remote sensing

MDPI

Article

S2Looking: A Satellite Side-Looking Dataset for Building Change Detection

Li Shen [1], Yao Lu [1,*], Hao Chen [2], Hao Wei [3], Donghai Xie [4], Jiabao Yue [4], Rui Chen [3], Shouye Lv [1] and Bitao Jiang [1]

1 Beijing Institute of Remote Sensing, Beijing 100011, China; shenli@bjirs.org.cn (L.S.); lvshouye@bjirs.org.cn (S.L.); jiangbitao@bjirs.org.cn (B.J.)
2 Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China; justchenhao@buaa.edu.cn
3 School of Microelectronics, Tianjin University, Tianjin 300072, China; weihao@tju.edu.cn (H.W.); ruichen@tju.edu.cn (R.C.)
4 Institute of Resource and Environment, Capital Normal University, Beijing 100048, China; xiedonghai@cnu.edu.cn (D.X.); yuejiabao2019@cnu.edu.cn (J.Y.)
* Correspondence: yaolu@bjirs.org.cn

Abstract: Building-change detection underpins many important applications, especially in the military and crisis-management domains. Recent methods used for change detection have shifted towards deep learning, which depends on the quality of its training data. The assembly of large-scale annotated satellite imagery datasets is therefore essential for global building-change surveillance. Existing datasets almost exclusively offer near-nadir viewing angles. This limits the range of changes that can be detected. By offering larger observation ranges, the scroll imaging mode of optical satellites presents an opportunity to overcome this restriction. This paper therefore introduces S2Looking, a building-change-detection dataset that contains large-scale side-looking satellite images captured at various off-nadir angles. The dataset consists of 5000 bitemporal image pairs of rural areas and more than 65,920 annotated instances of changes throughout the world. The dataset can be used to train deep-learning-based change-detection algorithms. It expands upon existing datasets by providing (1) larger viewing angles; (2) large illumination variances; and (3) the added complexity of rural images. To facilitate the use of the dataset, a benchmark task has been established, and preliminary tests suggest that deep-learning algorithms find the dataset significantly more challenging than the closest-competing near-nadir dataset, LEVIR-CD+. S2Looking may therefore promote important advances in existing building-change-detection algorithms.

Keywords: change detection; remote sensing; benchmark dataset; neural networks

Citation: Shen, L.; Lu, Y.; Chen, H.; Wei, H.; Xie, D.; Yue, J.; Chen, R.; Lv, S.; Jiang, B. S2Looking: A Satellite Side-Looking Dataset for Building Change Detection. *Remote Sens.* **2021**, *13*, 5094. https://doi.org/10.3390/rs13245094

Academic Editor: Mohammad Awrangjeb

Received: 19 November 2021
Accepted: 13 December 2021
Published: 15 December 2021

1. Introduction

Building-change detection underpins a range of applications in domains such as urban expansion monitoring [1], land use and cover type change monitoring [2,3], and resource management and evaluation [4]. It is of particular importance in the contexts of military surveillance and crisis management [5], where changes in buildings may be indicative of a developing threat or areas in which to focus disaster relief. Change detection involves identifying changes and differences in an object or phenomenon at different times [6]. Remote-sensing-based change detection uses multitemporal remote-sensing image data to analyze the same area in order to identify changes in the state information of ground features according to changes in the images.

Many change-detection methods have been proposed over the years. Traditional methods tended to be either pixel-based [7,8] or object-based [9–11]. Pixel-based change-detection methods involve pixel by pixel analysis of spectral or textural information of input image pairs followed by threshold-based segmentation to obtain the detection results [7,8].

Object-based change-detection methods similarly rely on spectral and textural information, but also consider other cues, such as structural and geometric details [9,10]. However, while these methods can effectively extract geometric structural details and set thresholds, they are easily influenced by variations in image detail and quality. This undermines their accuracy [12]. In recent years, the principal methods used for remote-sensing-based change detection have therefore shifted towards deep learning. This reflects a wider revolution in computer vision research [13]. Change-detection methods based on deep-learning include dual-attention fully convolutional Siamese networks (DASNet) [14], image fusion networks (IFN) [15], end-to-end change detection based on UNet++ (CD-UNet++) [13], fully convolutional Siamese networks based on concatenation and difference (FC-Siam-Conc and FC-Siam-Diff) [16], and dual-task constrained deep Siamese convolutional networks (DTCDSCN) [17]. Each of these reduces the risk of error by eliminating the need for preprocessing of the images [18].

Although deep-learning-based change-detection methods generally outperform other change-detection methods, their performance is heavily dependent on the scale, quality, and completeness of the datasets they use for training. A strong demand has therefore emerged for large-scale and high-quality change-detection datasets. A number of open datasets for remote-sensing change detection have been developed to meet this demand, such as the Change Detection Dataset [19], the WHU Building Dataset [1], the SZTAKI Air Change Benchmark Set (SZTAKI) [20,21], the OSCD dataset (OSCD) [16], the Aerial Imagery Change Detection dataset (AICD) [22], and the LEVIR-CD dataset [23], which was released last year. In addition, the privacy concerns can be addressed by privacy-preserving deep-learning-based techniques while providing the real-world data from satellite images [24].

However, most of these change-detection datasets are based on near-nadir imagery. While this is sufficient for certain kinds of change detection, the observed building features in these datasets are relatively simple, limiting the scope for change-detection algorithms to serve a comprehensive range of practical applications.

A potential solution to this constraint is offered by the scroll imaging mode adopted by the cameras in a number of modern optical satellites. Figure 1 shows the basic way in which this works.

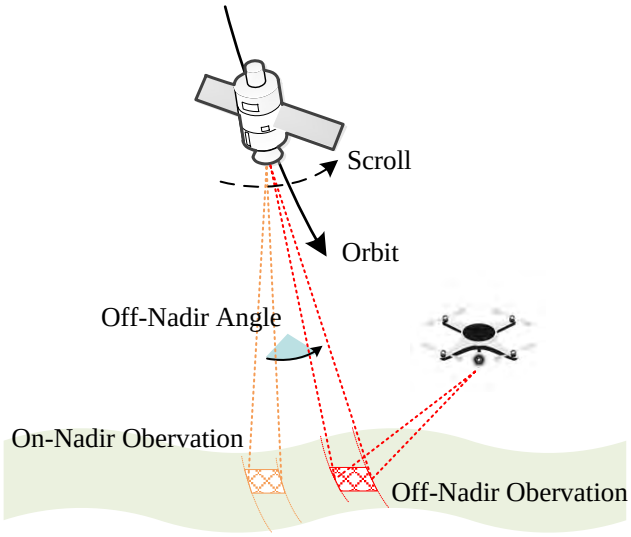

Figure 1. Scroll imaging. The satellite scrolls during the imaging process and obtains side-looking remote-sensing images.

In outline, as an optical satellite pursues its orbit, the onboard high-resolution cameras are able to scroll so that they can capture multiple images of the same object from different angles, rather than solely from overhead [25]. Unlike near-nadir satellite imagery, side-looking imagery can capture more relevant details of ground objects and yield more

potentially useful information. Scroll imaging endows surveillance satellites with a better imaging range and a shorter revisit period than conventional mapping satellites [26]. For example, the GF-1 satellite has a revisit period of 41 days, together with 4 days for $\pm 35^{\circ}$ off-nadir angles; the GF-2 satellite has a revisit period of 69 days, together with 5 days for $\pm 35^{\circ}$ off-nadir angles.

To date, the only study that has made a serious effort to develop a dataset consisting of multiangle satellite-derived images of specific objects in this way is SpaceNet MVOI [27]. However, this dataset is not focused on change detection, so it does not offer bitemporal images, and only contains 27 views from different angles at a single geographic location.

In this paper, we therefore introduce S2Looking, a building-change-detection dataset that contains large-scale side-looking satellite images captured at various off-nadir angles. The dataset consists of 5000 bitemporal pairs of very-high-resolution (VHR) registered images collected from the GaoFen (GF), SuperView (SV) and BeiJing-2 (BJ-2) satellites from 2017 to 2020. The imaged areas cover a wide variety of globally-distributed rural areas, with very different characteristics, as can be seen in Figure 2. The dataset shown in Figure 2 contains various scenes from all over the world, including villages, farms, villas, retail centers, and industrial areas, which relate to each row above, respectively. Each image in the pairs is 1024×1024 with an image resolution of $0.5 \sim 0.8$ m/pixels. The image pairs in the dataset are converted from the original TIFF format with 16 bit to PNG format with 8 bit. The pairs are accompanied by 65,920 expert-based annotations of changes and two label maps that separately indicate newly built and demolished building regions for each sample in the dataset. The side-view imaging and complex rural scenes in the dataset present whole new challenges for change-detection algorithms by making the identification and matching of buildings notably more difficult. Placing higher requirements on an algorithm's robustness by confronting it with more complex ground targets and imaging conditions increases its practical value if it can successfully meet such challenges. Thus, S2Looking offers a whole new resource for the training of deep-learning-based change-detection algorithms. It significantly expands upon the degree of richness offered by available datasets, by providing (1) larger viewing angles; (2) large illumination variances; and (3) the added complexity of the characteristics encountered in rural areas. It should also be noted that algorithms trained on S2Looking are likely to be better able to deal with other forms of aerially-acquired imagery, for instance by aircraft, because such images present similarly offset viewing angles. To facilitate use of the dataset, we have established a benchmark task involving the pixel-level identification of building changes in bitemporal images. We have also conducted preliminary tests of how existing deep-learning algorithms might perform when using the dataset. When this was compared with their performance on the closest-competing dataset, LEVIR-CD+, which is based on near-nadir images, the results revealed that S2Looking was substantially more challenging. This suggests it has the potential to induce step-change developments in building-change-detection algorithms that seek to address the challenges it presents.

