



Article

HSAA-CD: A Hierarchical Semantic Aggregation Mechanism and Attention Module for Non-Agricultural Change Detection in Cultivated Land

Fangting Li ^{1,2}, Fangdong Zhou ^{3,*}, Guo Zhang ¹ , Jianfeng Xiao ³ and Peng Zeng ⁴

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; lifangting@whu.edu.cn (F.L.); guozhang@whu.edu.cn (G.Z.)

² Hubei Institute of Photogrammetry and Remote Sensing, Wuhan 430074, China

³ Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China; jfxiao@mails.cnu.edu.cn

⁴ Hunan Institute of Land and Resources Planning, Changsha 410007, China; cyzeng@mails.cnu.edu.cn

* Correspondence: fdzhou@mails.cnu.edu.cn

Abstract: Cultivated land plays a fundamental role in the sustainable development of the world. Monitoring the non-agricultural changes is important for the development of land-use policies. A bitemporal image transformer (BIT) can achieve high accuracy for change detection (CD) tasks and also become a key scientific tool to support decision-making. Because of the diversity of high-resolution RSIs in series, the complexity of agricultural types, and the irregularity of hierarchical semantics in different types of changes, the accuracy of non-agricultural CD is far below the need for the management of the land and for resource planning. In this paper, we proposed a novel non-agricultural CD method to improve the accuracy of machine processing. First, multi-resource surveying data are collected to produce a well-tagged dataset with cultivated land and non-agricultural changes. Secondly, a hierarchical semantic aggregation mechanism and attention module (HSAA) bitemporal image transformer named HSAA-CD is performed for non-agricultural CD in cultivated land. The proposed HSAA-CD added a hierarchical semantic aggregation mechanism for clustering the input data for U-Net as the backbone network and an attention module to improve the feature edge. Experiments were performed on the open-source LEVIR-CD and WHU Building-CD datasets as well as on the self-built RSI dataset. The F1-score, intersection over union (IoU), and overall accuracy (OA) of these three datasets were 88.56%, 84.29%, and 68.50%; 79.84%, 73.41%, and 59.29%; and 98.83%, 98.39%, and 93.56%, respectively. The results indicated that the proposed HSAA-CD method outperformed the BIT and some other state-of-the-art methods and proved to be suitable accuracy for non-agricultural CD in cultivated land.

Keywords: change detection; non-agricultural change detection; bitemporal image transformer; high-resolution remote-sensing images



Citation: Li, F.; Zhou, F.; Zhang, G.; Xiao, J.; Zeng, P. HSAA-CD: A Hierarchical Semantic Aggregation Mechanism and Attention Module for Non-Agricultural Change Detection in Cultivated Land. *Remote Sens.* **2024**, *16*, 1372. <https://doi.org/10.3390/rs16081372>

Academic Editor: Salah Bourennane

Received: 4 March 2024

Revised: 1 April 2024

Accepted: 10 April 2024

Published: 13 April 2024



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

1. Introduction

The development of urbanization and the acceleration of human socioeconomic activities have led to a continuous decrease in cultivated land worldwide, which brings challenges to the rational utilization and protection of land resources [1]. Especially in China, the rapid urbanization process in the past decade has resulted in a continuous decline in arable land. The occupation of non-agriculturalization is the main factor for this change, including built-up areas, green corridors that exceed standards, decorative lakes, nature reserves, and non-agricultural constructions [2]. These non-agricultural changes threaten food security and pose a serious obstacle to sustainable development. The government and relevant ministries in China have issued relevant documents to rectify the non-agricultural occupation of farmland [3].

In general, non-agricultural change detection (CD) methods mainly include two approaches, namely traditional and remote-sensing-based methods. The traditional method of land survey requires a lot of human, material, and financial resources and takes a relatively long time to reveal the changes. The national land survey for land-cover/land-use investigation in China is conducted every ten years or even longer, and the annual regional surveys cannot keep pace with the real-time changes [4]. In recent years, CD from remote-sensing imagery has become a research hotspot and has been applied in many fields. With the widespread application of high-resolution remote-sensing images (RSIs), land-use-change research can bring higher precision and more efficient ways to detect non-agricultural changes in cultivated land. Domestic HRRSIs can be used to identify and detect water bodies, vegetable greenhouses, buildings, and road infrastructures [4].

Recently, the deep-learning methods for multiple CD methods have gained much attention because they take the high-level features of high-resolution RSIs into account, and weaken the influence of intrinsic spectral variability on CD to a certain extent. As a representative of deep-learning methods, a neural network is a cost-effective, time-efficient, short-cycle, and accurate approach. First, the results of land-cover/land-use change (LUCC) by using time series high-resolution RSIs can provide timely basic data for CD [5]. Secondly, the commonly used CNN methods can effectively reduce misjudgments due to the land-use classification standards varying in different periods. Finally, the sample libraries and the publicly available datasets under different classification standards can provide a fundamental basis for non-agricultural CD in cultivated land.

However, there are differences between non-agricultural CD and typical remote-sensing monitoring. On the one hand, there are various types of non-agricultural changes in cultivated land, but there is currently no corresponding dataset that can cover the non-agricultural CD task. The open-source datasets may get low accuracy when they are directly used for non-agricultural CD. On the other hand, the deep-learning methods used for LUCC can classify and identify specific types, but cannot aggregate multiple types into a spatial pattern. Thus, the existing CD methods are not suitable to be directly used for non-agricultural CD [6].

The main contributions of this paper can be summarized as follows:

- A hierarchical semantic structure of land-use types for non-agricultural changes in cultivated land is established, and the relationships between different types of changes are analyzed. Taking Hubei Province in China as an example, we select data from five regions to construct a dataset for detecting such changes. The dataset is suitable for cultivated land-change detection. We hope that the dataset will contribute to the innovation of farmland-change monitoring methods and their application
- Aiming at the problems of scattered results and disordered hierarchies in current networks for non-agricultural CD, a hierarchical semantic aggregation mechanism and attention module (HSAA) is proposed. The scattered classification results are aggregated by adding a semantic aggregation layer, and the aggregated types are enhanced by an attention mechanism, thus the accuracy of CD is further improved.

The rest of this article is organized as follows. Section 2 introduces the related work. Section 3 describes the overall structure of the proposed method. Section 4 presents the experimental datasets and results. Section 5 discusses the major findings and limitations of the study. Finally, a brief conclusion is presented in Section 6.

2. Related Work

In this section, we briefly reviewed the corresponding datasets and deep-learning methods for change detection. The limitations of current methods are then discussed.

2.1. RSI Datasets for Change Detection (CD)

Very high-resolution (VHR) images contain detailed spatial information and are often used in multiple change detection (CD) tasks. Thus, a series of RSI datasets have been constructed for multiple purposes and different application scenarios, including land-use

classification, semantic segmentation, object recognition, and CD. The information on existing RSI datasets is shown in Table 1. According to data sources, RSI can be mainly classified into aerial and satellite images, based on the aerial RSI dataset, multiple image pairs for CD are constructed by two or more images of the same region at different times.

Table 1. Information of the existing datasets for change detection.

Dataset	Image Size	Resolution	Number of Images Pairs	Tasks and Change Types	Data Source	Time Span
SZTAKI [7]	952 × 640	1.5 m	13	Built-up regions, buildings, planting of trees, etc.	Aerial image, FOMI, Google Earth	2000–2005, 2000–2007, 1984–2007
AICD [8]	800 × 600	0.5 m	1000	trees and buildings, etc.	Aerial images	/
WHU-Building [9]	32,207 × 15,354	0.2 m	16,077	Buildings	Aerial images	2012–2016
SYSU-CD [10]	256 × 256	0.5 m	20,000	Urban buildings, change of vegetation	Aerial Images	2007–2014
LEVIR_CD [11]	1024 × 1021	0.5 m	637	BCD tasks	Google Earth	2002–2018
DSIFN [12]	512 × 512	2 m	442	City Area change	Google Earth	2001–2018
GZCD [13]	1006 × 1168–4936 × 5224	0.55 m	19	BCD tasks	Google Earth	2013–2017
OSCD [10,14]	600 × 600	10 m	24	Urban growth changes	Sentinel-2 satellites Multispectral images	2015–2018
HRCUS-CD [15]	256 × 256	0.5 m	11,388	Built-up areas and new urban areas. BCD tasks	Satellite image	2010–2018 2019–2022
S2Looking-CD [16]	1024 × 1024	0.5–0.8 m	5000	BCD tasks	Satellite image	10 years
WXCD [17]	7840 × 6160	0.2/0.5 m	/	BCD tasks	UAV/SuperView-1	2012–2018
SVCD [18]	256 × 256	0.03–1 m	16,000	Object detection	Synthetic and real images	/

For some of the datasets, the time of the data source is not mentioned in the original article, and in the Time Span column of the table we denote it with “/”.

