International Journal of
Geo-Information

Article

Enhancing Adversarial Learning-Based Change Detection in Imbalanced Datasets Using Artificial Image Generation and Attention Mechanism

Amel Oubara [1], Falin Wu [1,*], Reza Maleki [1], Boyi Ma [1], Abdenour Amamra [2] and Gongliu Yang [3]

[1] SNARS Laboratory, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China; oubara@buaa.edu.cn (A.O.); maleki@buaa.edu.cn (R.M.); maboyi@buaa.edu.cn (B.M.)
[2] Ecole Militaire Polytechnique, Bordj El-Bahri BP 17, Algiers 16000, Algeria; amamra.abdenour@gmail.com
[3] School of Mechanical Engineering, Zhejiang University, Hangzhou 310030, China; yanggl007@zju.edu.cn
* Correspondence: falin.wu@buaa.edu.cn; Tel.: +86-10-82313929

Abstract: Deep Learning (DL) has become a popular method for Remote Sensing (RS) Change Detection (CD) due to its superior performance compared to traditional methods. However, generating extensive labeled datasets for DL models is time-consuming and labor-intensive. Additionally, the imbalance between changed and unchanged areas in object CD datasets, such as buildings, poses a critical issue affecting DL model efficacy. To address this issue, this paper proposes a change detection enhancement method using artificial image generation and attention mechanism. Firstly, the content of the imbalanced CD dataset is enhanced using a data augmentation strategy that synthesizes effective building CD samples using artificial RS image generation and building label creation. The created building labels, which serve as new change maps, are fed into a generator model based on a conditional Generative Adversarial Network (c-GAN) to generate high-resolution RS images featuring building changes. The generated images with their corresponding change maps are then added to the CD dataset to create the balance between changed and unchanged samples. Secondly, a channel attention mechanism is added to the proposed Adversarial Change Detection Network (Adv-CDNet) to boost its performance when training on the imbalanced dataset. The study evaluates the Adv-CDNet using WHU-CD and LEVIR-CD datasets, with WHU-CD exhibiting a higher degree of sample imbalance compared to LEVIR-CD. Training the Adv-CDNet on the augmented dataset results in a significant 16.5% F1-Score improvement for the highly imbalanced WHU-CD. Moreover, comparative analysis showcases the superior performance of the Adv-CDNet when complemented with the attention module, achieving a 6.85% F1-Score enhancement.

Keywords: building change detection; data imbalance; remote sensing image generation; GAN; adversarial learning; attention module

Citation: Oubara, A.; Wu, F.; Maleki, R.; Ma, B.; Amamra, A.; Yang, G. Enhancing Adversarial Learning-Based Change Detection in Imbalanced Datasets Using Artificial Image Generation and Attention Mechanism. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 125. https://doi.org/10.3390/ijgi13040125

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 12 January 2024
Revised: 3 April 2024
Accepted: 6 April 2024
Published: 9 April 2024

1. Introduction

Change Detection (CD) through the analysis of Remote Sensing (RS) images stands as an indispensable tool across a multitude of disciplines, encompassing agriculture, urban planning, and environmental surveillance. The fundamental principle involves the analysis of two or more images captured over different time instances, all within the same geographical area, aimed at discerning temporal changes [1]. The continuous evolution of remote sensing technology has significantly eased the task of detecting changes in even small-scale objects, such as buildings, leveraging the capabilities of Very High-Resolution (VHR) Images. In the sector of urban management, specifically Land Use Land Cover (LULC), identification of illegal construction, and disaster evaluation, the application of building change detection proves to be significantly useful. Furthermore, the insights derived from change detection analysis offer valuable solutions to policymakers for the effective planning and monitoring of sustainable urban development, ensuring the preservation of ecological balance and the development of cities [2].

Considering the importance of building CD, several traditional and Deep-Learning (DL)-based methods have been proposed to accurately accomplish this task. Traditional techniques for CD rely on manual interpretation and image differencing, practices that, though established, are susceptible to consuming significant time and are prone to errors. However, the Remote Sensing Change Detection (RSCD) task has been profoundly reshaped by the strides taken in deep learning. These advancements have launched a new era of methodologies characterized by heightened efficiency and precision [3].

Depending on how the deep features are extracted or the hidden patterns are learned from the bi-temporal data, DL techniques for detecting changes can be categorized into two main approaches: single-stream and double-stream [4]. The single-stream approach typically involves combining the bi-temporal input images and subsequently conducting a classification task to generate a binary or multiclass Change Map (CM). However, this configuration poses two significant research challenges: determining the data fusion strategy and optimizing the DL classifier. In contrast to the single-stream model, which operates with a single network, the double-stream architecture comprises two subnetworks with identical structures. These subnetworks are concurrently treated and trained to discern the deep features inherent in the two input images. The outcomes are then concatenated to formulate the CM. This configuration, founded upon the Siamese convolutional network, finds widespread application [5] owing to its capability to simultaneously train the two subnetworks and learn the deep features of bitemporal input images. However, the current dual-stream networks exhibit certain limitations, including elevated complexity and the need for heightened precision in generating the final CM.

To address this challenge, we introduce a single-stream architecture that leverages adversarial learning. Our approach involves concurrently training two sub-networks within an adversarial learning, one tasked with generating a CM and the other designed to evaluate the quality of the generated CM. The model we propose, named Adversarial Change Detection Network (Adv-CDNet), adopts the adversarial learning principles of Generative Adversarial Networks (GANs) [6]. The foundational architecture of our model draws inspiration from the Pix2Pix model, renowned for its ability to translate an input image from a source domain to a desired target domain [7]. In the same idea of processing the CD task as an image-to-image translation problem [8], our model operates by employing the resulting six-channel image, obtained by concatenating the two bi-temporal images, as the source input, while the CM serves as the intended target image. A schematic representation of our proposed model is depicted in Figure 1.

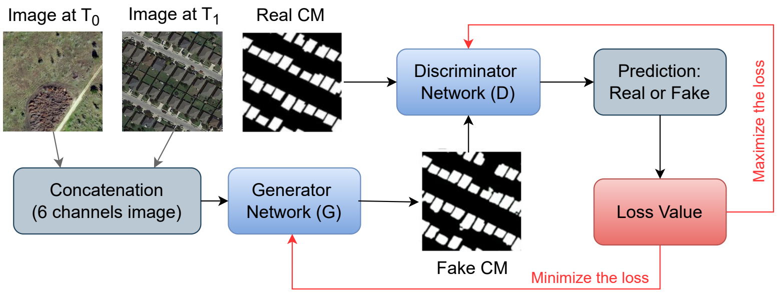


Figure 1. Adversarial learning of the change detection model.