In the next two subsections, we discuss work relating to change detection based on remote-sensing images and change-detection datasets, with the latter playing an important role in the analysis and processing required for the prior.

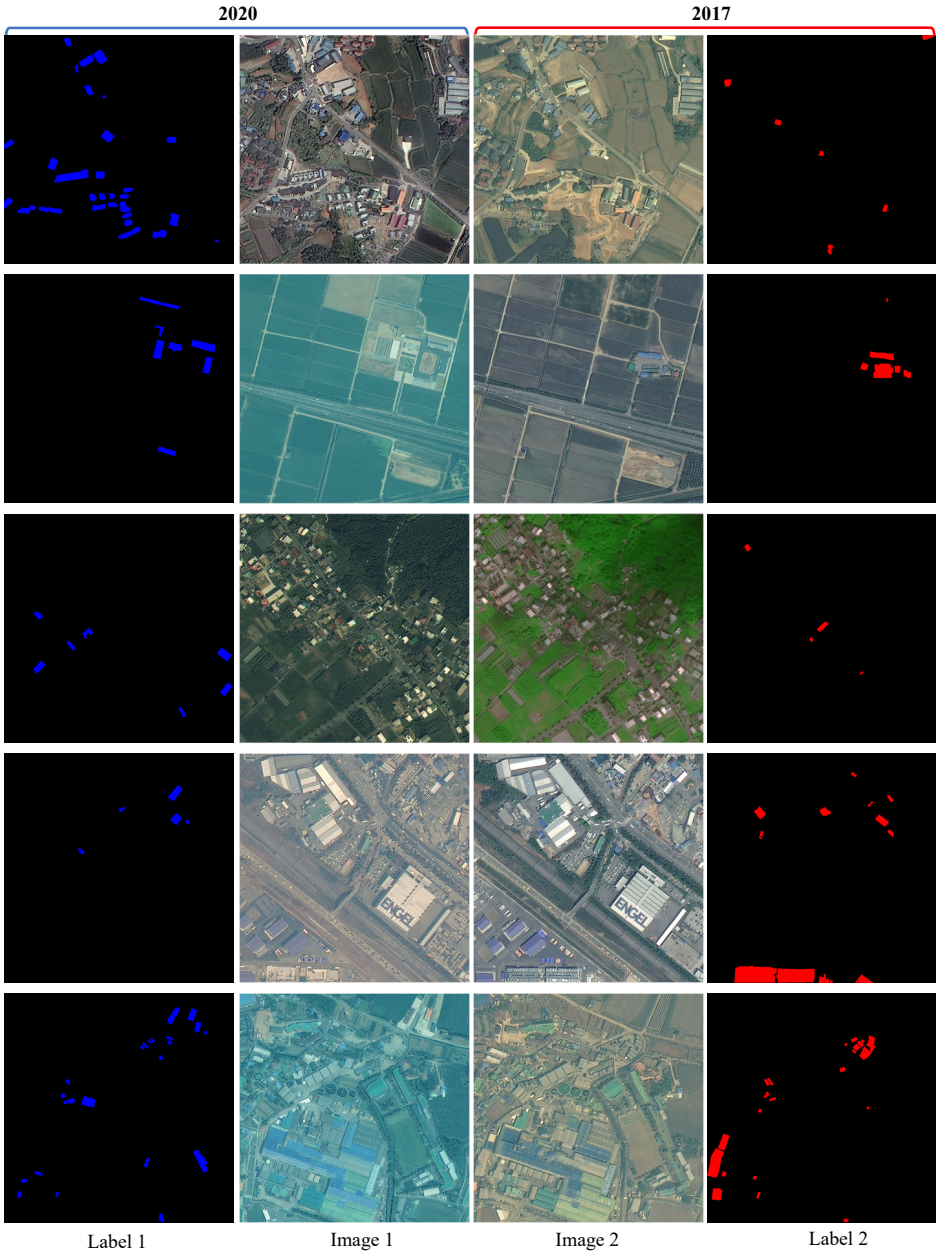

Figure 2. Samples from the S2Looking dataset. Images 1 and 2 in Figure 2 are bitemporal remote-sensing images, while Labels 1 and 2 are the corresponding annotation maps. Labels 1 and 2 indicate pixel-precise newly built and demolished areas of buildings, respectively.

1.1. Change Detection Methods

Traditional change-detection methods were originally either pixel-based [7,8] or object-based [9–11]. Traditional methods of change detection in remote-sensing images are designed on the basis of handcrafted features and supervised classification algorithms. These methods were capable of extracting geometric structural features from images and then applying thresholds that would indicate some kind of change when they were cross-compared. However, as pointed out in [12], these kinds of methods are very susceptible to variations in the images and their quality, making their accuracy questionable. However, with the rapid development of computer hardware and artificial intelligence, deep-learning-based methods have also been widely studied and applied. When methods based on deep learning began to hold sway in image processing, it became apparent that they could be used to address these issues, leading to their rapid adoption in change detection.

Deep-learning-based change detection can be roughly divided into metric-based methods and classification-based methods. Metric-based methods involve learning a parameterized embedding space where there is a large distance between the changed pixels and a small distance between the unchanged pixels. A change map can then be obtained by calculating the distances between embedded vectors at different times in the same position. Zhan et al. [28] used a deep Siamese fully convolutional network with weight sharing to learn the embedding space and extract features from images captured independently at different times. Saha et al. [29] proposed an unsupervised deep change vector analysis method based on a pretrained convolutional neural network (CNN) and contrastive/triplet loss functions [30,31]. Chen et al. [14] proposed DASNet to overcome the influence of pseudo change information in the recognition process. Chen et al. [23] proposed a spatiotemporal attention neural network (STANet) based on the FCN-network (STANet-Base). This enabled them to produce two improved models with self-attention modules; one with a basic spatiotemporal attention module (STANet-BAM), the other with a pyramid spatiotemporal attention module (STANet-PAM).

Classification-based methods typically use CNNs to learn the mapping from bitemporal data to develop a change probability map that classifies changed categories in a pixelwise fashion. A changed position has a higher score than an unchanged position. Zhang et al. [15] proposed the IFN model, which relies on a deep supervised difference discrimination network (DDN) to detect differences in the proposed image features. Peng et al. [13] developed an improved automatic coding structure based on the UNet++ architecture and proposed an end-to-end detection method with a multilateral fusion strategy. Daudt et al. [16,32] proposed a fully convolutional early fusion (FC-EF) model, which concatenates image pairs before passing them through a UNet-like network. They also developed the FC-Siam-Conc and FC-Siam-Diff models, which concatenate image pairs after passing them through a Siamese network structure. Liu et al. [17] proposed a dual-task constrained deep Siamese convolutional network DTCDSCN, within which spatial and channel attention mechanisms are able to obtain more discriminatory features. This network also incorporates semantic segmentation subnetworks for multitask learning. Chen et al. [33] proposed a simple yet effective change-detection network (CDNet) that uses a deep Siamese fully convolutional network as the feature extractor and a shallow fully convolutional network as the change classifier for feature differences in the images. Very recently, Chen et al. [34] proposed an efficient transformer-based model (BiT), which leverages a transformer to exploit the global context within a token-based space, thereby enhancing the image features in the pixel space.

The main difference between metric-based and classification-based approaches is the network structure. A classification-based approach offers more diverse choices than a metric-based one. For example, the former mostly makes use of a late fusion architecture, where the bitemporal features are first extracted and then compared to obtain the change category. The latter, however, can make use of both late and early fusion. As a result, classification-based methods are more commonly used than metric-based methods. In terms of the loss function, the former usually uses contrastive loss, which is designed for paired bitemporal feature data, while the latter uses cross-entropy loss for the fused features. Overall, deep learning approaches are very robust when it comes to variability in the data, but one of their key drawbacks is that they are only as good as the datasets that they use for training [35–37]. This places a heavy emphasis upon the development of high-quality and comprehensive datasets, with change detection being no exception.

1.2. Change Detection Datasets

There are many large-scale benchmark datasets available for detection, recognition, and segmentation based on everyday images. These include ImageNet [38], COCO [39], PIE [40], MVTec AD [41], Objects365[42], WildDeepfake [43], FineGym [44], and ISIA Food-500 [45]. There are also large-scale datasets containing satellite and aerial imagery, such as ReID [46], which contains 13 k ground vehicles captured by unmanned aerial vehicle

(UAV) cameras, and MOR-UAV [47], which consists of 30 UAV videos designed to localize and classify moving ground objects. WHU [48] consists of thousands of multiview aerial images for multiview stereo (MVS) matching tasks. DeepGlobe [49] provides three satellite imagery datasets, one for building detection, one for road extraction, and one for land-cover classification. xBD [50] and Post-Hurricane [51] consist of postdisaster remote-sensing imagery for building damage assessment. SensatUrban [52] contains urban-scale labeled 3D point clouds of three UK cities for fine-grained semantic understanding. FAIR1M [53] provides more than 1 million annotated instances labeled with their membership in 5 categories and 37 subcategories, which makes it the largest dataset for object detection in remote-sensing images presented so far. The stereo satellite imagery [54] and light detection and ranging (LiDAR) point clouds [55] provide the base data for the 3D city modeling technology. However, in the above datasets, each area is covered by only a single satellite or aerial image. Change-detection models need to be trained on datasets consisting of bitemporal image pairs, which typically correspond to different sun/satellite angles and different atmospheric conditions.