The first type of dataset is constructed by aerial images. The SZTAKI Air Change Benchmark Set [7] is the earliest widely used CD dataset. It has 13 pairs of 952 × 640 pixels optical aerial images with a spatial resolution of 1.5 m. The aerial image change detection (AICD) dataset [8] is built by 100 simulated scenes by realistic aerial images rendered artificially, which includes 1000 pairs of images 800 × 600 pixels in size, with a resolution of 0.5 m, and containing major change objects such as forest trees and buildings. The WHU-building CD dataset [9] consists of a pair of aerial images containing an aerial image of 12,796 buildings in 2012 and an aerial image of 16,077 buildings in 2016, 32,207 × 15,354 pixels in size with a resolution of 0.2 m. The Sun Yat-Sen University (SYSU-CD) dataset [10] consists of 800 pairs of images 1024 × 1024 pixels in size with a resolution of 0.5 m, which were captured in Hong Kong in 2007 and 2014, respectively. The SYSU-CD dataset contains a number of different sorts of modifications, including newly constructed urban buildings, suburban dilatation, groundwork before construction, change in vegetation, road extension, and sea development.

The second type of dataset is built from Google Earth images. The LEVIR-CD dataset [11] includes 637 pairs of RSIs 1024 × 1024 pixels in size with a resolution of 0.5 m and is used in building-change detection (BCD) tasks very frequently. The DSIFN dataset [12] is manually collected from Google Earth and contains 442 pairs of images 512 × 512 pixels in size with a resolution of 2 m from six Chinese cities, including Beijing, Shenzhen, Chongqing, Wuhan, Chengdu, and Xi’an. The Google dataset for CD (GDSCD) [13] contains 19 pairs of satellite images with a resolution of 0.55 m. The image pairs range in size and year, spanning from 1006 × 1168 pixels to 4936 × 5224 pixels and spanning the period 2006–2019.

The third type of dataset is constructed by satellite images. The Onera satellite change detection (OSCD) [14] dataset collects 24 pairs of Sentinel-2 multispectral satellite images taken between 2015 and 2018, approximately 600 × 600 pixels in size with a resolution of 10-m. The S2Looking-Cd dataset [16] comprises expansive side-looking satellite images taken at different off-nadir angles, and it includes approximately 65,920 annotated examples of angles. The HRCUS-CD dataset [15] contains 11,388 pairs of cropped HRRSIs

256 × 256 pixels in size with a resolution of 0.5 m as well as more than 12,000 labeled change instances. All the labels are manually annotated by experienced annotators in RSI interpretation. This dataset was collected in Zhuhai, China, which covers an area of 1736.45 km² and with a resident population of approximately 2.44 million (as of November 2020).

The other type of CD dataset is constructed by multiple sources of images, such as unmanned aerial vehicle (UAV) images and syntheses images. The WXCD dataset [17] is handcrafted from UAV and SuperView-1 (SV-1) images as the original image in the RGB band, and the building areas with significant changes in two temporal images are manually annotated as vectors using ArcGIS software 10.2 and converted to the Tiff format. The unique natural and human environment of the site, its complex and diverse building forms and scenes, and the sensor characteristics of the bitemporal images with different shooting angles and lighting variations pose a greater challenge to the building CD task than other publicly available datasets. Season-varying change detection (SVCD) [18] consists of two types of variations: real RSIs and synthetic images. The most commonly used is the real RSIs with seasonal variations, which contain 16,000 pairs of images with a resolution of 0.03–1 m and a size of 256 × 256 pixels.

The application focuses on different datasets. For example, datasets for BCD tasks provide more accurate image pairs of buildings before and after the change to support high-precision detection of building changes, such as LEVIR-CD, WHU-CD, HRCUS-CD, etc. In addition, detecting changes in land features and land types is also the main support function of the dataset. Detection of changing objects and geographical meanings is also an important direction for recent applications, such as WXCD.

2.2. Deep-Learning Methods for Change Detection (CD)

The classical convolutional neural networks (CNNs) were applied in CD for land-cover/land-use tasks, such as the Siamese CNNs [19] and axial cross-attention branch fusion networks [20]. USCD-MiBi [21], a simple and practical unsupervised change-detection method based on multi-indices and bitemporal remote-sensing image pairs, allows users to choose the most suitable change indices to deal with large-scale LUCC accurate change detection. The COCRF framework [22], consisting of a binary CD task and a multiclass CD task, is proposed to reduce the influence of spectral variability. U-Net applied in CD for VHR images [23], such as Siamese NestedUNet (SNUNet) [24], have been comprehensively investigated for land-cover CD with bitemporal images. Buildings change detection (BCD) was one of the hot research issues, not only for the plentiful datasets but also the networks, such as Siam-EmNet [25], AERNet [15], CMGFNet [26], etc.

For the different paths to get the change from images. ChangeNet [27] processes the change detection by fusion of the transposed convolution and multiscale difference. STANet [11] applied a spatial attention mechanism to enhance the feature extraction ability. SNUNet-CD [18] combined the Siamese network and Nested UNet for multiscale feature fusion to resolve the problem of small object detection.

Lots of methods put the attention into consideration. Scholars have tried different loss functions to obtain better accuracy, including contrast loss, triplet loss, etc. Some existing approaches determine whether a change has occurred by comparing the parametric distances of dual-temporal image pixel pairs, and L1 and L2 distances are frequently used to determine whether changes have taken place. For instance, STANet [11] uses a pyramid attention mechanism to enhance dual-temporal images and reduce false detections due to alignment errors. DASNet [28] uses a unique dual attention mechanism to enhance the ability of the network with a newly designed unique loss function. DSAMNet [27] focuses on the pseudo change and noise problems in the CD process. The scale and relation-aware Siamese network (SARASNet) [29] proposed relation-aware, scale-aware, and cross-transformer modules to deal with spatial information and scaling changes between objects. The change guiding network (CGNet) [30] tackles the insufficient expression problem of change features in the conventional U-Net structure to remove the edge integrity and internal holes phenomenon of change features. The ECFNet [31] can better utilize the fine-

grained information in the multiscale feature map for result prediction, which can improve the detection performance of small objects and reduce the false detection around the edge pixels of change objects. SDMNet [32] deals with the problems of effectively distinguishing interesting changes and pseudo changes in high-resolution remote-sensing images and forming accurate and robust CD results. Denoising diffusion probabilistic models for change detection (DDPM-CD) [33] trains lightweight change detection classifiers on a large number of existing RSIs using diffusion model decoders to detect accurate changes. TINYCD directly shares the weight of the existing network to achieve the purpose of the Siamese network [34]. The method based on feature interaction and multitask learning (FMCD) [35] can improve the ability to detect changes in complex scenes, by modeling the context information of features through a multilevel feature interaction module, so as to obtain representative features, and to improve the sensitivity of the model to changes. The change gradient image (CGI) [36] first embeds a multiscale information attentional module in U-Net to achieve multiscale information and adds the position channel attention module to pay more attention to the spectral and spatial information in the multiscale fused feature map. A composite higher-order attention network with multiple encoding paths named MCHA-Net [37] can improve the generalizability and detection accuracy of the network and outperforms state-of-the-art methods in both visual interpretation and quantitative evaluation. An unsupervised single-temporal change detection framework based on intra- and inter-image patch exchange (I3PE) [38] allows for training deep-change detectors on unpaired and unlabeled single-temporal remote-sensing images that are readily available in real-world applications.

The bitemporal image transformer (BIT) [39] uses a transformer to model long-range context in bitemporal images, enhancing the discrimination of paired features. ChangeFormer [40] unified the layered transformer encoders and multilayer perceptual decoders into a Siamese network structure to effectively extract features required for change detection without CNN. H-SALENet [41], a hierarchical self-attention augmented Laplacian pyramid expanding network, combines a deep convolutional module with a hierarchical and long-range context augmentation module (HLAM) to extract the deep features of bitemporal images and a Laplacian pyramid expansion module (LPEM) to catch change information at different scales and reduce high-frequency information loss simultaneously. The dual-perspective change contextual network (DPCC-Net) [42] emphasizes the process of extraction and optimization of change features by bitemporal feature fusion and contextual modeling. An attention-based multiscale transformer network (AMTNet) [43] utilizes a CNN-transformer structure to address complex textures, seasonal variations, climate changes, and new requirements issues. This model employed attention and transformer modules to model contextual information in bitemporal images effectively [32]. The coarse-to-fine boundary refinement network (CBR-Net) [44] can accurately extract building footprints from remote-sensing imagery. A bitemporal remote-sensing image change detection network based on a Siamese-attention feedback architecture, referred to as SAFNet [45], a global semantic module (GSM) on the encoder network, generates a low-resolution semantic change map to capture the changed objects, a temporal interaction module (TIM) and two auxiliary modules—the change feature extraction module (CFEM) and the feature refinement module (FRM)—to learn the fine boundaries of the changed target. The SAFNet algorithm exhibits state-of-the-art performance. A network based on feature differences and attention mechanisms (DAFNet) [46] includes a Siamese architecture-encoding network that encodes bitemporal images, a difference feature-extraction module (DFEM) for extracting difference features from two periods, an attention-regulation module (ARM) with an enhanced attention mechanism to optimize the ability to extract difference features, and a cross-scale feature-fusion module (CSFM) for merging features from different encoding stages. This method effectively alleviates issues of target misdetection, false alarms, and blurry edges.