The underlying framework of the pix2pix architecture relies on a supervised deep-learning methodology that depends on the availability of extensive labeled datasets. However, this approach faces challenges when applied to the RS building CD task, which suffers from the lack of large established datasets. The process of annotating large-scale CD datasets is marked by its time-intensive and labor-demanding nature. Additionally, the rarity and sparsity of the buildings changes (considered as the positive class) render the acquisition of compelling bitemporal images a formidable task. Compounding the issue, currently available building CD datasets, such as those highlighted in [9,10], often encompass only restricted geographic regions and are constrained by limited variations in image conditions.

To address the challenge posed by insufficient CD datasets, the adoption of data augmentation techniques emerges as a viable solution. Traditional approaches, predominantly rooted in image processing methodologies encompassing geometric and color transformations, image blending, and related techniques, have been explored [11]. However, these methods primarily involve geometric transformations or simply change the pixel values within RGB channels. Consequently, they fail in enhancing the semantic information fidelity of RS images, particularly when deployed in tasks demanding nuanced interpretation like CD [12]. To overcome these limitations, some works have leveraged imaging simulation systems to generate synthetic RS samples, which are subsequently combined with original data [13]. These innovative methodologies effectively address concerns such as data diversity, blurriness, and distortions. However, it is important to note that the generated images often exhibit compromised quality [3]. Despite these efforts, the challenge of generating high-quality synthetic images for bolstering CD datasets remains.

Recently, the performance of DL in image generation has manifested across computer vision tasks, demonstrating the capacity to produce high-quality and diverse samples that augment the original dataset. This technique has found substantial utility in generating RS samples as well. In the field of RS, DL-based approaches for image generation are based on various techniques, including Variational Auto Encoding (VAE) and adversarial learning such as the application of GANs [11]. For instance, Lv et al. [12] introduced a modified GAN, termed Deeply supervised GAN (D-sGAN), to synthesize new RS training samples for soil-moving detection. Expanding on this, Singh and Bruzzone [14] enhanced the generative adversarial network with class-based spectral indices, facilitating the generation of multispectral RS images. Addressing the specific task of aircraft detection within RS images, Liu et al. [15] devised a multiscale attention Cycle GAN to create novel samples. Xu et al. [16], on the other hand, proposed a data augmentation strategy combining a modified pix2pix model with the copy–paste operator for Solid Waste Detection. Notably, these endeavors primarily center around generating RS images intended for RS classification and object detection tasks.

Despite these advancements, generating new samples for CD tasks remains a formidable challenge within the DL method. Seo et al. [17] tackled this by synthesizing changes through diverse mechanisms like random building cropping, inpainting for building suppression, and copy–paste instance labeling. Another work proposed by Chen et al. [18] employed a GAN-based approach to create new building CD samples. Their methodology involved training a GAN model on a building dataset, followed by transferring generated instances of varying styles onto the synthesized images. The authors additionally introduced context-aware blending techniques for realistic building-background composites, concluding with context-aware color transfer for the final output. Similarly, Li et al. [19] proposed a method called Image-level Sample Pair Generation (ISPG) based on Label Translation GAN (LT-GAN) to address the challenges of limited data volume and severe class-imbalance issues in building change detection datasets.

Within the scope of our research, as illustrated in Figure 2, we introduce a framework based on the GAN model, designed to generate new images at time T_1 (post-change instance), that contains basically buildings objects by taking building labels as input. This is achieved through the creation of a novel building label, extrapolated from an image devoid of buildings taken at time T_0 (pre-change instance). Subsequently, the generated image at T_1 and its respective image at T_0, both accompanied by their corresponding created building masks, are integrated with the original dataset to create the balance. This concerted effort serves to amplify the quantity and variety of building CD samples, effectively remedying the data imbalance highlighted earlier.

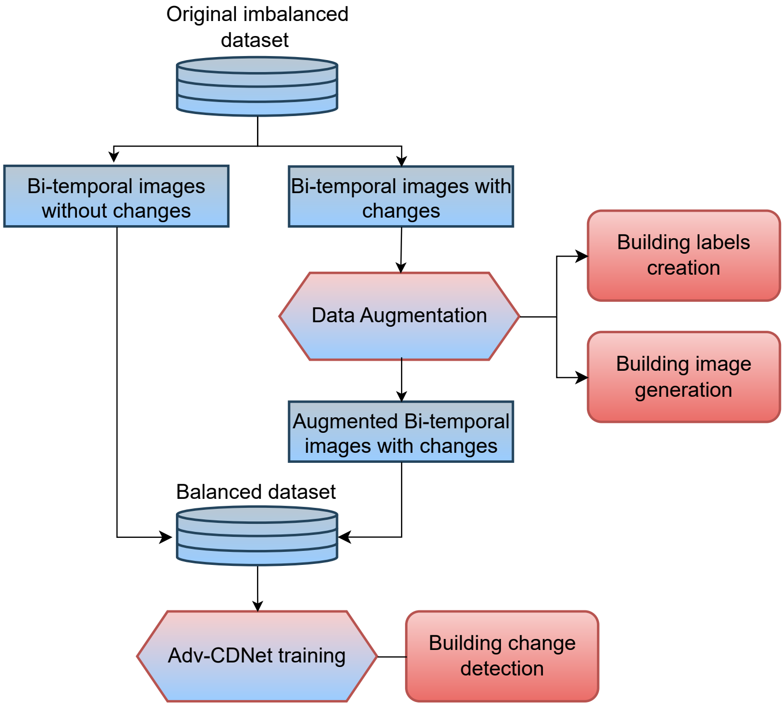

Original imbalanced dataset
Bi-temporal images without changes
Bi-temporal images with changes
Data Augmentation
Building labels creation
Building image generation
Augmented Bi-temporal images with changes
Balanced dataset
Adv-CDNet training
Building change detection

Figure 2. Comprehensive workflow: building change augmentation and detection.

Furthermore, beyond addressing data imbalance issues using data augmentation techniques, our study establishes the efficacy of introducing an attention module into a deep-learning model. This technique, accomplished through the integration of an attention mechanism, serves as a powerful strategy for rectifying the imbalance between changed and unchanged pixels. The utility of attention mechanisms lies in their capacity to enhance model detection capabilities by emphasizing specific features. In the context of RSCD, this mechanism ensures a heightened focus on changes within an image. For instance, a recent work by Feng et al. [20] introduced a dual-branch multilevel intertemporal network leveraging self and cross-attention mechanisms to effectively capture change representations, particularly in cases where foregrounds and backgrounds vary. Similarly, Li et al. [21] proposed a progressive feature aggregation approach with supervised attention, embedded within MobileNet architecture. This technique demonstrated high accuracy while maintaining a reduced parameter count and shorter training times for CD tasks. Due to its efficiency, we integrated an attention module into our Adv-CDNet model. This amelioration significantly improves the accuracy of change detection while concurrently rectifying the imbalance between changed and unchanged areas.