In Table 1, we present the statistics relating to existing change-detection datasets, together with those for our S2Looking dataset. It can be seen that SZ-TAKI [20,21] contains 12 optical aerial image pairs. These focus on concerns such as building changes and the planting of forests. OSCD [16] focuses on urban regions and includes 24 multispectral satellite image pairs. Unlike other datasets, AICD [22] is a synthetic change-detection dataset containing 500 image pairs. The LEVIR-CD dataset [23] consists of 637 manually-collected image patch pairs from Google Earth. Its recently expanded version, LEVIR-CD+, contains 985 pairs. The Change Detection Dataset [19] is composed of 11 satellite image pairs. The WHU Building Dataset [1] consists of just one max-width aerial image collected from a region that suffered an earthquake and was then rebuilt in the following years. Our own dataset, S2Looking, contains 5000 bitemporal pairs of rural images. Generally, S2Looking has the most image pairs; WHU has the largest size and best resolution; S2Looking offers the most change instances and change pixels. Apart from these general change-detection datasets, a river change-detection dataset [56] has also been released that specifically concentrates on the detection of changes in rivers that are locatable through hyperspectral images.

Table 1. Statistical characteristics of existing change-detection datasets.

Dataset	Pairs	Size	Is Real?	Resolution/m	Change Instances	Change Pixels
SZTAKI [20,21]	12	952 × 640	√	1.5	382	412,252
OSCD [16]	24	600 × 600	√	10	1048	148,069
AICD [22]	500	600 × 800	×	none	500	203,355
LEVIR-CD [23]	637	1024 × 1024	√	0.5	31,333	30,913,975
Change Detection Dataset [19]	7/4	4725 × 2700/1900 × 1000	√	0.03 to 1	1987/145	9,198,562/400,279
WHU Building Dataset [1]	1	32,507 × 15,354	√	0.075	2297	21,352,815
LEVIR-CD+	985	1024 × 1024	√	0.5	48,455	47,802,614
S2Looking	5000	1024 × 1024	√	0.5∼0.8	65,920	69,611,520

The datasets described above (excepting our new S2Looking) have played an important part in promoting the development of change-detection methods. However, they share a common drawback in that most of the images they contain have been captured at near-nadir viewing angles. The absence of large-scale off-nadir satellite image datasets limits the scope for change-detection methods to advance to a point where they can handle more subtle incremental changes in objects such as buildings. This is the underlying motivation behind the development of the S2Looking dataset, as described in more detail below.

1.3. Contributions

Overall, the primary contributions of this paper are as follows: (1) a pipeline for constructing satellite remote-sensing image-based building-change-detection datasets; (2) presentation of a unique, large-scale, side-looking, remote-sensing dataset for building-

change detection covering rural areas around the world; (3) a benchmark test upon which existing algorithms can assess their capacity to undertake the monitoring of building changes when working with large-scale side-looking images; (4) a preliminary evaluation of the added complexity presented by the dataset. On the basis of this, we reflect briefly upon potential future developments in building-change detection based on surveillance satellites.

2. Materials and Methods

As noted in [37], the lack of public large-scale datasets that cover all kinds of satellites, all kinds of ground surfaces, and from a variety of angles, places limits upon how much progress can be made in change-detection algorithm research and the resulting applications. This has driven our development of the S2Looking dataset. The situation was somewhat improved by the release of the LEVIR-CD dataset in 2020 [23], which presented a good range of high-resolution change instances. LEVIR-CD has very recently been superseded by LEVIR-CD+ (also available at https://github.com/S2Looking/, accessed 1 November 2021), with an even larger number of bitemporal image pairs. However, the LEVIR-CD+ dataset mainly targets urban areas as captured in survey or mapping satellite data from Google Earth. It also retains the focus of prior datasets upon near-nadir imagery. As LEVIR-CD+ was, until the development of S2Looking, the richest dataset available for testing change-detection algorithms, S2Looking has, in many ways, been framed against it, hence the focus in S2Looking upon mainly rural targets captured by surveillance or reconnaissance satellites at varying off-nadir angles. S2Looking thus provides both the largest and the most challenging change-detection dataset available in the public domain. Throughout the remainder of the paper, active comparisons are made between S2Looking and LEVIR-CD+ because the latter constitutes the immediately preceding state-of-the-art. Over the course of this section, we look in more detail at the objectives we were pursuing in developing S2Looking and its data-processing pipeline.

2.1. Motivation for the New S2Looking Dataset

As noted above, S2Looking was actively developed to expand upon the change-detection affordances of LEVIR-CD+. LEVIR-CD+ is itself based on the LEVIR-CD dataset presented in [23]. In comparison with the 637 image patch pairs in the LEVIR-CD dataset, LEVIR-CD+ contains more than 985 VHR (0.5 m/pixel) bitemporal Google Earth images, with a size of 1024 × 1024 pixels. These bitemporal images are from 20 different regions located in several cities in the state of Texas in the USA. The capture times of the image data vary from 2002 to 2020. Images of different regions were taken at different times. Each pair of bitemporal images has a time span of 5 years.

Unlike the LEVIR-CD+ dataset, S2Looking mostly targets rural areas. These are spread throughout the world and were imaged at varying large off-nadir angles by optical satellites. The use of Google Earth by LEVIR-CD+ actually places certain limits upon it, because, while Google Earth provides free VHR historical images for many locations, its images are obtained by mapping satellites to ensure high resolution and geometric accuracy. In contrast, images captured by optical satellites, especially surveillance satellites, can benefit from their use of scroll imaging to achieve a better imaging range and a shorter revisit period (as shown in Figure 1). For example, when a disaster occurs, scroll imaging by satellites is used to support analysts who require quick access to satellite imagery of the impacted region, instead of waiting until a satellite reaches orbit immediately above. By varying the camera angle on subsequent passes, it is also possible to revisit the impacted region more often to monitor recovery and look for signs of further events. For military applications, scroll imaging by satellites is frequently used to obtain a steady fix on war zones [57]. However, a challenge arising from using off-nadir angles that needs to be properly met by new change-detection algorithms is that there is a lack of geometric consistency between tall objects in the gathered satellite images. This effect can be seen in Figure 3. The subvertical objects highlighted in yellow in Figure 3 become more visible

and have a larger spatial displacement proportionate with the off-nadir angle in the middle and right images.

Figure 3. Examples of side-looking remote-sensing images. (**a**) Near-nadir imagery from Google Earth. (**b**) Side-looking imagery at an off-nadir angle of 10° from S2Looking. (**c**) Side-looking imagery at an off-nadir angle of 18° from S2Looking.

A further value of developing an off-nadir-based dataset for change detection is that aerial imagery, airborne and missile-borne, also tends to be captured at large off-nadir angles to maximize visibility according to the flight altitude (see, for instance, [58]). As such imagery presents the same problem of there being a lack of geometric consistency caused by the high off-nadir angles, an image processing model trained on side-looking satellite imagery will more easily adapt to aerial imagery, raising the possibility of it supporting joint operations involving satellites and aircraft.

Figure 4 illustrates the geospatial distribution of our new dataset. Most remote-sensing images in S2Looking relate to rural areas near the marked cities. The chosen cities have been a particular focus of satellite imagery, so it was easier to acquire bitemporal images for these locations. Together, the adjacent rural areas cover most types of rural regions around the world.

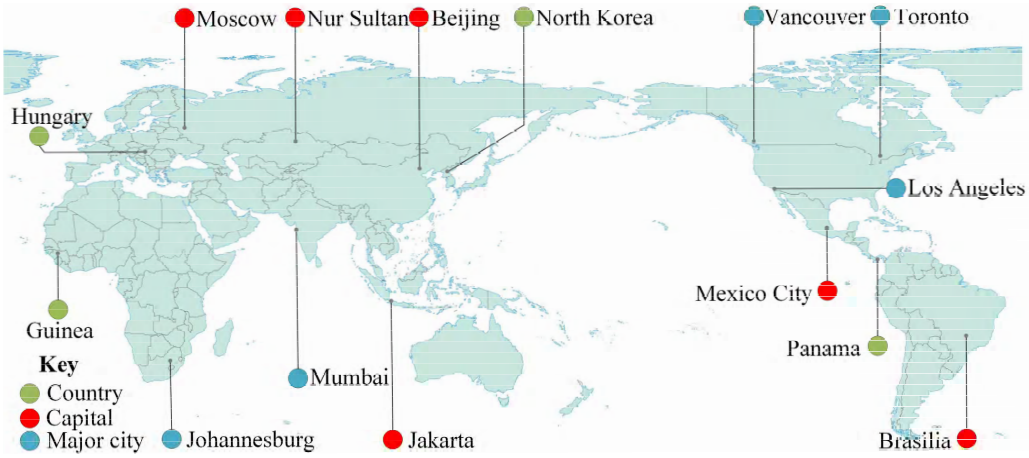

Figure 4. The global locations covered by the bitemporal images in the S2Looking dataset.