3. The Proposed Method

Our proposed method proposed two modules to enhance the BIT [39] method. In this section, the proposed architecture of HSAA-CD for non-agricultural change detection was presented.

3.1. Overview of the Proposed Architecture

In this paper, two different modules, FAM and CBAM, are applied to construct the architecture of HSAA-Net. As shown in Figure 1, the proposed framework consisted of two parts: a CNN backbone network and a bitemporal image transformer. The input of the network is a bitemporal remote-sensing image. Two ResNet50 [47] with shared weights are used as a feature extractor for the input bitemporal remote-sensing image pairs, aggregating different levels of features of each input image through the feature aggregation module for multiscale representation. Image feature maps extracted by ResNet are converted into semantic tokens by using the VT [48]. Then the tokens are fed into the transformer-encoder to obtain the global semantic information and capture long-range relations to comprehend the global semantic information for each temporal. Afterward, the transformer-decoder is used to project the corresponding semantic tokens into the pixel space to obtain deep features for each temporal refinement. We use CBAM in ResNet's bottle block to obtain more efficient features. Meanwhile, we use FAM to merge the feature information of the underlying semantics to generate the extracted primary feature maps. Then, the high-level feature maps obtained from the transformer-decoder are concatenated with the extracted primary feature maps from FAM. Finally, we fed the obtained variation feature map into PM to obtain the change map.

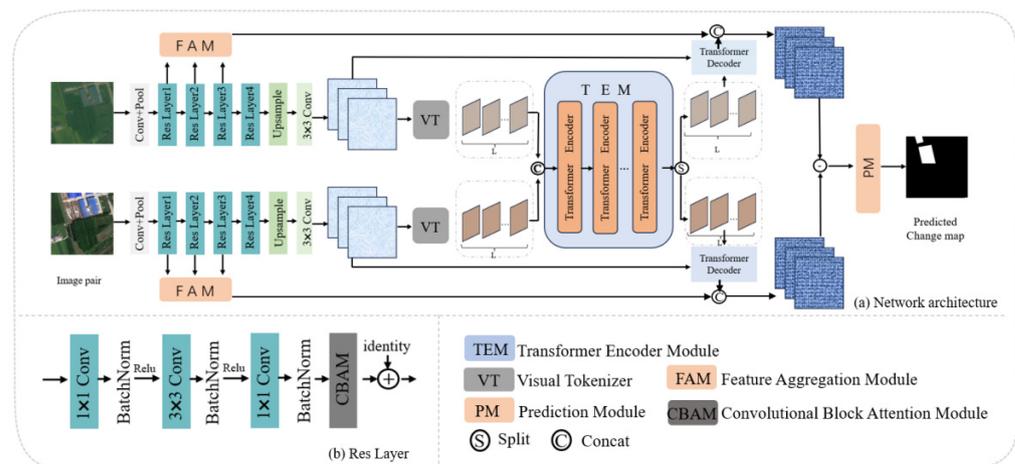


Figure 1. Overall structure of the proposed network model.

3.2. Feature Aggregation Module (FAM)

The feature aggregation module is used to extract a low-level feature map of the input remote-sensing image pair with the same network structure. The structure of the feature aggregation module is shown in Figure 2. In our method, each ResLayer contains three convolutional layers, three batch normalization layers, two rectified linear unit (ReLU) functions, and one CBAM layer. The specific approach of FAM is as follows: we input reslayer1, reslayer2, reslayer3, and the up-sampled feature maps into FAM, and the feature maps of each layer in FAM are processed by two 1×1 convolutional layers followed with BatchNorm (BN) action and a 3×3 convolutional layer. This is followed by BN action, one ReLU action, which concatenates the same size feature maps of four different layers, and finally passes through a LastConv layer. The LastConv layer contains two identical 1×1 convolutional layers with BN and ReLU action. Finally, the low-level feature maps are generated.

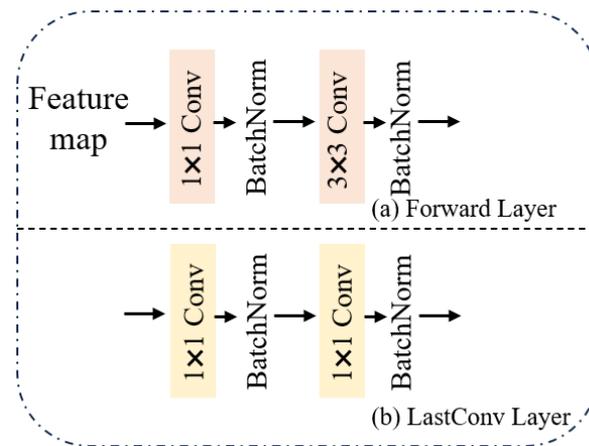


Figure 2. The overview of FAM. The module has two sub-layers: Forward and LastConv.

3.3. Convolutional Block Attention Module (CBAM)

In this paper, we improved ResNet50 by removing the fully-connected layers of the classical ResNet50 and adding CBAM modules to each of the remaining two convolutional layers and four residual layers (ResLayers) to extract bitemporal image feature maps. The convolutional block attention module (CBAM) [49] sequentially extrapolates intermediate feature maps along independent dimensions (channel and spatial) to infer attention maps and these attention maps are then multiplied with the input feature maps for adaptive feature optimization. The structure of the CBAM is shown in Figure 3. We integrate the CBAM module after the BatchNorm layer in ResNet’s bottleneck layer, and the outputs from the BatchNorm layer are subjected to the global max pooling and global average pooling based on width and height, respectively, and then passed through the MLP, respectively, and the MLP outputs are subjected to the element-wise feature optimization. The output features of MLP are subjected to element-wise summation and sigmoid activation to generate the final channel attention feature map. Subsequently, the channel attention feature map and input feature map are subjected to element-wise multiplication to generate. The channel attention feature map and input feature map are subsequently multiplied element-wise in order to generate the input features required by the spatial attention module.

$$\begin{aligned} \mathbf{M}_c(\mathbf{F}) &= \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))) \\ &= \sigma(\mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{\text{avg}}^c)) + \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{\text{max}}^c))) \end{aligned} \quad (1)$$

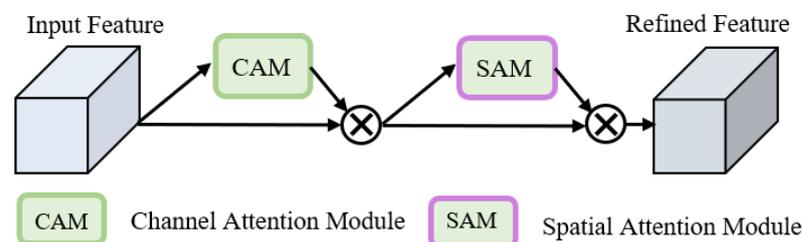


Figure 3. The overview of CBAM. The module includes two sequential sub-modules: Channel and Spatial.

The spatial attention module takes the output feature map of the CAM as the input feature map. Firstly, we perform a channel-based global max pooling and global average pooling, and then we perform a concatenate operation on these two results based on the channel. Then we convolve them to 1 channel, generate the spatial attention feature by

sigmoid, and finally multiply the feature with the input feature of this module to get the final generated feature.

$$\begin{aligned} \mathbf{M}_s(\mathbf{F}) &= \sigma(f^{7 \times 7}([\text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F})])) \\ &= \sigma\left(f^{7 \times 7}\left(\begin{bmatrix} \mathbf{F}_{\text{avg}}^s; \mathbf{F}_{\text{max}}^s \end{bmatrix}\right)\right) \end{aligned} \quad (2)$$

4. Experiments and Results

4.1. Datasets

4.1.1. Self-Built Image Dataset

The change-detection dataset applied to agriculture is relatively lacking. There are lots of reasons. The first one is the high costs of collecting, processing, and standardizing large-scale data. The second reason is that the types of changes detection in agricultural land vary greatly for climate, season, and the similarity of vegetation.