This paper contributed significantly in the following ways:

1. Firstly, we propose a data augmentation strategy designed to effectively generate new CD samples featuring diverse changes in numerous buildings. By employing building label creation and artificial image generation, we enhance the existing CD dataset, ultimately mitigating the risk of class imbalance challenges commonly encountered during the training of DL models for remote sensing building change detection.
2. Secondly, we present an innovative adversarial training framework called Adv-CDNet, which utilizes a modified Pix2Pix model and integrates a channel attention mechanism. This model can directly map bi-temporal images to a CM while extracting more discriminative features, in the imbalanced dataset, for the CD task.
3. Thirdly, we assess the performance of our high-resolution image generation framework on datasets with severe class imbalance. Experimentation involves two distinct publicly available Remote Sensing (RS) building Change Detection (CD) datasets. Comprehensive comparisons between our Adv-CDNet and other state-of-the-art methods show its

effectiveness over alternative approaches. Furthermore, the empirical findings from the evaluations of the incorporation of our data augmentation technique demonstrate significant improvement in the performance of our proposed model.

2. Related Works

2.1. Deep-Learning-Based Methods for Remote Sensing Change Detection

DL-based CD models are commonly structured around two key components. The primary and pivotal element is the feature extractor, tasked with harnessing the capabilities of deep neural networks to derive effective feature representations. The second element is the feature discriminator, responsible for classifying areas as changed or unchanged based on the extracted features. DL techniques are broadly categorized into supervised and unsupervised learning paradigms, depending on the presence or absence of labeled training data. In both learning approaches, a multitude of architectural frameworks has been proposed to address the RSCD task. For instance, Deep Belief Networks (DBNs) have emerged as a notable architecture within various unsupervised change detection methodologies [22]. In a similar way, Auto Encoders (AEs), being unsupervised feedforward neural networks, exhibit the capacity to generate sparse representations of input data. AE encompasses two integral components: the encoder, responsible for condensing the input image into a lower-dimensional representation, and the decoder, which reconstructs the encoder's output into an image closely resembling the original input. Capitalizing on their robust feature learning capabilities, diverse AE variants have been harnessed as feature extractors within the context of change detection. These variants encompass stacked AEs, which incorporate multiple layers of autoencoders; stacked denoising AEs; stacked Fisher AEs; sparse AEs; denoising AEs; fuzzy AEs; and contractive AEs [4]. Convolutional Neural Networks (CNNs) have garnered notable success in a multitude of image-processing tasks, primarily attributed to their intrinsic ability to autonomously learn deep features. These networks, both classical and their refined iterations [23], are extensively exploited as adept classifiers or feature extractors within the realm of change detection. Noteworthy examples encompass VGGNet, CaffeNet, SegNet, UNet, InceptionNet, and ResNet [4].

Recently, some developments have witnessed the integration of adversarial learning strategies to tackle the intricacies of remote sensing change detection. A pertinent instance is the conditional adversarial network proposed by Niu et al. [8], tailored for the challenge of change detection within heterogeneous images. Moreover, the implementation of GANs has emerged as a promising avenue for accurate urban change detection [24]. This marks a distinct shift in the landscape of change detection methodologies, wherein the alliance of adversarial learning principles and remote sensing image analysis showcases remarkable potential. In the context of our study, we extend our efforts by introducing a tailored modification to the pix2pix model, a distinctive form of conditional GAN renowned for its prowess in image-to-image translation tasks within the computer vision tasks [7]. We deploy this model specifically to address the building change detection task.

Furthermore, in a concerted effort to boost the efficiency of change detection tasks, various attention modules have been developed and integrated into change detection models. These modules enable neural networks to selectively focus on subsets of inputs or features by effectively selecting the most relevant ones, modeling intricate bi-temporal features, and enhancing feature representation. As a result, the attention mechanism serves to mitigate the influence of irrelevant information while emphasizing pertinent data, thereby improving overall performance [25]. Several papers have proposed different approaches that incorporate attention mechanisms in change detection, including the hierarchical attention network [26], supervised attention [27], and channel self-attention [28]. Moreover, Zhang et al. developed the Dual Cross-Attention-Transformer (DCAT) method, which utilizes the cross-attention mechanism to improve change feature discrimination and merges bi-temporal features [29]. Some works have explored multi-scale attention mechanisms. Zhang et al. proposed a Dual Multi-scale Attention model for change detection, incorporating a double-threshold automatic data equalization rule for data category imbalance [30]. However, multiple pooling operations can result in the loss of structural information. Ren et al. introduced the Dual-Attention-Guided Multi-scale Feature Aggregation Network for change detection, which addresses the problems of multiscale feature fusion and attention allocation strategy in DL methods. It utilizes cross-fusion of different scales and dual attention to guide fusion in space and channel information [31]. In our study, we incorporate channel attention module into the architecture of our adversarial change detection network. This amelioration aligns with the prevailing trend of including attention mechanisms into neural network structures to improve their performance in complex image analysis tasks, such as remote sensing change detection.

2.2. Class Imbalance Challenge in Change Detection

The quantity and quality of image data have a significant influence on the training process of DL models. Invariably, a wealth of high-quality training samples becomes imperative to ensure the precision and robustness essential to the execution of any RS task performed by DL-based methodologies. In the CD task, the acquisition of a substantial number of effective bitemporal images is a great challenge due to the rarity and sparsity of real-world changes. Consequently, most CD datasets, particularly those focused on building changes where the changed objects are small compared to the background, suffer from a pronounced class imbalance between areas with changes and those without changes

To counter this challenge, a prevalent approach involves the manipulation of weighted loss functions to compel the model to accord greater attention to samples exhibiting changes during the training phase. Illustratively, methodologies like weighted constrictive loss, as implemented in [9], or weighted cross-entropy loss, as in [32], are commonly leveraged. Additionally, Liu et al. [33] adopted a weighted focal loss, introducing nonlinear weights for different classes to reshape the original focal loss. Although these techniques, by strategically enhancing the significance attributed to changing samples, address the data imbalance issue, they can be more computationally expensive compared to standard cross-entropy loss, as they involve additional calculations. An alternative strategy to address data imbalance involves the utilization of data augmentation techniques to gather sufficient training samples. As delineated earlier, three distinct categories of data augmentation exist: traditional methods, imaging simulation methods, and image generation through the application of GANs. In this section, our focus is directed toward image generation utilizing GANs.

Proposed initially by Goodfellow et al. in 2014 [6], GAN stands as a prominent deep-learning architecture encompassing two neural networks: a generator and a discriminator. Operating collaboratively, these components strive to produce images closely resembling authentic ones. The predominant challenges inherent to GAN-based image generation pertain to maintaining training stability and achieving heightened image realism. Overcoming these challenges has been the driving force behind numerous advancements. For instance, Radford et al. [34] integrated Convolutional Neural Networks (CNNs) into GAN architectures, while the application of the Wasserstein distance [35] has been instrumental in stabilizing training dynamics. Furthermore, to enhance the quality of spectral images, Singh and Bruzzone [14] introduced the concept of spectral index GANs. Moreover, incorporating a conditional vector into both the generator and discriminator components enhances control over image attributes, including aspects such as quantity, shape, and type. This augmented architecture, aptly named conditional GAN, was originally put forth by Gauthier et al. [36]. This configuration has found extensive application in various works [12,16,18], all exploiting its potential to generate remote sensing images.