When it comes to remote sensing change detection, data from rural areas generally have more value than data from urban areas. There are several reasons for this. First of all, in the case of military surveillance, sensitive installations are usually built in isolated areas for safety reasons [25]. In the case of disaster rescue, satellite images of remote rural regions can be obtained faster than aerial photographs because satellites do not have a restricted flying range [59]. Therefore, rural images offer the scope to train remote-sensing

change-detection models that can find and update the status of military installations or destroyed buildings for disaster rescue teams. As is noted in Section 4 regarding the challenges posed by S2Looking, there are certain characteristics that accrue to rural images that can render it more difficult to recognize buildings, enhancing their value for training.

However, side-looking satellite imagery of rural areas is more difficult to collect than vertical-looking imagery of urban areas. The buildings can be imaged by satellites from different sides and projected along different angles into 2D images, as we see in Figure 3. Table 2 provides a summary of our dataset. The off-nadir angles have an average absolute value of 9.86° and a standard deviation of 12.197°. A frequency histogram of the dataset is also given in Figure 5. Here, it can be seen that the off-nadir angles ranged from $-35°$ to $+40°$, with the highest frequency of 1158 clustering between $-5°$ and $0°$, reflecting the optimal imaging angle. Optical satellites usually have on-nadir observation angles of within $\pm 15°$ to maximize their lifespan. The off-nadir mode typically relates to observation angles larger than 15°. Our dataset consists of 71.9% on-nadir and 28.1% off-nadir image pairs. This level of variation makes it difficult for a registration algorithm to match feature points between the bitemporal images. Additionally, irrelevant structures, such as greenhouses and small reservoirs, can interfere with the building annotation process.

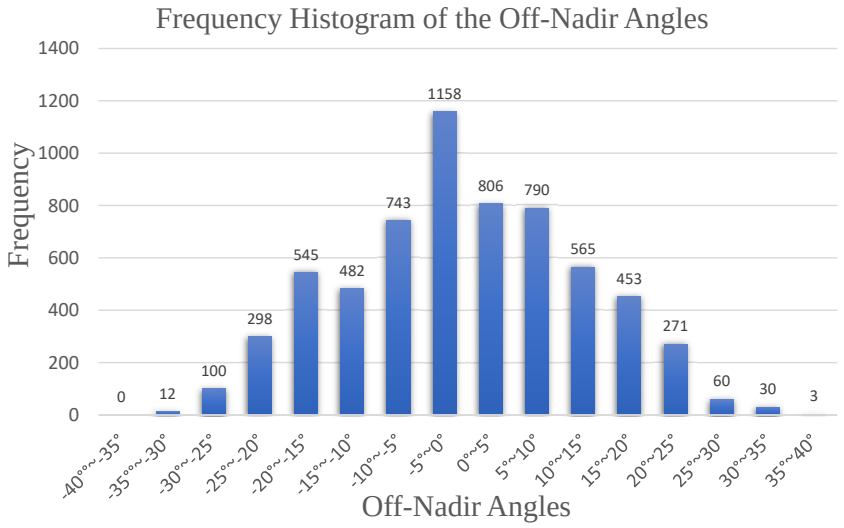

Figure 5. Frequency histogram of the off-nadir angles of the satellite images in the S2Looking dataset.

Table 2. A summary of the S2Looking dataset.

Type	Item	Value
Image Info	Total Image Pairs	5000
	Image Size	1024 × 1024
	Image Resolution	0.5∼0.8 m
	Time Span	1∼3 years
	Modality	RGB image
	Image Format	PNG 8bit
Off-Nadir Angle Info	Average Absolute Value	9.861°
	Median Absolute Value	9.00°
	Max Absolute Value	35.370°
	Standard Deviation	12.197°
Accuracy Info	Registration Accuracy	$\leq$ 8 pixels
	Annotation Accuracy	$\leq$ 2 pixels

2.2. The S2Looking Data Processing Pipeline

An illustration of the data-processing pipeline for our S2Looking dataset is shown in Figure 6. This pipeline is discussed in greater detail below:

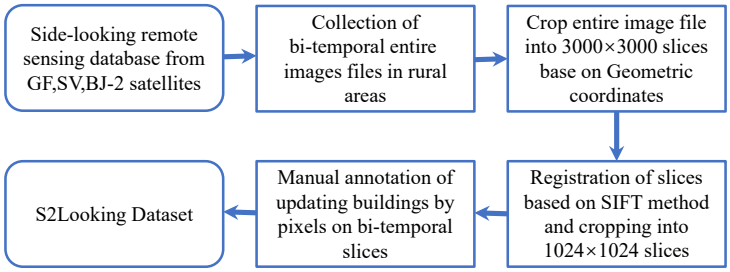

Figure 6. Data processing pipeline for the S2Looking dataset.

Image Selection. The images in the dataset were selected on the basis of both time and location. Locations were selected that covered typical types of rural areas around the world. The satellite images were then checked, and only locations that had been captured more than once in the past 10 years were retained.

Collection of bitemporal image files. Based on the chosen rural locations, we next checked our satellite image database and collected time-series images for each specific location. We then selected entire bitemporal image files with an intersection-over-union (IoU) greater than 0.7 and a time span greater than 1 year.

Cropping into slices based on geometric coordinates. Each remote-sensing image contained the rough coordinate information for each pixel, which was obtained by geometric inversion of the position of the satellite and the camera angle. Based on these geometric coordinates, the intersecting areas were selected and cropped into 3000 × 3000 slices. The final image size of 1024 × 1024 is better suited to GPU acceleration because it matches the number of GPU threads. The accuracy of the coordinates was about 20 m (about 40∼100 pixels), so we cropped the intersection areas into 3000 × 3000 slices with a stride of 1024 for the registration. The 3000 × 3000 slices therefore overlapped. After the registration process, the central 1024 × 1024 area was retained to form the bitemporal pairs in the S2Looking dataset.

Bitemporal image registration. Image registration is essential for the preparation of image pairs for a change-detection dataset [60]. Although accurate spatial coregistration for images captured at near-nadir angles can be achieved with accurate digital elevation models (DEMs) [61,62], the geolocalization accuracy deteriorates for large off-nadir angles, especially in the case of hilly/mountainous areas [63,64]. In addition, dense global DEMs (i.e., with a resolution of at least 2 m) were not available.

According to our experience, the building-change-detection task is achievable when the precision of the geometric alignment of bitemporal image pairs is less than 8 pixels. We therefore used a scale-invariant feature transform (SIFT) algorithm [65] to find and match the feature points in each pair of corresponding images. The SIFT algorithm constructs a multiscale image space by means of Gaussian filtering and searches for extreme points in the difference of Gaussian (DOG) image, which are used as feature points. For the feature description, the SIFT algorithm uses a gradient direction histogram to describe the feature points and uses the ratio of the distances to the nearest and second nearest points as the basis for matching. The SIFT algorithm is robust to illumination changes and can even handle some degree of affine or perspective distortion [66]. To improve the accuracy of the SIFT algorithm, we used floating-point variables to store the gray values of each image. This avoids ending up with missing values because of integer approximation when the Gaussian filtering and DOG results are calculated.

After this, incorrectly matched pairs were deleted according to a random sample consensus (RANSAC) homography transformation [67]. Finally, based on the matched points, the image pairs were resampled using a homography matrix to ensure the matched points had the smallest possible pixel distances in each image, as shown in Figure 7. Thus, the same features and buildings in the image pairs had the same pixel index in the bitemporal images and were suitable for change detection by recognizing any inconsistent buildings.

To avoid misaligned image pairs and guarantee the registration accuracy, all the bitemporal images in the S2Looking dataset were manually screened by remote sensing image-interpretation experts to ensure that the maximum registration deviation was less than 8 pixels. This means that misaligned areas are less than 1/16 of the average building change size in S2Looking (1056 pixels). Although absolutely accurate registration is probably impossible, further improvement of the registration of S2Looking is required. This would enhance the performance of change-detection models trained on the dataset. Registration could be improved, for instance, by manual annotation of the same points in the bitemporal pairs. A large number of registration methods are now available based on deep learning, and our hope is that the process can be iteratively refined by various researchers making use of the S2Looking dataset and undertaking their own registration activities prior to applying change-detection methods.

