, Y. A Transformer-Based Siamese Network and an Open Optical Dataset for Semantic Change Detection of Remote Sensing Images. *Int. J. Digit. Earth* **2022**, *15*, 1506–1525. [CrossRef]
53. Li, Q.; Zhong, R.; Du, X.; Du, Y. TransUNetCD: A Hybrid Transformer Network for Change Detection in Optical Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5622519. [CrossRef]
54. Wang, G.; Li, B.; Zhang, T.; Zhang, S. A Network Combining a Transformer and a Convolutional Neural Network for Remote Sensing Image Change Detection. *Remote Sens.* **2022**, *14*, 2228. [CrossRef]
55. Bandara, W.G.C.; Patel, V.M. A Transformer-Based Siamese Network for Change Detection. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Virtual, 17–22 July 2022; pp. 207–210. [CrossRef]
56. Yang, B.; Huang, Y.; Su, X.; Guo, H. MAEANet: Multiscale Attention and Edge-Aware Siamese Network for Building Change Detection in High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 4895. [CrossRef]
57. Yan, T.; Wan, Z.; Zhang, P. Fully Transformer Network for Change Detection of Remote Sensing Images. *arXiv* **2022**, arXiv:2210.00757.
58. Li, W.; Xue, L.; Wang, X.; Li, G. MCTNet: A Multi-Scale CNN-Transformer Network for Change Detection in Optical Remote Sensing Images. *arXiv* **2022**, arXiv:2210.07601. [CrossRef]
59. Mao, Z.; Tong, X.; Luo, Z.; Zhang, H. MFATNet: Multi-Scale Feature Aggregation via Transformer for Remote Sensing Image Change Detection. *Remote Sens.* **2022**, *14*, 5379. [CrossRef]
60. Chen, P.; Li, C.; Zhang, B.; Chen, Z.; Yang, X.; Lu, K.; Zhuang, L. A Region-Based Feature Fusion Network for VHR Image Change Detection. *Remote Sens.* **2022**, *14*, 5577. [CrossRef]
61. Jiang, M.; Zhang, X.; Sun, Y.; Feng, W.; Gan, Q.; Ruan, Y. AFSNet: Attention-guided full-scale feature aggregation network for high-resolution remote sensing image change detection. *Gisci. Remote Sens.* **2022**, *59*, 1882–1900. [CrossRef]
62. Ling, J.; Hu, L.; Cheng, L.; Chen, M.; Yang, X. IRA-MRSNet: A Network Model for Change Detection in High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5598. [CrossRef]
63. Mou, L.; Zhu, X.X. A Recurrent Convolutional Neural Network for Land Cover Change Detection in Multispectral Images. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 4363–4366. [CrossRef]
64. Caye Daudt, R.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. In Proceedings of the International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067. [CrossRef]
65. Zhang, W.; Lu, X. The Spectral-Spatial Joint Learning for Change Detection in Multispectral Imagery. *Remote Sens.* **2019**, *11*, 240. [CrossRef]
66. Fang, B.; Pan, L.; Kou, R. Dual Learning-Based Siamese Framework for Change Detection Using Bi-Temporal VHR Optical Remote Sensing Images. *Remote Sens.* **2019**, *11*, 1292. [CrossRef]
67. Peng, D.; Zhang, Y.; Guan, H. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [CrossRef]
68. Chen, H.; Member, S.; Wu, C.; Du, B.; Member, S.; Zhang, L. Change Detection in Multi-temporal VHR Images Based on Deep Siamese Multi-scale Convolutional Networks. *arXiv* **2019**, arXiv:1906.11479. [CrossRef].
69. Papadomanolaki, M.; Verma, S.; Vakalopoulou, M.; Gupta, S.; Karantzalos, K. Detecting Urban Changes with Recurrent Neural Networks from Multitemporal Sentinel-2 Data. *arXiv* **2019**, arXiv:1910.07778.