In line with this approach, our study takes a novel path, using a customized conditional GAN architecture with specific modifications adapted to generate image pairs for building change detection, all hinged upon created building labels.

3. Materials and Methods

3.1. Dataset Preparation

3.1.1. Change Detection Dataset

For a comprehensive assessment of our CD model, we conducted experiments on two widely acknowledged public datasets specifically designed for building CD using VHR RS images: the WHU Building Change Detection Dataset (WHU-CD) [10] and the LEVIR Building Change Detection Dataset (LEVIR-CD) [9]. These two datasets present highly imbalanced change label scenarios, as elucidated in Figure 3, where the ratio of changed pixels to unchanged pixels is notably skewed, thus giving rise to a pronounced class imbalance. This, in turn, amplifies the challenges associated with employing DL techniques for building change detection tasks.

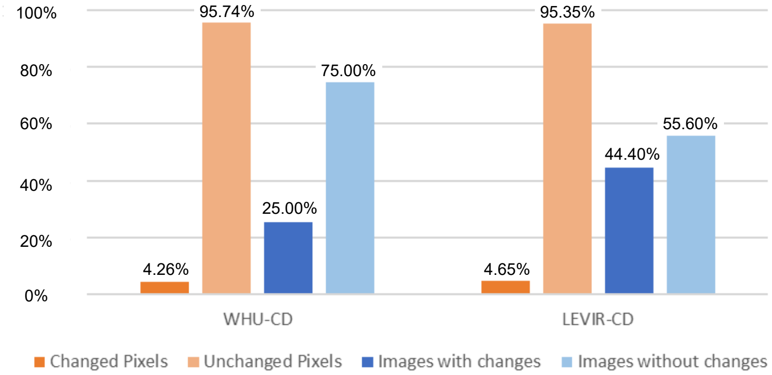


Figure 3. Visualization of the number of changed and unchanged pixels and images in the LEVIR-CD and WHU-CD datasets.

The WHU-CD dataset [10] comprises a single pair of VHR optical RS images (with a spatial resolution of 0.2 m) sized at 32,507 × 15,354 pixels. These bi-temporal RGB images, captured in 2012 (12,796 buildings) and 2016 (16,077 buildings) across a 20.5 km^2 area, were acquired following an earthquake in February 2011 to monitor post-earthquake building reconstruction. At a pixel level, our analysis identified 21,442,501 changed pixels, constituting only 4.26% of the total pixel count. The remaining 481,873,979 pixels were unchanged, accounting for 95.74%. This data distribution results in a highly imbalanced binary classification dataset (as depicted in Figure 3). To facilitate experimentation, we segmented these images into non-overlapping patches of size 256 × 256, yielding a collection of 7620 patch pairs. Notably, only 25.00% of their corresponding change maps depicted changes, intensifying the disparity between change maps containing building changes (changed maps) and those without (unchanged maps). We adopted a random 7:1:2 ratio to divide the dataset into training, validation, and testing subsets.

The LEVIR-CD dataset [9], on the other hand, encompasses 637 Google Earth image patch pairs, each characterized by a very high spatial resolution of 0.5 m and a size of 1024 × 1024 pixels. With a diverse range of building types, including villa residences, tall apartments, small garages, and large warehouses, this dataset showcases substantial land-use changes, featuring over 31,000 changed building instances. Similarly, when quantifying the changed and unchanged pixels across the entire dataset, our analysis unveiled 31,066,643 changed pixels, constituting 4.65% of the total, while 636,876,269 unchanged pixels constituted 95.35% (as depicted in Figure 3). To facilitate our analyses, we uniformly cropped these images into non-overlapping pairs of size 256 × 256. As a result, we accumulated a total of 10,192 image pairs. It is worth highlighting that, unlike the WHU-CD dataset, a substantial 44.40% of these pairs contained changes within their respective change maps. This balanced distribution between changed and unchanged maps distinguishes the LEVIR-CD dataset from its WHU-CD counterpart, which suffers from an imbalanced representation. To ensure consistent experimentation conditions, we performed a random split, allocating the dataset into training, validation, and test subsets using a 7:1:2 ratio, respectively.

3.1.2. Data Preparation for Remote Sensing Image Generation

Our RS image generation framework aims to generate synthetic RS images containing buildings. To achieve this objective, we initiated the process by assembling the requisite training data for our image generator model from both LEVIR-CD and WHU-CD.

Taking advantage from the provided buildings labels in the original bitemporal WHU-CD dataset images, we collected real remote sensing images with their corresponding building labels. In a systematic manner, we partitioned the original bitemporal images and their labels into discrete patches, each measuring 256 × 256. Only those patches that featured buildings were retained. This meticulous process yielded a total of 9542 pairs of real images along with their associated building labels, which served as the foundation for training our image generator model. For the LEVIR-CD dataset, we collected the training dataset by extracting images from the original CD dataset. Specifically, we selected images from the post-change instances that exhibited building modifications. This process yielded a total of 3160 pairs of real images with their corresponding building labels.

3.2. Data Augmentation Strategy

This section presents our data augmentation strategy, aimed at reinforcing the dataset's composition by infusing synthetic images predominantly featuring buildings. The core objective of data augmentation is to redress the balance within the CD dataset, striving for equilibrium between samples embodying changes and those without changes. To achieve this, we employ a conditional Generative Adversarial Network (c-GAN) to fabricate new samples replete with building-related changes. Our data augmentation approach comprises two pivotal components: the Building Change Detection Image Generator model and the Buildings Label Creation.

3.2.1. Building Change Detection Image Generator Model

Our model operates within the purview of the conditional GAN (c-GAN), wherein the architectural framework capitalizes on the building pattern within the input image as a conditioning factor to generate a fake output image rich in building features. The model encompasses two integral components: a generator (G) and a discriminator (D) [7]. In our context, the generator (G) undertakes the responsibility of translating a building label image into a tangible RS image brimming with building structures. Simultaneously, the discriminator (D) is tasked with distinguishing authentic images from the artificially generated counterparts.