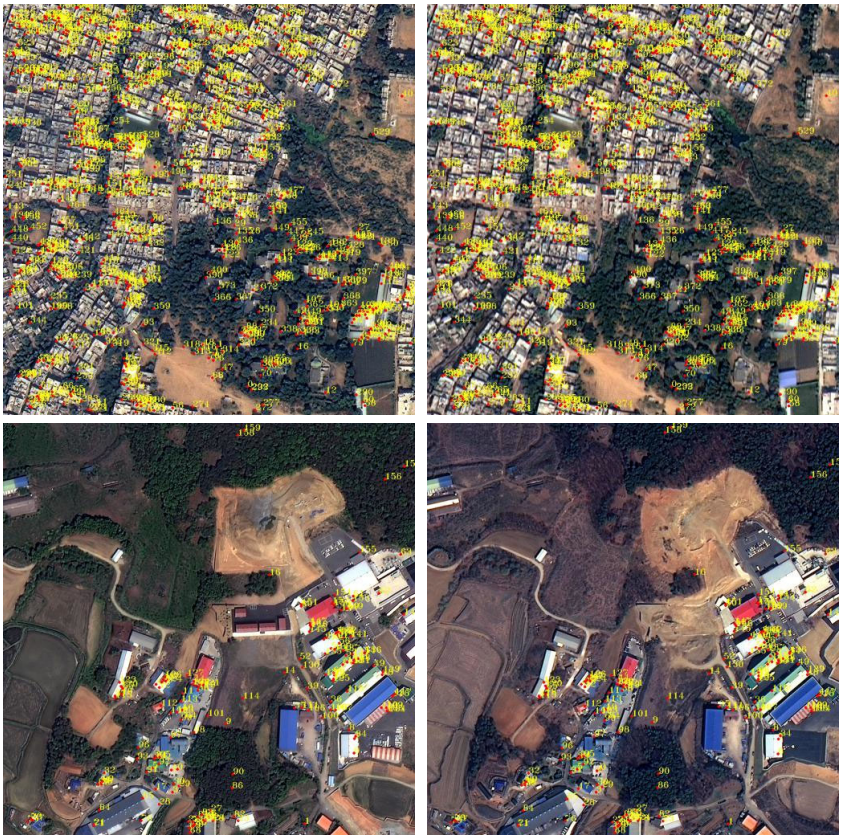

Figure 7. Registered feature points. The images in the left and right columns form bitemporal pairs. The numbered points are the matched feature points in the two images.

Annotation and Quality Control. The bitemporal images were also annotated by the remote sensing image-interpretation experts. All newly built and demolished building regions in the dataset were annotated at the pixel level in separate auxiliary graphs, thus making further registration processes possible. The task was jointly completed by 5 remote-sensing experts from the authors' research institution, 8 postgraduate students from partner universities and more than 20 well-trained annotators employed by the Madacode AI data service company (www.madacode.com, accessed 1 November 2021). The remote-sensing experts have been undertaking military surveillance and damage assessment for natural disasters such as fires for many years. The annotation accuracy was required to be higher than 2 pixels. All annotators had rich experience in interpreting remote-sensing images and a comprehensive understanding of the change-detection task. They followed detailed specifications made by the remote-sensing experts for annotating the images to yield consistent annotations. Moreover, each sample in our dataset was annotated by one annotator and then double-checked by another to maximize the quality

of the annotations. The labels were checked by the postgraduate students, then batches containing 20 bitemporal images were randomly reexamined by the remote-sensing experts. It took approximately 2000 person-days to manually annotate and review the entire dataset.

We saw some selected samples from the dataset in Figure 2. It should be noted that the construction of new buildings and the demolition of old buildings was annotated in separately labeled images. In this way, if the registration accuracy is improved, these labeled images can be simply adjusted without the need for any further annotation.

The S2Looking Dataset. As noted previously, the dataset can be downloaded from https://github.com/S2Looking/ accessed on 1 November 2021, with our hope being that innovative use of the dataset in the image-processing and change-detection communities will give rise to new requirements that can then be addressed through further dataset refinements.

2.3. The S2Looking Challenge

Buildings are representative manmade structures. During the last few decades, the areas from which our images were collected have seen significant land-use changes, especially in terms of rural construction. This presents more difficulties for change detection than construction in urban areas. Our VHR remote-sensing images provide an opportunity for researchers to analyze more subtle changes in buildings than simple construction and destruction. This might include changes from soil/grass/hardcore to building expansion and building decline. The S2Looking dataset therefore presents current deep learning techniques with a new and significantly more challenging resource, the use of which will likely lead to important innovation. To further promote change-detection research, we are currently organizing a competition based on the S2looking dataset (see https://www.rsaicp.com, accessed 1 November 2021), which will be addressed to the challenges identified below.

The S2Looking dataset was extracted from side-looking rural-area remote-sensing images, which makes it an expanded version of the LEVIR-CD+ dataset that will inevitably be harder to work with. Given S2Looking as training data, the challenge for the change-detection community is to create models and methods that can extract building growth and decline polygons quickly and efficiently. Furthermore, these models and methods will need to assign each polygon in a way that will accurately cover the range of building changes, meaning that each polygon and building region must be matched pixel by pixel.

Many methods have achieved satisfactory results on the LEVIR-CD dataset (F1-score > 0.85) [23,34], but the S2Looking dataset presents a whole new set of challenges that change-detection algorithms will need to be able to address. These are summarized below:

Sparse building updates. The changed building instances in the S2Looking dataset are far sparser than those in the LEVIR-CD+ dataset due to the differences between rural and urban areas. Most rural areas are predominantly covered with farmland and forest, while urban areas are predominantly covered with buildings that are constantly being updated. The average number of change instances in S2Looking is 13.184, while the average number of change instances in LEVIR-CD is 49.188 [23]. This makes it more difficult for networks to extract building features during the training process.

Side-looking images. The S2Looking dataset concentrates on side-looking remote-sensing imagery. This makes the change-detection problem different from ones relating to datasets consisting of Google Earth images. The buildings have been imaged by satellites from different sides and projected along varying off-nadir angles into 2D images, as we see in Figure 3. This means that a change-detection model is going to have to identify the same building imaged from different directions and detect any updated parts.

Rural complexity. Seasonal variations and land-cover changes unrelated to building updates are more obvious in rural areas than in urban areas. Farmland is typically covered by different crops or withered vegetation, depending on the season, giving

it a different appearance in different remote-sensing images. A suitable change-detection model needs to distinguish building changes from irrelevant changes in order to generate fewer false-positive pixels.

Registration accuracy. The registration process for the bitemporal remote-sensing images in S2Looking is not completely accurate due to the side-looking nature of the images and terrain undulations. Based on the manual screening by experts, we have managed to bring the registration accuracy to 8 pixels or better, but this necessitates a change-detection model that can tolerate slightly inaccurate registration.

2.4. Challenge Restrictions

To better accommodate operational use cases and maintain fairness, the geographic coordinate information has been removed from the data used for inference in the challenge. Thus, no geographic base map database can be applied in the change-detection challenge. The models for change detection and subsequent image registration are only allowed to extract information from the images themselves.

3. Results

In order to provide a benchmark against which change-detection algorithms can assess their performance with regard to meeting the above challenges, we have established a set of evaluation metrics. To gain a sense of the extent to which S2Looking has moved the requirements imposed on change-detection algorithms beyond those posed by the current baseline dataset, LEVIR-CD+, we undertook a thorough evaluation of benchmark and state-of-the-art deep-learning methods using both LEVIR-CD+ and S2Looking. This exercise confirmed the scope of our dataset to establish a new baseline for change detection. We report below upon the benchmark metrics and the evaluation that was undertaken. We conclude this section with a close analysis of the evaluation results and the ways in which existing change-detection algorithms are failing to meet the identified challenges.

3.1. Benchmark Setup

Train/Val/Test Split Statistics. We evaluated the performance of four classic (FC-EF, FC-Siam-Conc, FC-Siam-Diff, and DTCDSCN) and five state-of-the-art (STANet-Base, STANet-BAM, STANet-PAM, CDNet, and BiT) deep-learning methods on both the LEVIR-CD+ dataset and S2Looking dataset. The specific state-of-the-art methods were chosen because they had previously performed well for change detection on the LEVIR-CD+ dataset. As noted in Section 2, the LEVIR-CD+ dataset contains 985 image patch pairs. We designated 65% of these pairs as a training set and the remaining 35% as a test set. This is consistent with our previous work on LEVIR-CD+. The S2Looking dataset consists of 5000 image patch pairs, which we split into a training set of 70%, a validation set of 10%, and a test set of 20%. Strictly speaking, a validation set is not necessary for the training process and it can be merged with the training set. However, the validation set was able to capture the effect of each iteration during the training process, as shown in Figure 8. After each iteration, the algorithm was tested on the validation set to assess the level of convergence. In view of the relative difficulty of the S2Looking dataset, it was felt that it would need a larger training set. The relative proportions of 70%, 10%, and 20% are also widely used in other deep-learning-based studies.

Evaluation Metrics. In remote sensing change detection, the goal is to infer changed areas between bitemporal images. To this end, we took three-channel multispectral image pairs as the input, then output a single-channel prediction map. The Label 1 and Label 2 maps, when combined into one map, form a pixel-precise ground-truth label map. The performance of a change-detection method is reflected in the differences between the prediction maps and the ground-truth maps. To evaluate an algorithm's performance, Precision, Recall, and F1-scores can be used as evaluation metrics:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{2}$$

$$\text{F1-score} = \frac{2}{\text{Precision}^{-1} + \text{Recall}^{-1}} \tag{3}$$

Here, TP, FP, and FN correspond to the number of true-positive, false-positive, and false-negative predicted pixels for class 0 or 1. These are standard metrics used in the evaluation of change-detection algorithms [68]. Both Precision and Recall have a natural interpretation when it comes to change detection. Recall can be viewed as a measure of the effectiveness of the method in identifying the changed regions. Precision is a measure of the effectiveness of the method at excluding irrelevant and unchanged structures from the prediction results. The F1-score provides an overall evaluation of the prediction results; the higher the value, the better.