This article constructs a dataset for detecting non-agricultural changes in cultivated land using multi-source data, including a series of high-resolution satellite remote-sensing images from different satellites, the vector data of geographical and national monitoring results from the administering department, and the land-use classification results of the National Land Use Surveying. High-resolution satellite remote-sensing data was selected from domestic optical high-resolution satellite images in the third and fourth quarters of 2020, including the Resource Satellite series, High Resolution series, and Beijing 2 (Figure 4). The changes in the land cover of rice and corn agricultural land are particularly significant in these two seasons and are difficult to distinguish from changes in the surrounding environment. When we choose to detect in the datasets of these two seasons and achieve high accuracy, it indicates that the detection of non-agricultural changes in this region or similar environments has a relatively high adaptability. Source-specific information from the high-resolution satellite remote-sensing image data is presented in Table 2. The data covered two full-province coverage surveys of geographical conditions in 2015 and 2018 in Hubei Province. To ensure consistency in the resolution of images before and after the interpretation of sample changes, all images were uniformly corrected to 2 m resolution. Remote-sensing monitoring data of land use which contains 8 class 1 and 47 class 2 categories was downloaded from the Resources Environment Science and Data Center, and vector boundaries were extracted from multi-temporal land-use classification grid data to select the vector range of arable land, which was used as a reference to select the range of changes in arable land. Geographic country surface coverage classification system information is presented in Table 3.

Table 2. High-resolution satellite remote-sensing image data sources.

Satellite Name	Launch Time	Spectral Bands	Resolution (m)	Coverage Area	Orbit Information
ZY-3	January 2012	Visible, Near-Infrared	2.1–5.8	Global	Sun-synchronous orbit
GF-1	April 2013	Visible, Near-Infrared	2–16	Global	Sun-synchronous orbit
GF-2	August 2014	Visible, Near-Infrared	1–4	Global	Sun-synchronous orbit
GF-6	June 2018	Visible, Near-Infrared, Mid-Infrared	2–8	Global	Sun-synchronous orbit
GF-7	November 2019	Visible, Near-Infrared, Mid-Infrared	0.8–3.2	Global	Sun-synchronous orbit
BJ-1	September 2008	Visible, Near-Infrared	4	Global	Sun-synchronous orbit

Based on the consistency of data available, we select Qianjiang City, Shayang County, Zhijiang City, Qichun County, and Hong'an County of Hubei Province for the image dataset. The HRRSI were acquired in 2015 and 2018 and the change reports from satellite monitoring of land use nationwide in 2020 and 2021. The non-agricultural use of arable land includes occupying cultivated land to plant trees, building houses, lakes or water, building roads, building greenhouses, building photovoltaics, building landscape parks,

filling land, and building other agricultural facilities. Table 4 shows the 9 main types of arable land conversion to other land uses and the spectral characteristics of types of non-agricultural arable land changes.

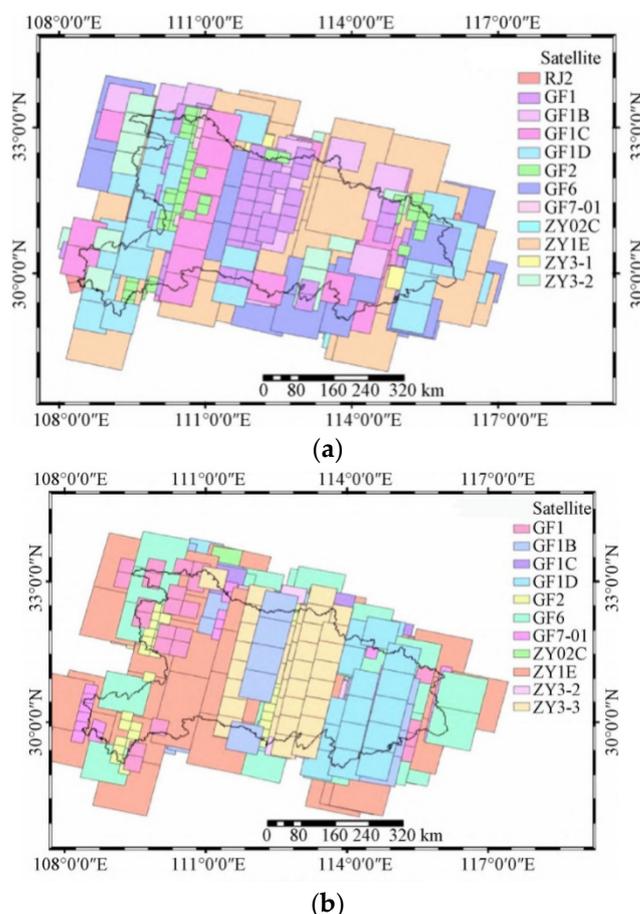
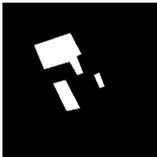
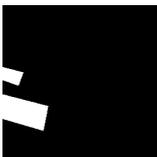
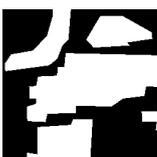
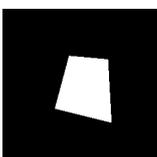


Figure 4. Image coverage in the third quarter and fourth quarter. (a) Image coverage in the third quarter and (b) Image coverage in the fourth quarter.

Table 3. Geographic country surface coverage classification system.

No.	Class I	Class II	
1	Buildings	Multi-story house building area, low house building area, abandoned house building area, multi-story and above independent house building, low building	Urban construction land
2	Railroads and Roads	Railroads, highways, city roads, country roads, ramps	Urban construction land
3	Structures	Hardened surfaces, hydraulic facilities, transportation facilities, city walls,	Urban construction land
4	Manually excavated land	Open-pit extraction sites, stockpiles Construction sites, other man-made stockpiles Paddy fields	Urban construction land
5	Cultivated Land	early land, orchards, tea plantations, mulberry plantations, rubber plantations, seedling paintings, flower weeks, other economic seedlings	Ecological living land
6	Forest and Grass Cover	Tree forests, shrub forests, mixed tree and shrub forests, bamboo forests, open forests, young planted forests, sparse shrubs and grasslands, natural grasslands, artificial grasslands	Ecological land
7	Water	Rivers, canals, lakes, reservoirs, lakes, glaciers and permanent snow cover	Ecological land
8	Deserts and Bare Ground	Deserts and bare ground Saline surface, clay surface, sandy surface, rocky surface Rocky surface	Ecological land

Table 4. Samples of Cultivated land transfer to other types.

Sample Type	Pre-Image	Post-Image	Ground Truth
To Forest and Grassland			
To Greenhouse			
To Buildings			
To Traffic Roads			
To Filling Land			
To Lake or Water			
To Photovoltaic Power Station			
To Park			
To Other Agricultural Facility Land			

In tasks with a limited number of training samples, data augmentation was critical for the invariance and robustness of the training network. In this experiment, the original image data and ground truth image data were preprocessed, including data cutting and

data augmentation. The original remote-sensing images were cut into 256×256 pixels blocks, sample label images were created, and then data augmentation processes such as rotation, displacement, random clipping, and random scaling were applied to the images. Finally, we obtained 2036 sample image pairs 256×256 pixels in size. The training, validation, and test sets of the dataset are 70%, 20%, and 10%, respectively.

Figure 5 displays part of the image maps in the dataset, Figure 5a,b show the original bitemporal image maps and Figure 5c shows the label images obtained for different land-use types.

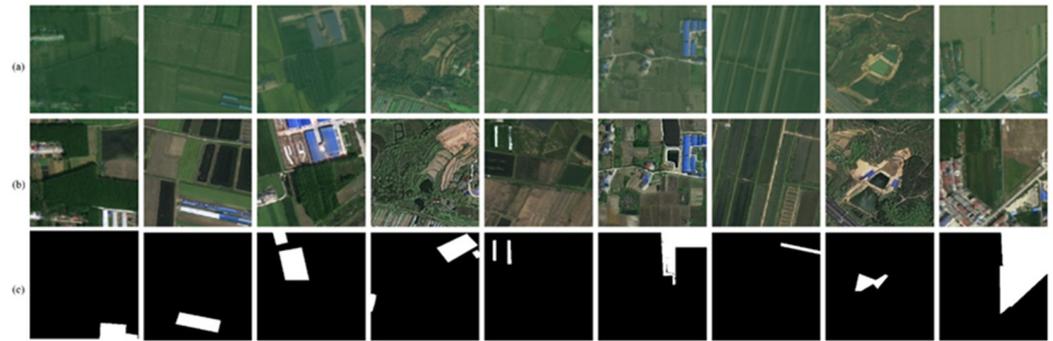


Figure 5. The AGRI-CD dataset: (a) Remote-sensing image at the moment of T1. (b) Remote-sensing image at the moment of T2. (c) Ground truth.