The model is trained in a supervised manner. The training dataset assumes the form of corresponding image pairs denoted as ($\mathrm{l_i}$, $\mathrm{r_i}$), where $\mathrm{l_i}$ signifies the building label image and $\mathrm{r_i}$ represents the associated authentic remote sensing image. Throughout the training phase, the generator improves its ability to create more realistic samples to deceive the discriminator. In parallel, the discriminator is fortified to perceive the subtle nuances that differentiate authentic images from their fabricated counterparts. This dynamic interplay between generator and discriminator transpires through the following objective function designed to measure the degree of authenticity for generated images,

$$L(G,D) = Arg\ min_G max_D L_c(G,D) + \lambda L_{L_1}(G), \tag{1}$$

where L_c is the commonly used objective function of the c-GAN, defined as:

$$\mathbb{E}_{(l,r)}[\log(D(l,r))] + \mathbb{E}_l[\log(1 - D(l,G(l)))]. \tag{2}$$

The variable L_{L_1} embodies the disparity between the generated image and the real image. This additional loss encourages the generator model to create plausible translations of the input label image. The formulation of L_{L_1} is expressed as follows:

$$L_{L_1}(G) = \mathbb{E}_{(l,r)}[|r - G(l)|]. \tag{3}$$

The generator model is based on a U-Net architecture, while the discriminator adopts a patch-based fully convolutional network, as outlined in [7]. Our training approach couples adversarial learning for both generator and discriminator, aimed at mapping the input building labels to high-resolution images by optimizing the adversarial loss function expressed in Equation (1). The adversarial loss function encourages the generator to produce images that are indistinguishable from real images, thereby ensuring that the generated images are of high quality and contain realistic building changes. Nonetheless, we encountered challenges with image quality and training stability, both of which fell below the desired performance.

To surmount these hurdles and generate superior-quality remote sensing images, we introduce refinements to this foundational architecture. Specifically, for the generator, we replace the U-Net with a sequence of SPatially-Adaptive DEnormalization (SPADE) residual blocks fortified with upsampling layers [37]. Illustrated in Figure 4, the generator takes the latent vector as input, with each SPADE residual block incorporating the semantic label of the building to uphold its semantic integrity. This novel architecture yields heightened performance metrics with a reduced parameter count, achieved through the omission of downsampling layers. As for the discriminator, we adopt the architecture proposed in the pix2pixHD model [38], which uses a multiscale discriminator to engender high-resolution images. The multi-scale discriminator model consists of several discriminator networks operating at different spatial scales, enabling it to capture both local and global features of the input images. By using a multi-scale discriminator model, we can ensure that the generator model produces high-resolution images with building features that are consistent with the original ones.

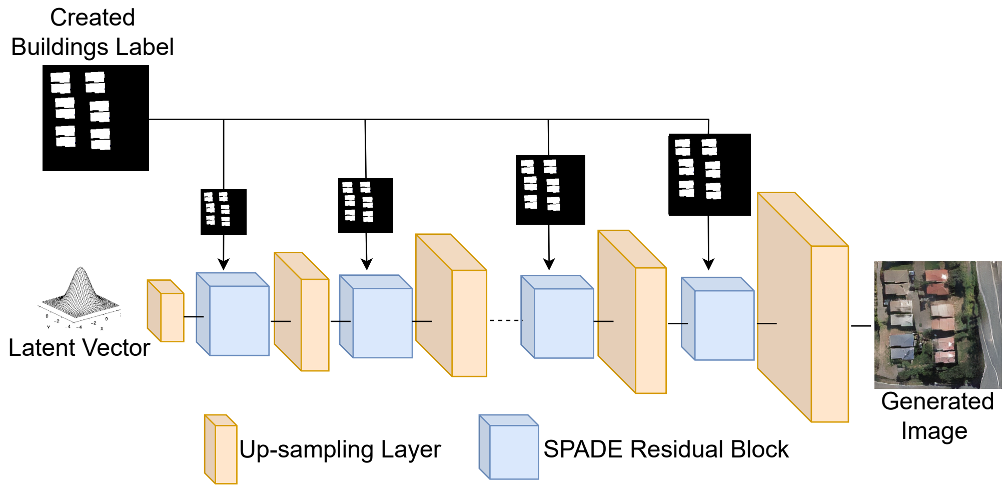


Figure 4. The architecture of the buildings generator.

To mitigate the training stability, the initial objective function, expressed in Equation (1), underwent transformation through the adoption of the Hinge loss [39]. Further modifications encompass the inclusion of feature matching loss and perceptual loss, concepts integral into the pix2pixHD model [38]. This resultant new loss function is defined as follows:

$$\begin{aligned} L(G,D) =& min_G max_D L_{H-GAN}(G,D) \\ &+ \lambda_F L_F(G,D) + \lambda_P L_P(G,D), \end{aligned} \tag{4}$$

where λ_F and λ_P denote weight factors that control the significance of each term. To mirror the original pix2pixHD model [38], both these weights were set at 10. Through these strategic refinements, we strive to bridge the gap between image quality and training stability, laying the foundation for a more robust and enhanced model.

3.2.2. Building Label Creation for Change Detection

Building labels are essential for accurately detecting changes in building structures within remote sensing imagery. By using diverse building masks, we can enhance the robustness and accuracy of our change detection model. These labels can be used to generate synthetic building changes, which can be added to the change detection training data. To generate novel building CD pairs, accompanied by their corresponding labels, via the previously outlined generation model, it becomes imperative to engender new building semantic labels or masks. These labels subsequently serve as the annotated CM for our comprehensive CD dataset. Our change detection dataset comprises pairs of images captured at distinct moments, denoted as T_0 and T_1, accompanied by their respective building CM. Notably, a considerable portion of these CM labels do not encompass building changes, thus instigating a disparity in the distribution of changed and unchanged images, as depicted in Figure 3.

In our work, showcased in Figure 5, we introduce building changes to the change maps that previously lacked such changes. The building label map thus created is then funneled into the generator to fabricate an image that integrates the presence of buildings. This ensuing generated image takes its place as a freshly generated image at T_1. In this construct, the image at T_0 is selected from the pair of images devoid of changes, while the generated image occupies the T_1 slot. Concurrently, the role of the change label is assumed by the newly created building label map. This iterative process continues until equilibrium is established between images manifesting changes and those lacking, which is a critical stride in fortifying the dataset's balance.

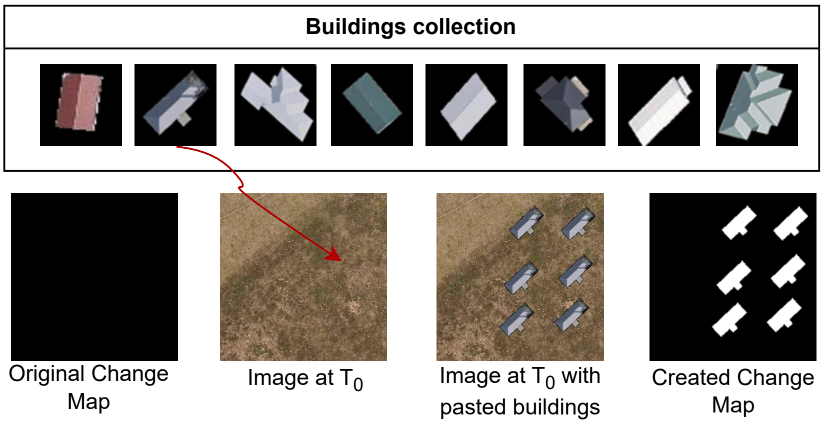


Figure 5. Illustration of building label creation.