Training Process. For their implementation, we followed the default settings for each of the classic and state-of-the-art methods. Due to memory limitations, we kept the size of the original input images to 1024 × 1024 for the classic methods (FC-EF, FC-Siam-Conc, FC-Siam-Diff, and DTCDSCN) and cropped the images to 256 × 256 for the state-of-the-art methods (STANet-Base, STANet-BAM, STANet-PAM, CDNet, and BiT). It can be seen from Figure 8 that the performance of the model improves when iteration number N increases. We also observe that the marginal benefit on the model performance is declining with the number of instances increasing. Moreover, there is an upper limit for N when the image does not have more space to superimpose more building instances. Therefore, depending on the recommended settings for each of the methods, we set maximum number N = 50 for the classic methods and N = 180 for the state-of-the-art methods. We trained the detection models on a system with a Tesla P100 GPU accelerator and an RTX 2080 Ti graphics card. All of the methods were implemented with PyTorch.

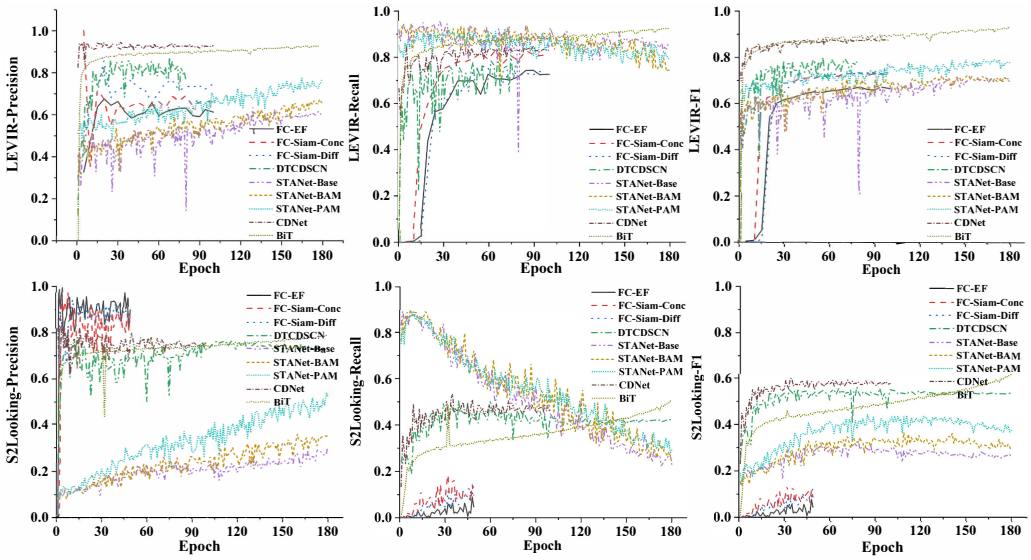

Figure 8. Results of fine-tuning the algorithms on the LEVIR-CD+ and S2Looking datasets. The top row presents the Precision, Recall, and F1-score metrics for the LEVIR-CD+ dataset. The bottom row presents the metrics for the S2Looking dataset.

3.2. Benchmark Results

Figure 8 shows the results of fine-tuning the various change-detection algorithms on the LEVIR-CD+ and S2Looking datasets.

Each algorithm took more epochs to converge and obtained lower F1-scores on S2Looking than on LEVIR-CD+. The evaluation metrics and visualizations of the results of the remote sensing change detection for all methods and dataset categories are presented in Table 3 and Figure 9, respectively. The F1-scores of the evaluated methods for the S2Looking dataset were at least 25% lower than the scores for LEVIR-CD+. This confirms that S2Looking presents a far more difficult challenge than LEVIR-CD+.

We also conducted an experiment where only a subset of the pixels with angles close to on-nadir, i.e., $\pm 15^{\circ}$, were used. This subset contained 3597 image pairs. The evaluated methods performed much better on the on-nadir subset of S2Looking, helping us to quantify the effect of using strongly off-nadir pixels. The evaluation metrics for the on-nadir subset are presented in Table 4. The F1-scores of the evaluated methods were about 46.4% greater than their average scores for the overall dataset. This confirms that the principal difficulty confronting change-detection algorithms when using the S2Looking dataset arises from the side-looking images.

Table 3. Results for the evaluated methods.

Method	LEVIR-CD+			S2Looking		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
FC-EF[16]	0.6130	0.7261	0.6648	0.8136	0.0895	0.0765
FC-Siam-Conc [16]	0.6624	0.8122	0.7297	0.6827	0.1852	0.1354
FC-Siam-Diff [16]	0.7497	0.7204	0.7348	0.8329	0.1576	0.1319
DTCDSCN [17]	0.8036	0.7503	0.7760	0.6858	0.4916	0.5727
STANet-Base [23]	0.6214	0.8064	0.7019	0.2575	0.5629	0.3534
STANet-BAM [23]	0.6455	0.8281	0.7253	0.3119	0.5291	0.3924
STANet-PAM [23]	0.7462	0.8454	0.7931	0.3875	0.5649	0.4597
CDNet [33]	0.8896	0.7345	0.8046	0.6748	0.5493	0.6056
BiT [34]	0.8274	0.8285	0.8280	0.7264	0.5385	0.6185

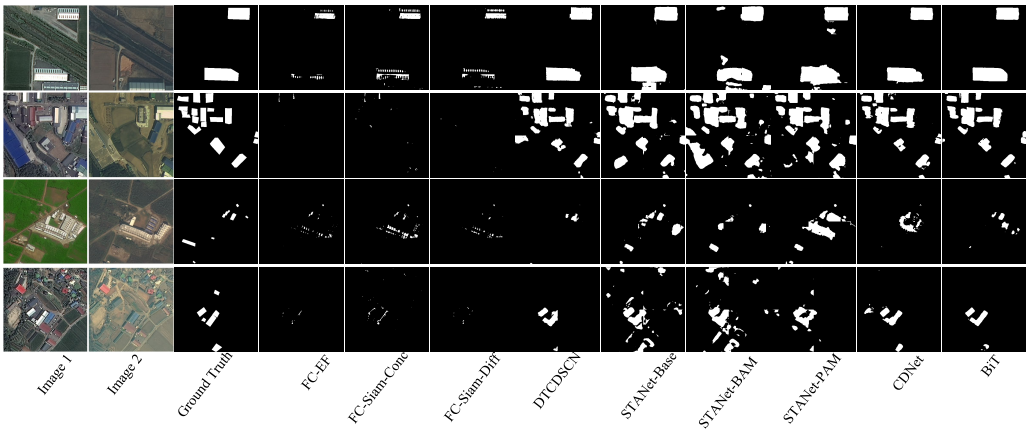

Figure 9. Visualizations of the results of the different methods on the S2Looking dataset.

Table 4. Evaluation metrics of on-nadir subset of S2Looking.

	Precision	Recall	F1-Score
FC-EF	0.7605	0.1155	0.1825
FC-Siam-Conc	0.7461	0.1663	0.2541
FC-Siam-Diff	0.6609	0.115	0.1749
DTCDSCN	0.7403	0.6155	0.6436
STANet-Base	0.3852	0.7344	0.4822
STANet-BAM	0.4366	0.7215	0.5206
STANet-PAM	0.5103	0.7477	0.5865
CDNet	0.8036	0.7375	0.7545
BiT	0.8512	0.733	0.7738

4. Discussion

A comparison of the results of the evaluation on the benchmark set for the two datasets is able to reveal the current failings of various algorithms with regard to how they handle the challenges posed by S2Looking. This, in turn, can provide insights into how they might be improved. We, therefore, now look at the evaluation results for each individual method.

FC-Net [16] consists of three different models, namely, FC-EF, FC-Siam-Conc, and FC-Siam-Diff. As shown in Table 3, FC-Siam-Conc and FC-Siam-Diff performed better on both datasets than FC-EF. FC-Net performed poorly on S2Looking compared to LEVIR-CD+. This is because the structure of FC-Net is too simple to be effectively trained on the complex problems presented by the S2Looking dataset. Unlike the DTCDSCN model, which has 512 channels, the deepest layer of FC-Net has only 128 channels. As a result, the ability of FC-Net to capture feature representations is limited. From the predictions produced for the S2Looking test set, we can see that the change-detection performance of FC-Net was largely premised upon image contrast, which fails to recognize building structures. This results in the detection of small objects with strong contrast, such as white cars, darkened windows, and the shadows of changed buildings, rather than whole building boundaries. This is evident in Figure 9.

DTCDSCN [17] is also a Siamese network, but it combines the task of change detection with semantic detection. DTCDSCN contains a change-detection module and two semantic segmentation modules. It also has an attention module to improve its feature-representation capabilities. Compared with FC-Net, DTCDSCN was better able to identify changed regions and was more robust to side-looking effects and building shadows. In addition, DTCDSCN was better at detecting small changes in building boundaries between the bitemporal images. Therefore, DTCDSCN performed much better than FC-Net on the S2Looking dataset. However, DTCDSCN failed to recognize a number of small prefabricated houses, as can be seen in the third and fourth rows of Figure 9. Additionally, because of the complexity of the rural scenes, some large greenhouses, cultivated land, and hardened ground were misrecognized as changed buildings.