4.1.2. LEVIR-CD Dataset and WHU-CD Dataset

We also tested the proposed approach on two commonly used and high-quality RSI benchmark datasets: the LEVIR-CD Dataset and the WHU-CD Dataset.

The LEVIR-CD dataset was proposed by Beihang in 2020 for change detection in public buildings [11]. The dataset consists of 637 pairs of images, and the number of training, validation, and test dataset images in the dataset are 445, 64, and 128, respectively. The spatial resolution is 0.5 m/pixel, and the image size is 1024×1024 pixels. We crop the image from the original 1024×1024 pixels in the dataset into 16 sub-blocks 256×256 pixels in size without overlapping regions. Then we divide the dataset according to the same ratio as the self-built dataset. The training, validation, and test sets of the dataset are 70%, 20%, and 10%, respectively. Finally, the number of image pairs in the three datasets is 7120, 2048, and 1024, respectively.

The Wuhan University dataset [9] is a widely used public building-change-detection dataset proposed by Wuhan University. The dataset is collected from two aerial remote-sensing RGB images of New Zealand in 2012 and 2016, containing 12,796 buildings and 16,077 buildings, respectively. The size is $32,507 \times 15,354$ pixels at a resolution of 0.2 m. We crop the two images into pairs which are 256×256 pixels in size without overlapping areas. Then, we randomly divide the cropped images into three parts: 6096, 762, and 762 in the ratio of 8:1:1. These are used as the training dataset, test dataset, and validation dataset, respectively.

On the two publicly available datasets, we performed data preprocessing including data cutting and dividing the dataset, the images in the data set were evenly cut into image blocks of 256×256 pixels.

4.2. Experiment Setting and Evaluation Metrics

The stochastic gradient descent (SGD) optimizer was used to optimize the experimental process. Learning rate (LR) and batch size (BS) were obtained experimentally. Cross entropy loss function is adopted for the loss function. The learning rate strategy was the polynomial decay strategy, using the formula:

$$lr = lr_0 \times \left(1 - \frac{i}{\max_i}\right)^p \quad (3)$$

where lr is the learning rate, lr_0 is the initial learning rate and was set to 0.001, i refers to the current iteration number, max_i refers to the maximum number of iterations, and p is the learning rate strategy index, which is set to 0.9 in the experiment.

In order to make an effective evaluation of the experimental results, this paper used the overall accuracy (OA), F1-score ($F1$), precision (Pre), recall (Rec), and intersection over union (IoU) as evaluation indices. The formulas are as follows:

$$OA = \frac{TP + TN}{P + N} \quad (4)$$

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec} \quad (5)$$

$$Pre = \frac{TP}{TP + FP} \quad (6)$$

$$Rec = \frac{TP}{TP + FN} \quad (7)$$

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

where P , N , TP , TN , FP , and FN represent the positive, negative, true positive, true negative, false positive, and false negative pixels in the prediction result, respectively.

4.3. Comparison of Most Recent Networks

To verify the effectiveness and superiority of HSAA-CD, we selected several previous change-detection methods as comparative methods on three datasets (LEVIR, WHU-CD, and self-built RSI datasets). The methods used for comparison include the SNUNet [24], DTCDCN [47], and BIT methods [39].

SNUNet [24] uses the Siamese UNet++ network as a feature extraction tool and uses the integrated channel attention module that has been comprehensively investigated for land cover CD. DTCDCN [47] uses two different encoders and a dual attention module to extract more context features in the decoder part to obtain more detailed difference features. The BIT [39] method is a transformer-based approach that uses a transformer to model long-range context in bitemporal images, enhancing the discrimination of paired features are state-of-the-art methods.

We compare these three change-detection methods mentioned above with our proposed HSAA-CD method on three datasets and analyze the results qualitatively and quantitatively. To validate the effectiveness of our enhanced BIT model, we set the original BIT model as a baseline for comparison.

4.4. Experiments on Self-Built Dataset

Table 5 shows the quantitative results of different methods on the self-built dataset. As shown in Table 5, our method is superior in most evaluation indicators on the self-built dataset compared to other methods. Specifically, our proposed method has achieved the best performances in precision, F1, IoU, and OA, reaching 77.96%, 68.50%, 59.29%, and 93.56%, respectively. Compared with the BIT model, our method achieves performance improvements of 1.99% and 1.62% in the F1 and IoU metrics, respectively.

Figure 6 shows the change-detection results of different methods in the self-built dataset. As shown in the first two rows of Figure 6, although the BIT model and our model detect the wrong change area, the prediction of building on agricultural land is accurate, while the other two methods have errors in detecting the change area or fail to detect the change area. It can be seen that in the third row of the large agricultural land area change, our method monitors the complete area change, while the other methods fail or partially detect the change area. By use of the channel and spatial attention mechanism added in ResNet and the feature aggregation module, our method is able to better perceive the rich low-level and high-level features, and the algorithm is able to better determine the extent

of the agricultural land and has better performance in detecting the changes of both small and large agricultural land.

Table 5. Experimental results of different methods on the AGRI-CD dataset.

Method	Pre (%)	Rec (%)	F1 (%)	IoU (%)	OA (%)
SNUNet	64.88	66.30	60.09	57.96	92.76
DTCDCN	72.82	64.33	65.32	58.16	93.05
BIT	73.64	63.11	66.51	57.67	93.23
BaseLine	74.24	63.51	67.41	58.27	93.31
+CBAM	77.33	63.73	68.31	59.14	93.42
+CBAM + AM	77.96	63.76	68.50	59.29	93.56

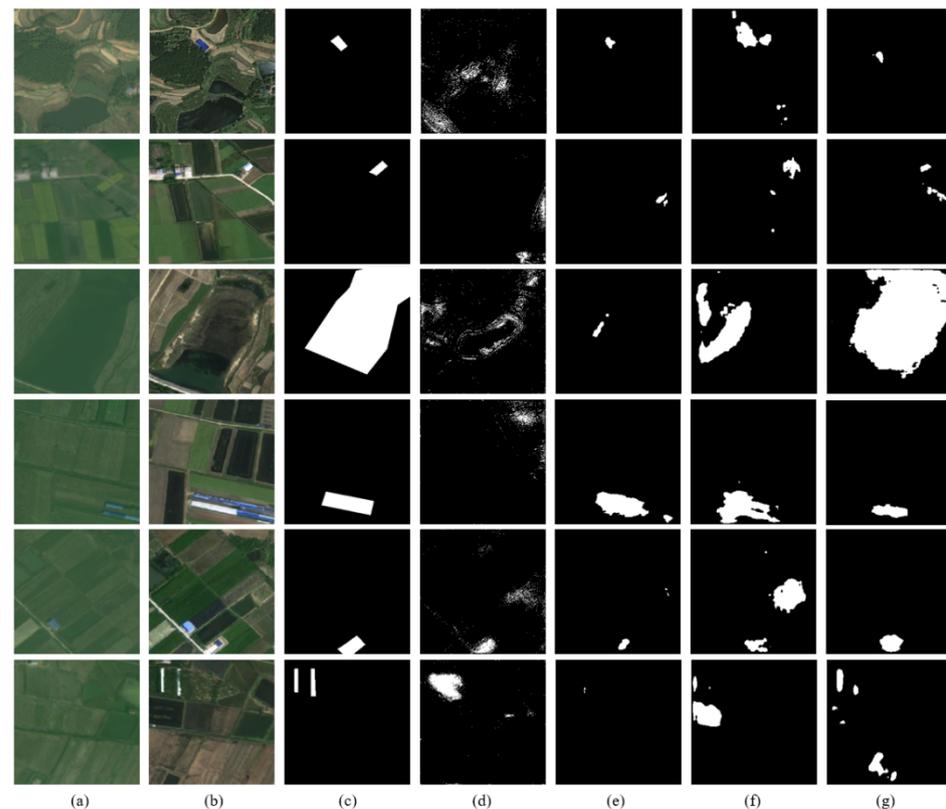


Figure 6. Change maps by different methods on the self-built datasets and qualitative comparison of the results. (a,b) The input bitemporal remote-sensing images. (c) Ground truth. (d) SNUNet. (e) DTCDCN. (f) BIT. (g) Ours. In the change map, white pixels indicate actual changes, and black pixels indicate no changes.