70. 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]
71. Qian, J.; Xia, M.; Zhang, Y.; Liu, J.; Xu, Y. TCDNet: Trilateral Change Detection Network for Google Earth Image. *Remote Sens.* **2020**, *12*, 2669. [CrossRef]

72. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [CrossRef]
73. Zhao, W.; Chen, X.; Ge, X.; Chen, J. Using Adversarial Network for Multiple Change Detection in Bitemporal Remote Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8003605. [CrossRef]
74. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8007805. [CrossRef]
75. 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]
76. Liu, M.; Shi, Q.; Marinoni, A.; He, D.; Liu, X.; Zhang, L. Super-Resolution-Based Change Detection Network with Stacked Attention Module for Images with Different Resolutions. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4403718. [CrossRef]
77. Zhang, H.; Lin, M.; Yang, G.; Zhang, L. ESCNet: An End-to-End Superpixel-Enhanced Change Detection Network for Very-High-Resolution Remote Sensing Images. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 28–42. [CrossRef] [PubMed]
78. Xu, Q.; Chen, K.; Zhou, G.; Sun, X. Change Capsule Network for Optical Remote Sensing Image Change Detection. *Remote Sens.* **2021**, *13*, 2646. [CrossRef]
79. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [CrossRef]
80. Diakogiannis, F.I.; Waldner, F.; Caccetta, P. Looking for Change? Roll the Dice and Demand Attention. *Remote Sens.* **2021**, *13*, 3707. [CrossRef]
81. Ye, Y.; Zhou, L.; Zhu, B.; Yang, C.; Sun, M.; Fan, J.; Fu, Z. Feature Decomposition-Optimization-Reorganization Network for Building Change Detection in Remote Sensing Images. *Remote Sens.* **2022**, *14*, 722. [CrossRef]
82. Zhang, X.; Yu, W.; Pun, M.O. Multilevel Deformable Attention-Aggregated Networks for Change Detection in Bitemporal Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5621518. [CrossRef]
83. Chen, T.; Lu, Z.; Yang, Y.; Zhang, Y.; Du, B.; Plaza, A. A Siamese Network Based U-Net for Change Detection in High Resolution Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 2357–2369. [CrossRef]
84. Li, Z.; Yan, C.; Sun, Y.; Xin, Q. A Densely Attentive Refinement Network for Change Detection Based on Very-High-Resolution Bitemporal Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4409818. [CrossRef]
85. Basavaraju, K.S.; Sravya, N.; Lal, S.; Nalini, J.; Reddy, C.S.; Dell'Acqua, F. UCDNet: A Deep Learning Model for Urban Change Detection From Bi-Temporal Multispectral Sentinel-2 Satellite Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5408110. [CrossRef]
86. Zheng, H.; Gong, M.; Liu, T.; Jiang, F.; Zhan, T.; Lu, D.; Zhang, M. HFA-Net: High frequency attention siamese network for building change detection in VHR remote sensing images. *Pattern Recognit.* **2022**, *129*, 108717. [CrossRef]
87. Cheng, G.; Wang, G.; Han, J. ISNet: Towards Improving Separability for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5623811. [CrossRef]
88. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9351, pp. 234–241. [CrossRef]
89. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
90. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef] [PubMed]
91. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
92. 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. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
93. Abdani, S.R.; Zulkifley, M.A.; Mamat, M. U-Net with Spatial Pyramid Pooling Module for Segmenting Oil Palm Plantations. In Proceedings of the 2020 IEEE second International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 26–27 September 2020; pp. 1–5. [CrossRef]
94. Sobel, I.; Feldman, G. An Isotropic 3x3 Image Gradient Operator. *Pattern Classification and Scene Analysis*; Wiley: New York, NY, USA, 1973; pp. 271–272.
95. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Virtual, 16–20 November 2020; pp. 38–45. [CrossRef]
96. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929. [CrossRef].
97. Zhang, C.; Wang, L.; Cheng, S.; Li, Y. SwinSUNet: Pure Transformer Network for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5224713. [CrossRef]

98. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 9992–10002. [CrossRef]
99. Hou, B.; Wang, Y.; Liu, Q. Change Detection Based on Deep Features and Low Rank. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2418–2422. [CrossRef]
100. Gong, M.; Yang, Y.; Zhan, T.; Niu, X.; Li, S. A Generative Discriminatory Classified Network for Change Detection in Multispectral Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 321–333. [CrossRef]
101. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised deep change vector analysis for multiple-change detection in VHR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3677–3693. [CrossRef]
102. Holgado Alvarez, J.L.; Ravanbakhsh, M.; Demir, B. S2-cGAN: Self-Supervised Adversarial Representation Learning for Binary Change Detection in Multispectral Images. In Proceedings of the IEEE 2020 International Geoscience and Remote Sensing Symposium (IGARSS 2020), Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2515–2518. [CrossRef]
103. Wu, C.; Chen, H.; Du, B.; Zhang, L. Unsupervised Change Detection in Multitemporal VHR Images Based on Deep Kernel PCA Convolutional Mapping Network. *IEEE Trans. Cybern.* **2021**, *52*, 12084–12098. [CrossRef]
104. Zhang, M.; Shi, W. A Feature Difference Convolutional Neural Network-Based Change Detection Method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232–7246. [CrossRef]
105. Leenstra, M.; Marcos, D.; Bovolo, F.; Tuia, D. Self-supervised pre-training enhances change detection in Sentinel-2 imagery. *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 12667, pp. 578–590. [CrossRef]
106. Chen, H.; Li, W.; Shi, Z. Adversarial Instance Augmentation for Building Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5603216. [CrossRef]
107. Saha, S.; Mou, L.; Zhu, X.X.; Bovolo, F.; Bruzzone, L. Semisupervised Change Detection Using Graph Convolutional Network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 607–611. [CrossRef]
108. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2281. [CrossRef] [PubMed]
109. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
110. Burchfield, D.R.; Petersen, S.L.; Kitchen, S.G.; Jensen, R.R. sUAS-Based Remote Sensing in Mountainous Areas: Benefits, Challenges, and Best Practices. *Pap. Appl. Geogr.* **2020**, *6*, 72–83. [CrossRef]
111. Orusa, T.; Cammareri, D.; Borgogno Mondino, E. A Scalable Earth Observation Service to Map Land Cover in Geomorphological Complex Areas beyond the Dynamic World: An Application in Aosta Valley (NW Italy). *Appl. Sci.* **2022**, *13*, 390. [CrossRef]
112. Zhu, L.; Zhang, Y.; Wang, J.; Tian, W.; Liu, Q.; Ma, G.; Kan, X.; Chu, Y. Downscaling Snow Depth Mapping by Fusion of Microwave and Optical Remote-Sensing Data Based on Deep Learning. *Remote Sens.* **2021**, *13*, 584. [CrossRef]
113. Southworth, J.; Muir, C. Specialty Grand Challenge: Remote Sensing Time Series Analysis. *Front. Remote Sens.* **2021**, *2*, 770431. [CrossRef]