STANet [23] is a Siamese-based spatiotemporal attention network designed to explore spatial–temporal relationships. Its design includes a base model (STANet-Base) that uses a weight-sharing CNN to extract features and measure the distance between feature maps to detect changed regions. STANet also includes a basic spatiotemporal attention module (STANet-BAM) and a pyramid spatiotemporal attention module (STANet-PAM) that can capture multiscale spatiotemporal dependencies. As shown in Table 3, STANet-PAM performed better (for its F1-score) than STANet-BAM and STANet-Base on both datasets. We also found that STANet had a relatively high Recall but low Precision compared to other methods. This may be because the batch-balanced contrastive loss that it employs in the training phase gives more weight to the misclassification of positive samples (change), resulting in the model having a tendency to make more positive predictions. Note that STANet-PAM performed better than DTCDSCN on the LEVIR-CD+ dataset but worse than DTCDSCN on the S2Looking dataset. From this, we conclude that STANet is more vulnerable to side-looking effects and illumination differences, which are more severe in the S2Looking dataset. Thus, it was more frequently misrecognizing the sides of buildings as building changes, which influenced the FP value in Eq. 1 and reduced its Precision.

CDNet [33] is a simple, yet effective, model for building-change detection. It uses a feature extractor (a UNet-based deep Siamese fully convolutional network) to extract image features from a pair of bitemporal patches. It also has a metric module to calculate bitemporal feature differences and a relatively simple classifier (a shadow fully convolutional network), which can produce change-probability maps from the feature difference images. CDNet produced better detection results than the previous methods on both datasets. This may be because the structure of CDNet, including its deep-feature extractor, can better handle moderate illumination variances and small registration errors, enabling it to produce high-resolution, high-level semantic feature maps. Note in Figure 9, for instance, that CDNet was robust in relation to misregistered hilly regions. However, it

failed to predict some small prefabricated houses and structures that appeared bright in one bitemporal image and dim in another, which were misrecognized as changed buildings. CDNet has a pixelwise change-discrimination process that is performed on the two feature maps. This model is not well-equipped to deal with large side-looking angles.

BiT [34] is an efficient change-detection model that leverages transformers to model global interactions within the bitemporal images. As with CDNet, the basic model [34] has a feature extractor and a prediction head. Unlike CDNet, however, it has a unique module (a bitemporal image transformer) that can enhance features. BiT consists of a Siamese semantic tokenizer to generate a compact set of semantic tokens for each bitemporal input, a transformer encoder to model the context of semantic concepts into token-based spacetime, and a Siamese transformer decoder to project the corresponding semantic tokens back into the pixel space and thereby obtain refined feature maps. Table 3 shows that BiT outperformed all the other methods on the two datasets. Only small prefabricated houses on hills were misrecognized, due to the lower registration accuracy, as can be seen in the fourth row of Figure 9. Most incorrect pixel predictions produced by BiT were due to side-looking effects associated with the expansion of building boundaries, which make building boundaries harder to accurately recognize in remote-sensing images.

Consequently, more sophisticated change-detection models are going to be needed to efficiently tackle the challenges posed by the S2Looking dataset. It should be noted that the change-detection methods here evaluated were all basically robust to seasonal and illumination variations in S2Looking, therefore building-change detection using the dataset is a solvable problem.

5. Conclusions

This paper has introduced the S2Looking dataset, a novel dataset that makes use of the camera scroll facility offered by modern optical satellites to assemble a large collection of bitemporal pairs of side-looking remote-sensing images of rural areas for building-change detection. The overall goal was to create a new baseline against which change-detection algorithms can be tested, thereby driving innovation and leading to important advances in the sophistication of change-detection techniques. A number of classic and state-of-the-art change-detection methods were evaluated on a benchmark test that was established for the dataset. To assess the extent to which the S2Looking dataset has added to the challenges confronting change-detection algorithms, deep-learning-based change-detection methods were applied to both S2Looking and the expanded LEVIR-CD+ dataset, the most challenging change-detection dataset previously available. The results show that contemporary change-detection methods perform much less well on the S2Looking dataset than on the LEVIR-CD+ dataset (by as much as 25% lower in terms of their F1-score). Analysis of the performance of the various methods enabled us to identify potential weaknesses in their change-detection pipelines that may serve as a source of inspiration for further developments.

There are several things that could to be actioned to improve upon the S2Looking dataset. First of all, the level of variation in the captured image pairs makes it difficult for a registration algorithm to match feature points between them. Improving current registration techniques would enhance the performance of change-detection models trained on the dataset. We therefore hope to see S2Looking driving a move towards innovative new image-registration methods and models. There are also flaws in current change-detection models, whereby irrelevant structures such as greenhouses and small reservoirs can be misidentified as building changes. Generally, by bringing this work to the attention of image-processing researchers with an interest in change detection, we feel that refinement of both the dataset and existing change-detection methodologies can be driven forward.

Author Contributions: All coauthors made significant contributions to the manuscript. L.S. conceptualized the problem and the technical framework, developed the collection of bitemporal data and analyzed the benchmark results, and wrote the paper. Y.L. conceptualized the problem and the technical framework, ran the annotation program, selected locations for satellite images, designed

the collection of bitemporal data and analyzed the dataset, developed the collection of bitemporal data, and made the specifications in the annotation. H.C. helped conceptualize the problem and the technical framework, helped develop specifications in annotation and trained annotator from AI data processing company, and conducted benchmark experiments with CDNet and BiT methods. H.W. ran benchmark experiments with FC-Net and DTCDSCN and analyzed the data, and managed training and test datasets for both LEVIR-CD+ and S2Looking. D.X. ran benchmark experiments with STANet and analyzed data, and conducted the process of registering the bitemporal images in the S2Looking Data Processing Pipeline. J.Y. ran benchmark experiments and analyzed data, and completed the process of registering. R.C. gave advice on the change detection problem, and reviewed and edited the submitted version of the article. S.L. make specifications in annotation and reexamined the quality of the dataset annotations. B.J. gave trained annotators from the AI data processing company to annotate the dataset, led the project and gave final approval of the version to be published. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. The datasets can be found here: https://github.com/S2Looking/.

Acknowledgments: We would like to thank Ao Zhang and Yue Zhang for collecting the rich satellite images. We thank Chunping Zhou, Dong Liu, Yuming Zhou from Beijing institution of remote sensing for helping on making specifications in annotation; We also thank Ying Yang and Yanyan Bai from Madacode company for their efforts in the annotation process.

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

References

1. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 574–586. [CrossRef]
2. Afify, H.A. Evaluation of change detection techniques for monitoring land-cover changes: A case study in new Burg El-Arab area. *World Pumps* **2011**, *50*, 187–195. [CrossRef]
3. Demir, B.; Bovolo, F.; Bruzzone, L. Updating Land-Cover Maps by Classification of Image Time Series: A Novel Change-Detection-Driven Transfer Learning Approach. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 300–312. [CrossRef]
4. Hégarat-Mascle, S.L.; Ottlé, C.; Guérin, C. Land cover change detection at coarse spatial scales based on iterative estimation and previous state information. *Remote Sens. Environ.* **2018**, *95*, 464–479. [CrossRef]
5. Brunner, D.; Lemoine, G.; Bruzzone, L. Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2403–2420. [CrossRef]
6. Singh, A. Review Article Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]
7. Tan, K.; Jin, X.; Plaza, A.; Wang, X.; Xiao, L.; Du, P. Automatic Change Detection in High-Resolution Remote Sensing Images by Using a Multiple Classifier System and Spectral—Spatial Features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3439–3451. [CrossRef]
8. Gómez-Candón, D.; López-Granados, F.; Caballero-Novella, J.J.; Pea-Barragán, J.M.; García-Torres, L. Understanding the errors in input prescription maps based on high spatial resolution remote sensing images. *Precis. Agric.* **2012**, *13*, 581–593. [CrossRef]
9. Zhang, Y.; Peng, D.; Huang, X. Object-Based Change Detection for VHR Images Based on Multiscale Uncertainty Analysis. *IEEE Geoence Remote Sens. Lett.* **2017**, *PP*, 1–5. [CrossRef]
10. Leichtle, T.; Gei, C.; Wurm, M.; Lakes, T.; Taubenbck, H. Unsupervised change detection in VHR remote sensing imagery—An object-based clustering approach in a dynamic urban environment. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *54*, 15–27. [CrossRef]
11. 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]
12. Tewkesbury, A.P.; Comber, A.J.; Tate, N.J.; Lamb, A.; Fisher, P.F. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sens. Environ.* **2015**, *160*, 1–14. [CrossRef]
13. 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]
14. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Li, H. DASNet: Dual attentive fully convolutional siamese networks for change detection of high resolution satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 1194–1206. [CrossRef]
15. 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]
16. 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. [CrossRef]