In this test, most of the errors of change detection are concentrated in seasonal changes in rice fields to lakes and ponds, which have strong confusion between rice fields and ponds in the spectrum. The missed identification focuses on the conversion of arable land to landfills, mainly due to the similar spectral characteristics of arable land and landfills, resulting in missed identification of arable land that has actually become landfills. The number of samples has a significant impact on the accuracy of the model. For example, the number of samples in the study area for farmland to garden forest and grass and farmland to landscape park is relatively small, resulting in lower recognition accuracy for farmland to garden forest and grass and farmland to landscape park. Through the application results in five regions, it can be seen that the sample classification proposed in this article is more targeted towards the issue of non-agricultural conversion of cultivated land compared to other studies.

4.5. Experiments on LEVIR-CD Dataset and WHU-CD Dataset

Table 6 shows the quantitative results of different methods on the LEVIR-CD dataset. As shown in Table 6, our method is superior in most evaluation indicators on the LEVIR-CD dataset compared to other methods. Specifically, our proposed method has achieved the best performances in precision, F1, IoU, and OA, reaching 89.14%, 88.56%, 79.84%, and 98.83%, respectively. Compared with the BIT model, our method achieves performance improvements of 0.93% and 0.33% in the F1 and IoU metrics, respectively.

Table 6. Experimental results of different methods on the LEVIR-CD dataset.

Method	Pre (%)	Rec (%)	F1 (%)	IoU (%)	OA (%)
SNUNet	89.14	87.40	87.72	78.37	98.75
DTCDCSN	88.16	86.50	87.32	77.35	98.02
BIT	88.67	88.66	87.63	79.51	98.61
BaseLine	88.72	88.76	87.86	79.62	98.69
+CBAM	88.89	88.85	88.55	79.83	98.79
+CBAM + AM	89.14	88.83	88.56	79.84	98.83

Figure 7 shows the change-detection results of different methods on the LEVIR-CD dataset. As shown in the first two rows of Figure 7, our model is more accurate in predicting edges compared to the BIT method change-detection results. As shown in the sixth row of the figure, our method predicts complex building changes more completely and accurately than other methods. Our method can learn the building characteristics better than other methods, and the detection results contain correct edge information and complete change areas.

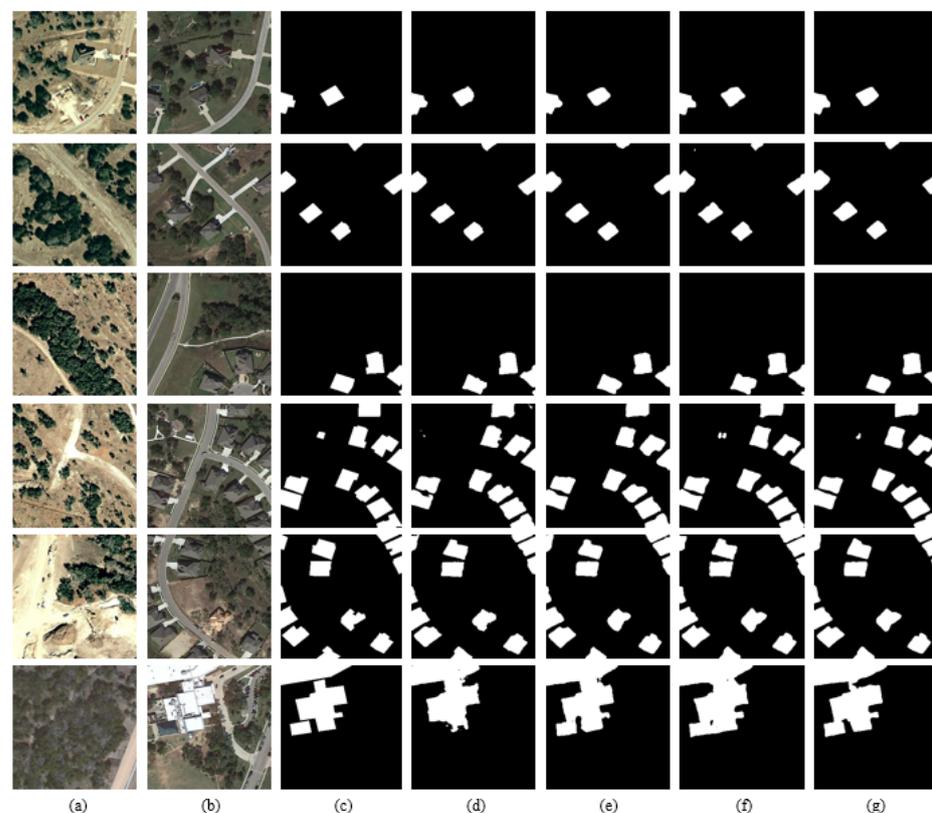


Figure 7. Change maps by different methods on the LEVIR datasets and qualitative comparison of the results. (a,b) The input bitemporal remote-sensing images. (c) Ground truth. (d) SNUNet. (e) DTCDCSN. (f) BIT. (g) Ours. In the change map, white pixels indicate actual changes, and black pixels indicate no changes.

Table 7 shows the quantitative results of different methods on the WHU-CD dataset. The WHU-CD dataset is also a building-change-detection dataset that contains only the change areas of buildings. We selected large building change areas for change detection. Our method achieves optimal results in all metrics of precision, recall, F1, OA, and IoU, reaching 85.55%, 83.54%, 84.29%, 98.39%, and 73.41%, respectively. Compared with other methods, our method improves F1 and IoU by at least 0.52% and 0.51%, respectively.

Table 7. Experimental results of different methods on the WHU-CD dataset.

Method	Pre (%)	Rec (%)	F1 (%)	IoU (%)	OA (%)
SNUNet	78.37	82.20	73.34	71.09	97.62
DTCDCSCN	80.74	81.20	78.32	70.77	97.25
BIT	84.98	82.64	83.77	72.90	97.21
BaseLine	85.08	83.64	84.07	73.09	98.11
+CBAM	85.45	83.50	84.16	73.25	98.32
+CBAM + AM	85.55	83.54	84.29	73.41	98.39

Figure 7 shows the change-detection results of different methods in the WHU-CD dataset. As shown in Figure 8, our method obtained the best correct and complete prediction results among all the methods. In detail, as you can see in the second and fourth rows of the figure, the change-detection results generated by the BIT and DTCDCSCN methods are noisy and contain some unnecessary information. Our method generated less noisy and correct areas. In the sixth row, our method generates a hole, incorrectly treating the lower left corner of the after-time image as a house. Our method generates better results in most cases.

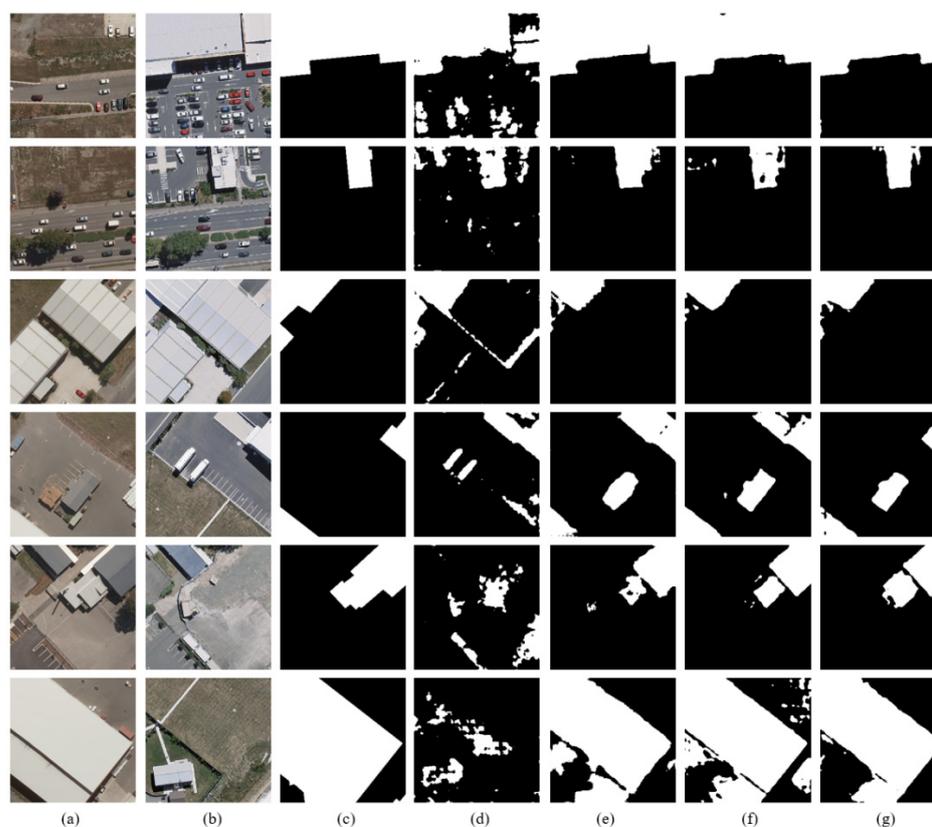


Figure 8. Change maps by different methods on the WHU datasets and qualitative comparison of the results. (a,b) The input bitemporal remote-sensing images. (c) Ground truth. (d) SNUNet. (e) DTCDCSCN. (f) BIT. (g) Ours. In the change map, white pixels indicate actual changes, and black pixels indicate no changes.