The creation of the building label map relies on the copy–paste technique. As illustrated in Figure 5, the process commences with the compilation of diverse building silhouettes extracted from the original dataset. These cropped building images are subsequently resized to dimensions of (64 × 64), a modification that aligns their geometric attributes with those of the original dataset (256 × 256). The ensuing step involves embedding their corresponding semantic masks within the fresh CM. The choice of particular buildings and their spatial positions within the generated change map (CM) depends on the image assigned for T_0, which is a crucial aspect of the process.

To summarize, Algorithm 1 outlines the comprehensive procedure for generating new CD samples. Initially, we randomly select an image at T_0 from the set D_{uc}, which consists of images without any changes from the original dataset D. Subsequently, we introduce building masks, chosen from B, into the created label L_{cr}, aligning them with the chosen position in the selected image. This modified label is then input into our pretrained generator model G-model to produce a newly generated image at T_1 with building changes. The image selected at T_0, the generated image at T_1, and their respective created label are subsequently incorporated into the original dataset, augmenting the set D_c with changed images. This process is repeated until we reach the desired number of generated samples (N).

Algorithm 1: Building Change Samples Generation

```
Input: $D = \{D_C, D_{UC}\}$ (original training dataset with changed and unchanged
    images); $D_C = \left\{ \left(T^k_{c0}, T^k_{c1}, T^k_c\right) \middle| k = 1 : K \right\}$ (a set of changed images with a
    size of $K$); $D_{UC} = \left\{ T^l_{uc0}, T^l_{uc1}, T^l_{uc} \middle| l = 1 : L \right\}$ (a set of unchanged images
    with a size of $L$ where $L \gg K$ for imbalanced dataset); $B = \left\{ b^i \middle| i = 1 : I \right\}$ (a
    set of extracted building masks with a size of $I$); $N$ (the desired number of
    images to be generated); $G$-model (the trained building generator model).
Output: $D_g = \left\{ \left(T^n_{uc0}, T^n_{g1}, L^l_{cr}\right) \middle| n = 1 : N \right\}$ (a set of generated changed images
    with a size of $N$); $D_{Aug}$ (augmented CD training set).
Initialize $D_g \leftarrow \phi$; $D_{Aug} \leftarrow D$; $L^n_{cr} \leftarrow L^n_{uc}$;
for $n \leftarrow 1$ to $N$ do
    // perform image generation
    $T^n_{uc0} \leftarrow$ select an image at $T_0$ from $D_{UC}$;
    while true do
        $b^i \leftarrow$ select building mask $i$ from $B$;
        $h, w \leftarrow$ size of $b^i$;
        $(x, y) \leftarrow$ seclect position in $T^n_{uc0}$ where to paste the building $b^i$;
        $L^n_{cr}(x : x + h, y : y + w) \leftarrow b^i$
    end
    $T^n_{g1} \leftarrow G-$model$(L^n_{cr})$;
    $D_g \leftarrow D_g \cup \left(T^n_{uc0}, T^n_{g1}, L^l_{cr}\right)$;
    $D_{Aug} \leftarrow D_{Aug} \cup D_g$
end
```

3.2.3. New Sample Generation for the Change Detection Dataset

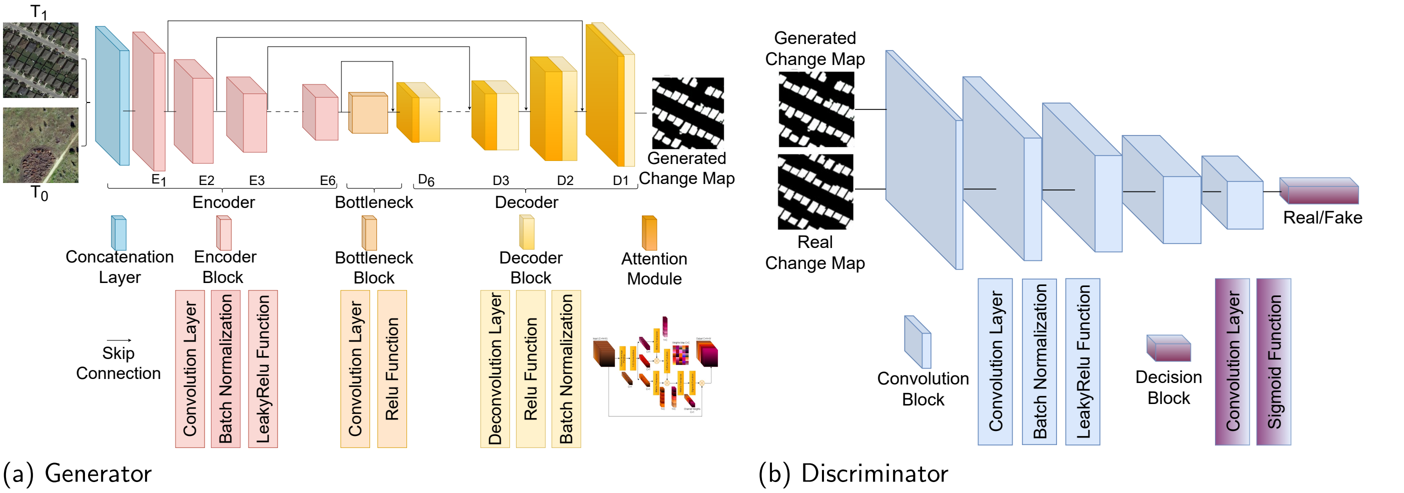
Aiming to achieve the desired ratio between changed and unchanged maps, we judiciously determined the count of image pairs that would incorporate changes in their change maps. As elaborated in the previous subsections, building masks are incorporated into the change maps that originally lacked building changes. This step yielded updated change maps, which subsequently acted as inputs to our pre-trained generator model, thus generating synthetic RS images at T_1. The original image at T_0, the corresponding generated image at T_1, and the newly created building CM were added to our CD dataset. This strategic combination successfully increased the quantity of changed maps, achieving a balance between the subsets of changed and unchanged images.

3.3. Building Change Detection Model

3.3.1. Adversarial Change Detection Network (Adv-CDNet)

Our approach for building change detection relies on an adversarial training paradigm, as depicted in Figure 1. The crux of our strategy is to treat change detection as an image-to-image translation task by translating the combined bi-temporal images from the input domain to produce a CM in the output. Opting for the Pix2Pix model, which excels at domain translation, we introduce a customized version of this model to execute the transformation of two-period images into a change detection map.