17. Liu, Y.; Pang, C.; Zhan, Z.; Zhang, X.; Yang, X. Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 811–815. [CrossRef]
18. 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]
19. Zhang, M.; Shi, W. A Feature Difference Convolutional Neural Network-Based Change Detection Method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7232-7246. [CrossRef]
20. 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]
21. Benedek, C.; SziráNyi, T. A Mixed Markov Model for Change Detection in Aerial Photos with Large Time Differences. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008.
22. Bourdis, N.; Marraud, B.; Sahbi, H. Constrained optical flow for aerial image change detection. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2011, Vancouver, BC, Canada, 24–29 July 2011.
23. 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]
24. Alkhelaiwi, M.; Boulila, W.; Ahmad, J.; Koubaa, A.; Driss, M. An Efficient Approach Based on Privacy-Preserving Deep Learning for Satellite Image Classification. *Remote Sens.* **2021**, *13*, 2221. [CrossRef]
25. Li, D.; Wang, M.; Jiang, J. China's high-resolution optical remote sensing satellites and their mapping applications. *Geo-Spat. Inf. Sci.* **2021**, *24*, 85–94. [CrossRef]
26. Habib, A.F.; Kim, E.M.; Kim, C.J. New Methodologies for True Orthophoto Generation. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 25–36. [CrossRef]
27. Weir, N.; Lindenbaum, D.; Bastidas, A.; Etten, A.V.; Tang, H. SpaceNet MVOI: A Multi-View Overhead Imagery Dataset. In Proceedings of the IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019.
28. Zhan, Y.; Fu, K.; Yan, M.; Sun, X.; Wang, H.; Qiu, X. Change Detection Based on Deep Siamese Convolutional Network for Optical Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1845–1849. [CrossRef]
29. 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]
30. Wang, M.; Tan, K.; Jia, X.; Wang, X.; Chen, Y. A Deep Siamese Network with Hybrid Convolutional Feature Extraction Module for Change Detection Based on Multi-sensor Remote Sensing Images. *Remote Sens.* **2020**, *12*, 205. [CrossRef]
31. Zhang, M.; Xu, G.; Chen, K.; Yan, M.; Sun, X. Triplet-Based Semantic Relation Learning for Aerial Remote Sensing Image Change Detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 266–270. [CrossRef]
32. Caye Daudt, R.; Le Saux, B.; Boulch, A. Fully Convolutional Siamese Networks for Change Detection. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4063–4067. [CrossRef]
33. Chen, H.; Li, W.; Shi, Z. Adversarial Instance Augmentation for Building Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5603216. [CrossRef]
34. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection With Transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, 1–14. [CrossRef]
35. Kaya, M.; Bilge, H.?. Deep Metric Learning: A Survey. *Symmetry* **2019**, *11*, 1066. [CrossRef]
36. Walsh, J.; Mahony, N.O.; Campbell, S.; Carvalho, A.; Riordan, D. Deep Learning vs. Traditional Computer Vision. In Proceedings of the Computer Vision Conference (CVC) 2019, Las Vegas, NV, USA, 2–3 May 2019.
37. Jiang, H.; Hu, X.; Li, K.; Zhang, J.; Gong, J.; Zhang, M. PGA-SiamNet: Pyramid Feature-Based Attention-Guided Siamese Network for Remote Sensing Orthoimagery Building Change Detection. *Remote Sens.* **2020**, *12*, 484. [CrossRef]
38. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Computer Vision & Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
39. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.
40. Rasouli, A.; Kotseruba, I.; Kunic, T.; Tsotsos, J. PIE: A Large-Scale Dataset and Models for Pedestrian Intention Estimation and Trajectory Prediction. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2020.
41. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD—A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
42. Shao, S.; Li, Z.; Zhang, T.; Peng, C.; Sun, J. Objects365: A Large-Scale, High-Quality Dataset for Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
43. Zi, B.; Chang, M.; Chen, J.; Ma, X.; Jiang, Y. WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection. In Proceedings of the MM '20: The 28th ACM International Conference on Multimedia, New York, NY, USA, 12–16 October 2020.
44. Shao, D.; Zhao, Y.; Dai, B.; Lin, D. FineGym: A Hierarchical Video Dataset for Fine-grained Action Understanding. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

45. Min, W.; Liu, L.; Wang, Z.; Luo, Z.; Wei, X.; Wei, X.; Jiang, S. ISIA Food-500: A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network. In Proceedings of the 28th ACM International Conference on Multimedia, New York, NY, USA, 12–16 October 2020; pp. 393–401. [CrossRef]
46. Wang, P.; Jiao, B.; Yang, L.; Yang, Y.; Zhang, S.; Wei, W.; Zhang, Y. In Proceedings of the Vehicle Re-identification in Aerial Imagery: Dataset and Approach. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
47. Mandal, M.; Kumar, L.K.; Vipparthi, S.K. MOR-UAV: A Benchmark Dataset and Baselines for Moving Object Recognition in UAV Videos. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20), New York, NY, USA, 12–16 October 2020; pp. 2626–2635. [CrossRef]
48. Liu, J.; Ji, S. A Novel Recurrent Encoder-Decoder Structure for Large-Scale Multi-View Stereo Reconstruction From an Open Aerial Dataset. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, 13–19 June 2020; pp. 6049–6058. [CrossRef]
49. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–17209. [CrossRef]
50. Gupta, R.; Hosfelt, R.; Sajeev, S.; Patel, N.; Goodman, B.; Doshi, J.; Heim, E.; Choset, H.; Gaston, M. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. *arXiv* **2019**, arXiv:1911.09296.
51. Chen, S.A.; Escay, A.; Haberland, C.; Schneider, T.; Staneva, V.; Choe, Y. Benchmark Dataset for Automatic Damaged Building Detection from Post-Hurricane Remotely Sensed Imagery. *arXiv* **2018**, arXiv:1812.05581.
52. Hu, Q.; Yang, B.; Khalid, S.; Xiao, W.; Trigoni, N.; Markham, A. Towards Semantic Segmentation of Urban-Scale 3D Point Clouds: A Dataset, Benchmarks and Challenges. *arXiv* **2021**, arXiv:2009.03137.
53. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A Benchmark Dataset for Fine-grained Object Recognition in High-Resolution Remote Sensing Imagery. *arXiv* **2021**, arXiv:2103.05569.
54. Pepe, M.; Costantino, D.; Alfio, V.S.; Vozza, G.; Cartellino, E. A Novel Method Based on Deep Learning, GIS and Geomatics Software for Building a 3D City Model from VHR Satellite Stereo Imagery. *ISPRS Int. J.-Geo-Inf.* **2021**, *10*, 697. [CrossRef]
55. Gao, L.; Shi, W.; Zhu, J.; Shao, P.; Sun, S.; Li, Y.; Wang, F.; Gao, F. Novel Framework for 3D Road Extraction Based on Airborne LiDAR and High-Resolution Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 4766. [CrossRef]
56. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-end Two-dimensional CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3–13. [CrossRef]
57. Hernández-López, D.; Piedelobo, L.; Moreno, M.A.; Chakhar, A.; Ortega-Terol, D.; González-Aguilera, D. Design of a Local Nested Grid for the Optimal Combined Use of Landsat 8 and Sentinel 2 Data. *Remote Sens.* **2021**, *13*, 1546. [CrossRef]
58. Singh, K.K.; Frazier, A.E. A meta-analysis and review of unmanned aircraft system (UAS) imagery for terrestrial applications. *Int. J. Remote Sens.* **2018**, *39*, 5078–5098. [CrossRef]
59. Miyamoto, T.; Yamamoto, Y. Using Multimodal Learning Model for Earthquake Damage Detection Based on Optical Satellite Imagery and Structural Attributes. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020.
60. Yamazaki, F.; Kouchi, K.I.; Kohiyama, M.; Muraoka, N.; Matsuoka, M. Earthquake damage detection using high-resolution satellite images. In Proceedings of the Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004.
61. Mi, W.; Chen, C.; Pan, J.; Ying, Z.; Chang, X. A Relative Radiometric Calibration Method Based on the Histogram of Side-Slither Data for High-Resolution Optical Satellite Imagery. *Remote Sens.* **2018**, *10*, 381.
62. Wang, M.; Cheng, Y.; Tian, Y.; He, L.; Wang, Y. A New On-Orbit Geometric Self-Calibration Approach for the High-Resolution Geostationary Optical Satellite GaoFen4. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1670–1683. [CrossRef]
63. Poli, D.; Toutin, T. Review of developments in geometric modelling for high resolution satellite pushbroom sensors. *Photogramm. Rec.* **2012**, *27*, 58–73. [CrossRef]
64. Barazzetti, L.; Brumana, R.; Cuca, B.; Previtali, M. Change detection from very high resolution satellite time series with variable off-nadir angle. In Proceedings of the SPIE the International Society for Optical Engineering, Paphos, Cyprus, 16–19 March 2015.
65. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
66. Parente, L.; Chandler, J.H.; Dixon, N. Automated Registration of SfM-MVS Multitemporal Datasets Using Terrestrial and Oblique Aerial Images. *Photogramm. Rec.* **2021**, *36*, 12–35. [CrossRef]
67. Ramu, G.; Babu, S. Image forgery detection for high resolution images using SIFT and RANSAC algorithm. In Proceedings of the International Conference on Communication & Electronics Systems, Coimbatore, India, 19–20 October 2017; pp. 850–854.
68. Radke, R.J.; Andra, S.; Al-Kofahi, O.; Roysam, B. Image change detection algorithms: A systematic survey. *IEEE Trans. Image Process.* **2005**, *14*, 294–307. [CrossRef] [PubMed]