Figures 7 and 8 show the change-detection results of the different methods on the LEVIR-CD and WHU datasets. It is observed that our method achieves the best performance on two publicly available building-change-detection datasets of different resolutions and sizes.

We used a graphics workstation with a configuration that includes a Nvidia RTX A6000ada professional graphics card with 48 G memory. The processing time is shown in Table 8.

Table 8. Process time of different methods on three datasets.

Method	Self-Built	LEVIR-CD	WHU-CD
SNUNet	1 h 02 min	13 h 42 min	11 h 10 min
DTCDCN	2 h 47 min	15 h 21 min	9 h 39 min
BIT	1 h 43 min	17 h 15 min	11 h 58 min
BaseLine			
+CBAM	1 h 38 min	15 h 56 min	11 h 28 min
+CBAM + AM	1 h 41 min	16 h 32 min	11 h 42 min

5. Discussion

Since the samples of the self-constructed image data are based on geographical and national monitoring results, this method can significantly improve the product efficiency in the practical application and the quality. The monitoring results can be used as the ground truth during the dataset generation, which can improve the accuracy of the cultivated land non-agricultural monitoring model. The samples collected based on geographical and national monitoring results have high spatiotemporal consistency in their change pattern vectors, remote-sensing image data, and land-use type attribute values. However, the manual collection of patterns or semi-automatic change recognition collection methods are limited to the differences in spectral characteristics of ground objects, which can easily lead to difficulty in distinguishing different objects that have the same spectrum and the same objects that have different spectrums. The cultivated land attribute of the geographical and national monitoring results includes fallow land, and this article classified it as cultivated land in the previous image during sample collection. Due to the temporal characteristics of geographical and national monitoring results, which are concentrated from March to July each year, this article does not consider the situation of winter fallow fields being transferred out. In subsequent research, the temporal range of data sources can be expanded and this type of sample can be collected for monitoring model training.

In this paper, we collected the data resources in Hubei province. The cultivated land or the farmland situation is in relevant provinces and regions in the middle and lower reaches of the Yangtze River. It was also popular for the cultivated land situation in the areas of the same latitude and altitude. In those areas, rice fields are the main cultivated crop. It has very obvious differences from the corn fields in the north. There are also significant differences compared to other rice fields in China. To improve the generality of the method in this article, it is necessary to add more samples of cultivated land in different regions and cases of changes.

On the other hand, the current definition of cultivated land types in this article uses the geographical and national monitoring classification standards, which have some differences from the land cover and utilization classification standards and the classification standards of land resource surveys. The same patch has different names and semantic differences in different classification standards. For the tasks of non-agriculture change detection, mapping lists for the different classes between the different classification standards can improve the quantity of the dataset by integrating the existing land-cover/land-use image datasets.

The unbalance in the number of different classes may reduce the accuracy of specific applications. The path for changes between some classes remains only a theoretical possibility. For those situations, we did not specifically enhance the sample counts and the transformation data for each type of sample.

Compared to other networks, the method proposed in this article focuses more on improving the practical accuracy of specialized application projects. By adding a semantic aggregation module, samples of the same type can be semantically merged without paying attention to the conversion of internal semantic type samples. This is mainly to meet the practical needs of engineering. Existing methods have the ability to detect changes in each subcategory of non-agricultural changes in cultivated land, but their overall accuracy in detecting changes is relatively low. The existing neural network models are susceptible to external factors during change detection, leading to missed detection results. Methods such as the BIT [39] method that use transformers to model long-range context in bitemporal images enhance the ability to discriminate pairs of features. However, transformer-based methods are prone to ignore the low-level semantics during change detection, resulting in unclear edges and failure to correctly identify change regions in the detection results. In practical applications, this often increases the workload and cannot be directly used. The semantic aggregation module and attention module in this article actually sacrifice the detection ability of internal type change transformation to improve the overall accuracy of change detection. To obtain detailed information and accuracy of the transformation of each subclass in the future, it may be necessary to divide it into two parts. By adding subclass mapping relationship diagrams or boundaries, while determining the overall detection accuracy, further refining the content of internal changes and the accuracy of each part.

6. Conclusions

In this paper, we propose a hierarchical semantic aggregation mechanism and an attention module bitemporal image transformer method named HSAA-CD for non-agricultural change detection in cultivated land. We first used multi-resource surveying data to produce a well-tagged and high-resolution cultivated land and non-agricultural image dataset. Then, we proposed a hierarchical semantic aggregation mechanism and an attention module bitemporal image transformer method named HSAA-CD for non-agricultural changes detecting in cultivated land. The HSAA-CD added a hierarchical semantic aggregation mechanism for clustering the input data for ResNet as the backbone network and an attention module to improve the feature edge.

We evaluated the proposed HSAA-CD on three datasets, the self-built image dataset, the LEVIR-CD dataset and the WHU dataset. Firstly, using the self-built image dataset, our method achieved 93.56% OA. Our method improved the OA and average *F1*-score by 0.25% and 1.09%, respectively, compared to the BIT method and was also higher than the SNUNet and DTCDSCN methods. Secondly, on the LEVIR-CD and WHU-CD datasets, the OAs of HSAA-CD were 98.83% and 98.39, and the average *F1*-scores were 88.56% and 84.29, which illustrated the effectiveness and feasibility of HSAA-CD in improving change-detection performance.

Nevertheless, the performance of HSAA-CD is still restricted due to the complex structure and non-uniform distribution of non-agricultural types. As a continuation of this work, some widely used methods, such as semantic information edge detection and the dual attention mechanism, will be considered in the future as methods to improve the performance of HSAA-CD for non-agricultural change of cultivated land. The proposed HSAA-CD method proved suitably accurate for high-resolution remote-sensing change detection of non-agricultural in cultivated land.

Author Contributions: Conceptualization, F.L. and G.Z.; Methodology, F.L. and F.Z.; Software, F.L., F.Z. and J.X.; Formal analysis, F.L.; Resources, P.Z.; Data curation, F.L. and J.X.; Writing—original draft, F.Z.; Writing—review & editing, F.L.; Visualization, J.X.; Supervision, G.Z.; Project administration, F.L.; Funding acquisition, G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Key Research and Development Program of China (NO. 2023YFB3905503).

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank all the anonymous reviewers for their helpful comments and suggestions to improve the manuscript.