As illustrated in Figure 6, our model consists of two crucial components: the generator and the discriminator. The generator (Figure 6a) is based on the U-Net architecture, which features encoder–decoder design and skip connections. These connections align encoder layers with corresponding decoder layers, ensuring that feature maps of the same dimensions are maintained. The U-Net framework effectively maps fused input images onto the latent feature space, capturing feature variations. The generator combines shallow and deep semantic features by harnessing skip connections, enhancing the information encapsulated within it. In contrast, the discriminator (Figure 6b) adopts a simple patch-based convolutional network form. Armed with either real ground truth change maps or synthetic change maps generated by the generator, the discriminator evaluates the authenticity of the presented change maps. It assigns binary values (0 or 1) to differentiate between change maps produced by the generator and real ground truth. Our objective function remains consistent with Equation (1). The real images ($\mathrm{r_i}$) are replaced by real change maps (ground truth), and the combined bi-temporal images replace the label image ($\mathrm{l_i}$). Crucially, the evaluation of the distance between the produced CM and the ground truth employs a loss function weighted by λ = 100. Employing the Adam optimizer with a learning rate of 0.0002 and β = 100 facilitates our optimization.



(a) Generator

(b) Discriminator

Figure 6. Building change detection model composed of generator and discriminator networks. The generator aims to generate a CM and the discriminator examines whether the generated CM is real or fake.

3.3.2. Channel Attention Module

In our work to emphasize the model's focus on changed features, we introduced a channel attention module to the decoder blocks, as illustrated in Figure 6a. This module harnesses the power of channel connections and channel weight recalibration to refine intricate features. The recalibrated channel-wise feature responses rely on newly generated channel weights, fostering a more nuanced representation. Noteworthy is the fact that channel attention, as elucidated in [40,41], holds immense potential for enhancing the performance of deep convolutional neural networks. This is particularly relevant in the context of RS change detection DL models, where its efficacy has been demonstrated [26,28,42].

As shown in Figure 7, in alignment with the Channel Self-Attention module introduced by Wang et al. [28], the process starts by the application of Global Average Pooling (GAP) to the input feature maps characterized by their dimensions (Channel (C) × Height (H) × Width (W)). Concretely, GAP computes the mean value for each feature map, as formulated by:

$$F_{GAP} = \frac{1}{W * H} \sum_{i}^{W} \sum_{j}^{H} F_{in}(i, j). \tag{5}$$

Subsequently, as articulated in Equation (6), the rest of the calculations encompass a 1 × 1 convolution ($conv_{(1D)}$), matrix multiplication operations (*), transpose operations (T), and the employment of two activation functions (Softmax and Sigmoid).

$$\begin{aligned} F_{out} =& Sigmoid((Softmax(conv_{1D}(F_{GAP}) \\ & * conv_{1D}(F_{GAP})) * (conv_{1D}(F_{GAP}))^{T})^{T}) * F_{in}. \end{aligned} \tag{6}$$

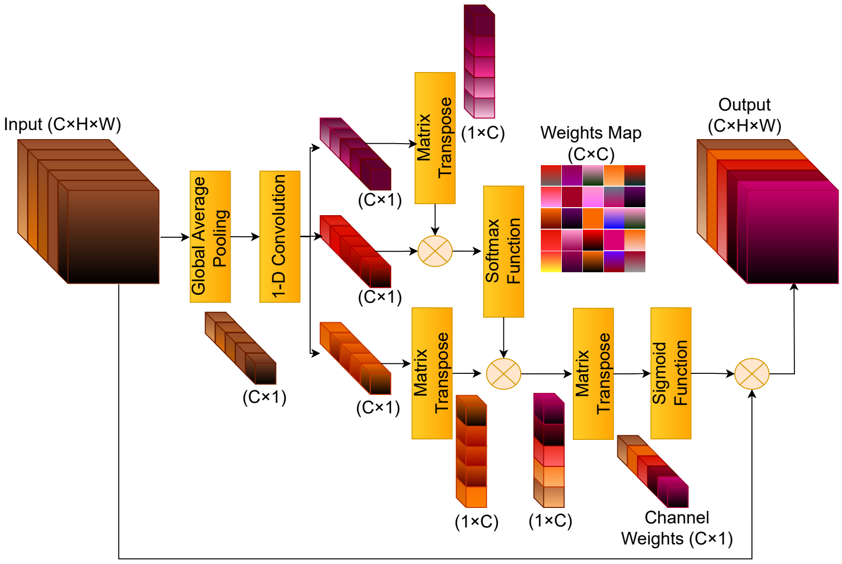


Figure 7. Channel attention module that can refine the detailed feature.

4. Experiments and Results

4.1. Experimental Setup

The implementation of change detection and image generation models was carried out using Torch. The CD model was executed on a single Nvidia GeForce RTX 3090 GPU, which is manufactured by Yunxuan Ltd. from Shanghai, China. In contrast, the image generation model utilized the power of two Nvidia GeForce RTX 3090 GPUs for increased processing capability due to its higher complexity compared with the CD model.

4.2. Evaluation

The evaluation of our change detection framework included the utilization of five key assessment metrics: Overall Accuracy (OA), Intersection over Union (IoU), Precision (Pre), Recall (Re), and F1-Score ($F1$).

$$OA = \frac{TP + TN}{N} \tag{7}$$

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

$$Pre = \frac{TP}{TP + FP} \tag{9}$$

$$Re = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2\ Pre\ Re}{Pre + Re} \tag{11}$$

where TP, FP, TN, FN, and N correspond to the counts of true positives, false positives, true negatives, false negatives, and the total number of pixels, respectively. These metrics collectively provided a comprehensive understanding of the framework's performance.

In the specific context of change detection, a noteworthy precision value signifies a limited occurrence of false alarms, while a substantial recall value indicates minimal instances of missed detections. Simultaneously, the F1-Score and overall accuracy serve as holistic performance indicators, with higher values indicative of superior performance. The intersection over Union metric gauges the degree of alignment between the predicted CM and the Ground Truth. By harnessing this comprehensive suite of metrics, our evaluation methodology offered a well-rounded perspective on the efficacy and capabilities of our change detection framework.

4.3. Comparison with the State-of-the-Art Change Detection Methods

To comprehensively assess the performance of our proposed change detection model, we conducted both quantitative and qualitative comparisons with State-of-the-Art (SOTA) change detection methods. These methods serve as benchmarks against which the efficacy of our model can be measured. The following SOTA methods were selected for evaluation:

1. Fully Convolutional Early Fusion (FC-EF) [43]: This method employs image-level fusion based on the U-Net architecture. The bi-temporal images are concatenated into a single input for the U-Net model, facilitating holistic feature extraction.
2. Fully Convolutional Siamese Concatenation (FC-Siam-Conc) [43]: In contrast to FC-EF, FC-Siam-Conc adopts feature-level fusion. It leverages two encoders with shared weights to extract features from bi-temporal images, concatenating them to the decoder at the same level.
3. Fully Convolutional Siamese Difference (FC-Siam-Diff) [43]: This method shares similarities with FC-Siam-Conc, differing primarily in the formation of skip connections. Instead of simple concatenation, FC-Siam-Diff transports the absolute value of the difference between bi-temporal features to the decoder.
4. Bitemporal Image Transformer (BIT) [44]: This network captures contextual information within the spatial–temporal domain. By leveraging transformer, BIT effectively models contexts between different temporal images, enhancing its ability to analyze and interpret complex spatial–temporal relationships.
5. Spatial–Temporal Attention Neural Network (STANet) [9]: STANet represents a metric-based Siamese FCN approach, enhanced with a spatial–temporal attention module to extract more discriminative features.
6. Hierarchical Attention Network (HANet) [26]: This model is a discriminative Siamese network, featuring a hierarchical attention network (HAN) with a lightweight and efficient self-attention mechanism, which is designed to integrate multiscale features and refine detailed features.