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

References

1. Zhou, Y.; Zhong, Z.; Cheng, G. Cultivated Land Loss and Construction Land Expansion in China: Evidence from National Land Surveys in 1996, 2009 and 2019. *Land Use Policy* **2023**, *125*, 106496. [[CrossRef](#)]
2. Liu, S.; Xiao, W.; Ye, Y.; He, T.; Luo, H. Rural Residential Land Expansion and Its Impacts on Cultivated Land in China between 1990 and 2020. *Land Use Policy* **2023**, *132*, 106816. [[CrossRef](#)]
3. Li, H.; Song, W. Spatial Transformation of Changes in Global Cultivated Land. *Sci. Total Environ.* **2023**, *859*, 160194. [[CrossRef](#)] [[PubMed](#)]
4. Wu, Y.; Wang, Y.; Li, Y.; Xu, Q. Optical Satellite Image Change Detection Via Transformer-Based Siamese Network. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Kuala Lumpur, Malaysia, 17–22 July 2022; Institute of Electrical and Electronics Engineers Inc.: Danvers, MA, USA, 2022; Volume 2022-July, pp. 1436–1439.
5. Abbott, R.; Robertson, N.M.; Martinez Del Rincon, J.; Connor, B. Unsupervised Object Detection via LWIR/RGB Translation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; Volume 2020-June, pp. 407–415.
6. Yang, L.; Chen, Y.; Song, S.; Li, F.; Huang, G. Deep Siamese Networks Based Change Detection with Remote Sensing Images. *Remote Sens.* **2021**, *13*, 3394. [[CrossRef](#)]
7. Benedek, C.; Szirányi, T. Change Detection in Optical Aerial Images by a Multilayer Conditional Mixed Markov Model. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3416–3430. [[CrossRef](#)]
8. Bourdis, N.; Marraud, D.; Sahbi, H. Constrained Optical Flow for Aerial Image Change Detection. In Proceedings of the IEEE International Geoscience & Remote Sensing Symposium, Vancouver, BC, Canada, 24 July 2011.
9. 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](#)]
10. Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A Deeply Supervised Attention Metric-Based Network and an Open Aerial Image Dataset for Remote Sensing Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5604816. [[CrossRef](#)]
11. 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](#)]
12. 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](#)]
13. 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](#)]
14. Daudt, R.C.; Saux, B.; Boulch, A.; Gousseau, Y. Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018.
15. Zhang, J.; Shao, Z.; Ding, Q.; Huang, X.; Wang, Y.; Zhou, X.; Li, D. AERNet: An Attention-Guided Edge Refinement Network and a Dataset for Remote Sensing Building Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5617116. [[CrossRef](#)]
16. Shen, L.; Lu, Y.; Chen, H.; Wei, H.; Xie, D.; Yue, J.; Chen, R.; Lv, S.; Jiang, B. S2looking: A Satellite Side-Looking Dataset for Building Change Detection. *Remote Sens.* **2021**, *13*, 5094. [[CrossRef](#)]
17. Li, S.; Wang, Y.; Cai, H.; Lin, Y.; Wang, M.; Teng, F. MF-SRCDNet: Multi-Feature Fusion Super-Resolution Building Change Detection Framework for Multi-Sensor High-Resolution Remote Sensing Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *119*, 103303. [[CrossRef](#)]
18. 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. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences—ISPRS Archives, Riva del Garda, Italy, 30 May 2018; International Society for Photogrammetry and Remote Sensing: Riva del Garda, Italy, 2018; Volume 42, pp. 565–571.
19. Zhang, Z.; Vosselman, G.; Gerke, M.; Tuia, D.; Yang, M.Y. Change Detection between Multimodal Remote Sensing Data Using Siamese CNN. *arXiv* **2018**, arXiv:1807.09562.
20. Song, L.; Xia, M.; Weng, L.; Lin, H.; Qian, M.; Chen, B. Axial Cross Attention Meets CNN: Bibranch Fusion Network for Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 32–43. [[CrossRef](#)]
21. Lei, G.; Li, A.; Bian, J.; Naboureh, A.; Zhang, Z.; Nan, X. A Simple and Automatic Method for Detecting Large-Scale Land Cover Changes without Training Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 7276–7292. [[CrossRef](#)]
22. Shi, S.; Zhong, Y.; Zhao, J.; Lv, P.; Liu, Y.; Zhang, L. Land-Use/Land-Cover Change Detection Based on Class-Prior Object-Oriented Conditional Random Field Framework for High Spatial Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5600116. [[CrossRef](#)]
23. Lv, Z.; Huang, H.; Gao, L.; Benediktsson, J.A.; Zhao, M.; Shi, C. Simple Multiscale UNet for Change Detection with Heterogeneous Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2504905. [[CrossRef](#)]
24. 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](#)]

25. Huang, L.; Tian, Q.; Tang, B.H.; Le, W.; Wang, M.; Ma, X. Siam-EMNet: A Siamese EfficientNet–MANet Network for Building Change Detection in Very High Resolution Images. *Remote Sens.* **2023**, *15*, 3972. [[CrossRef](#)]
26. Hosseinpour, H.; Samadzadegan, F.; Javan, F.D. CMGFNet: A Deep Cross-Modal Gated Fusion Network for Building Extraction from Very High-Resolution Remote Sensing Images. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 96–115. [[CrossRef](#)]
27. Varghese, A.; Gubbi, J.; Ramaswamy, A.; Balamuralidhar, P. ChangeNet: A Deep Learning Architecture for Visual Change Detection. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Munich, Germany, 8–14 September 2018; Springer: Munich, Germany, 2019; Volume 11130 LNCS, pp. 129–145.
28. 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](#)]
29. Chen, C.-P.; Hsieh, J.-W.; Chen, P.-Y.; Hsieh, Y.-K.; Wang, B.-S. SARAS-Net: Scale and Relation Aware Siamese Network for Change Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022.
30. Han, C.; Wu, C.; Guo, H.; Hu, M.; Li, J.; Chen, H. Change Guiding Network: Incorporating Change Prior to Guide Change Detection in Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 8395–8407. [[CrossRef](#)]
31. Zhu, S.; Song, Y.; Zhang, Y.; Zhang, Y. ECFNet: A Siamese Network with Fewer FPs and Fewer FNs for Change Detection of Remote-Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6001005. [[CrossRef](#)]
32. Li, X.; Yan, L.; Zhang, Y.; Mo, N. SDMNet: A Deep-Supervised Dual Discriminative Metric Network for Change Detection in High-Resolution Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5513905. [[CrossRef](#)]
33. Bandara, W.G.C.; Nair, N.G.; Patel, V.M. DDPM-CD: Remote Sensing Change Detection Using Denoising Diffusion Probabilistic Models. *arXiv* **2022**, arXiv:2206.11892.
34. Codegoni, A.; Lombardi, G.; Ferrari, A. TINYCD: A (Not So) Deep Learning Model for Change Detection. *Neural Comput. Appl.* **2022**, *35*, 8471–8486. [[CrossRef](#)]
35. Zhao, C.; Tang, Y.; Feng, S.; Fan, Y.; Li, W.; Tao, R.; Zhang, L. High-Resolution Remote Sensing Bitemporal Image Change Detection Based on Feature Interaction and Multitask Learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5511514. [[CrossRef](#)]
36. Lv, Z.; Zhong, P.; Wang, W.; You, Z.; Falco, N. Multiscale Attention Network Guided with Change Gradient Image for Land Cover Change Detection Using Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 2501805. [[CrossRef](#)]
37. Zhang, H.; Ma, G.; Zhang, Y.; Wang, B.; Li, H.; Fan, L. MCHA-Net: A Multi-End Composite Higher-Order Attention Network Guided with Hierarchical Supervised Signal for High-Resolution Remote Sensing Image Change Detection. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 40–68. [[CrossRef](#)]
38. Chen, H.; Song, J.; Wu, C.; Du, B.; Yokoya, N. Exchange Means Change: An Unsupervised Single-Temporal Change Detection Framework Based on Intra- and Inter-Image Patch Exchange. *ISPRS J. Photogramm. Remote Sens.* **2023**, *206*, 87–105. [[CrossRef](#)]
39. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [[CrossRef](#)]
40. 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), Kuala Lumpur, Malaysia, 17–22 July 2022; Institute of Electrical and Electronics Engineers Inc.: Danvers, MA, USA, 2022; Volume 2022-July, pp. 207–210.
41. Cheng, H.; Wu, H.; Zheng, J.; Qi, K.; Liu, W. A Hierarchical Self-Attention Augmented Laplacian Pyramid Expanding Network for Change Detection in High-Resolution Remote Sensing Images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *182*, 52–66. [[CrossRef](#)]
42. Shu, Q.; Pan, J.; Zhang, Z.; Wang, M. DPCC-Net: Dual-Perspective Change Contextual Network for Change Detection in High-Resolution Remote Sensing Images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102940. [[CrossRef](#)]
43. Liu, W.; Lin, Y.; Liu, W.; Yu, Y.; Li, J. An Attention-Based Multiscale Transformer Network for Remote Sensing Image Change Detection. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 599–609. [[CrossRef](#)]
44. Guo, H.; Du, B.; Zhang, L.; Su, X. A Coarse-to-Fine Boundary Refinement Network for Building Footprint Extraction from Remote Sensing Imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 240–252. [[CrossRef](#)]
45. Yin, H.; Ma, C.; Weng, L.; Xia, M.; Lin, H. Bitemporal Remote Sensing Image Change Detection Network Based on Siamese-Attention Feedback Architecture. *Remote Sens.* **2023**, *15*, 4186. [[CrossRef](#)]
46. Ma, C.; Yin, H.; Weng, L.; Xia, M.; Lin, H. DAFNet: A Novel Change-Detection Model for High-Resolution Remote-Sensing Imagery Based on Feature Difference and Attention Mechanism. *Remote Sens.* **2023**, *15*, 3896. [[CrossRef](#)]
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016. [[CrossRef](#)]
48. Wu, B.; Xu, C.; Dai, X.; Wan, A.; Zhang, P.; Tomizuka, M.; Keutzer, K.; Vajda, P. Visual transformers: Token-based image representation and processing for computer vision. *arXiv* **2020**, arXiv:2006.03677.
49. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module; Springer: Cham, Switzerland, 2018. [[CrossRef](#)]

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