Analysis of Table 1 reveals our Adv-CDNet boasts a higher parameter count than the three basic Siamese Networks and BIT, underscoring its ability to learn and represent complex patterns effectively. Interestingly, when coupled with the attention module, it exhibits fewer parameters than the intricate STA-Net, showcasing a balance between complexity and efficiency. Furthermore, our model demonstrates lower FLOPs compared to HANet and STANet, emphasizing computational efficiency.

Upon reviewing the experimental outcomes detailed in Table 1 for both WHU-CD and LEVIR-CD datasets, it becomes apparent that our method delivers satisfactory performance on the LEVIR-CD dataset, even without the attention module. The OA, IoU, Pre, Re, and $F1$ were 98.62%, 74.85%, 90.95%, 80.88%, and 85.62%, respectively. Compared to the three baseline networks (FC-EF, FC-Siam-Conc, and FC-Siam-Diff), our method provided improvements of 6.99%, 2.83%, and 8.23% in terms of the F1-Score, respectively. These results can be confirmed through the visual interpretations shown in Figure 8, where we see limited occurrences of false alarms (FP) and missed detections (FN) in our model's results compared to the sub-mentioned networks. The decreased FP and FN explain the improved precision and recall, respectively, leading to the increase in F1-Score. Similarly, as shown in Table 1, the test results of our Adv-CDNet on the WHU-CD dataset also outperform state-of-the-art methods, including FC-EF, FC-Siam-Conc, and FC-Siam-Diff, across various evaluation metrics such as Overall Accuracy, Intersection over Union, precision, and F1-Score. These results can be confirmed with the visual interpretation from Figure 8, where

FC-Siam-Conc and FC-Siam-Diff show a significant amount of occurrence of false positives, which explains the decrease of their precision compared to our model.

Table 1. Comparative results with other SOTA methods in terms of FLOPS, Parameters (Param), Overall Accuracy (OA), Intersection over Union (IoU), Precision (Pre), Recall (Re), and F1-Score ($F1$) on the LEVIR-CD and WHU-CD test sets. The highest metrics values in each dataset are marked in bold.

Methods	FLOPs (G)	Param (M)	WHU-CD					LEVIR-CD				
			OA	IoU	Pre	Re	$F1$	OA	IoU	Pre	Re	$F1$
FC-EF	2.29	1.35	95.30	46.12	58.47	68.58	63.12	97.81	64.79	78.16	79.12	78.63
FC-Siam-Conc	2.29	1.54	87.00	40.51	54.00	78.23	63.89	98.33	70.64	87.25	78.77	82.79
FC-Siam-Diff	2.29	1.35	96.77	49.91	60.24	51.53	55.54	97.91	63.11	85.99	70.35	77.39
Adv-CDNet	13.41	6.14	98.20	50.43	77.74	56.22	65.26	98.62	74.85	**90.95**	80.88	85.62
BIT	4.35	3.55	96.96	57.49	79.58	65.49	71.85	97.81	75.65	88.84	83.05	85.85
STA-Net	13.88	16.92	93.18	56.63	61.83	**82.57**	70.71	95.92	75.01	77.79	**95.56**	84.10
HANet	14.07	3.03	97.35	58.34	82.14	62.96	71.28	98.63	75.85	87.45	85.12	86.27
Adv-CDNet + Attention	13.41	6.80	**98.53**	**59.83**	**84.21**	63.02	**72.10**	**98.67**	**76.10**	90.54	82.92	**86.56**

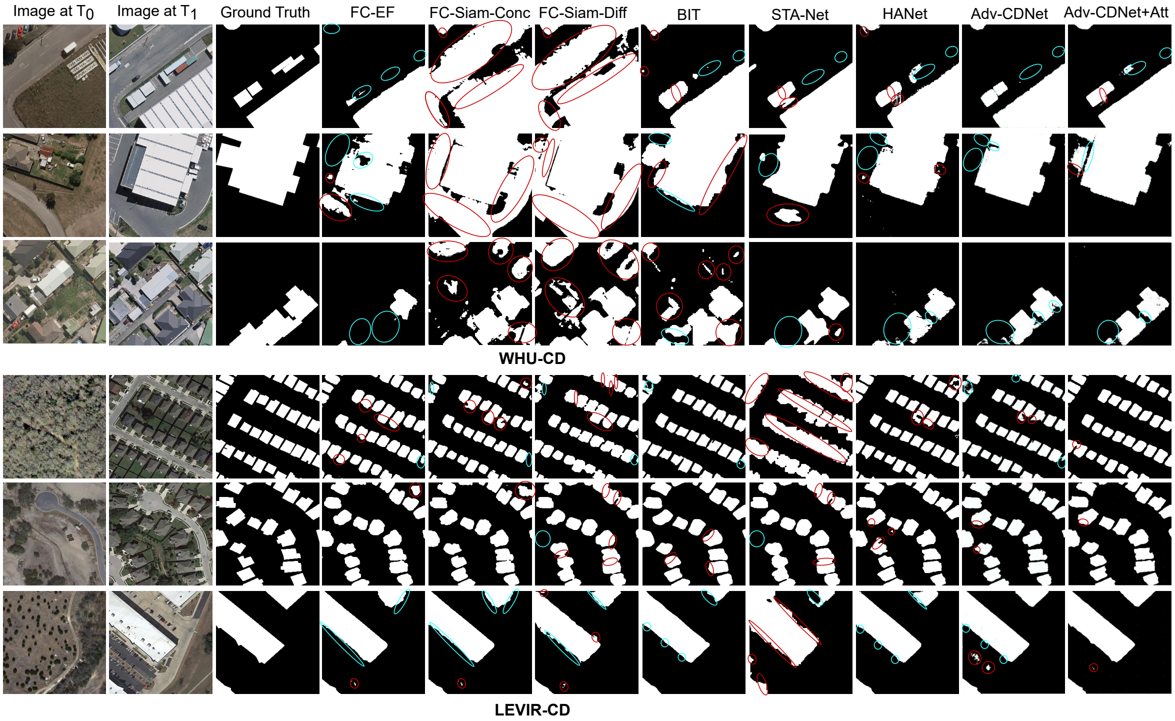


Figure 8. Comparative visualization of different methods on the LEVIR-CD and WHU-CD test sets. For easier comparison, some of the relevant detection errors have been marked with red circles for false positives (FP) and blue circles for false negatives (FN).

The main reason for these results is related to the fact that most of these sub-mentioned SOTA methods use a Siamese network, which is a double-stream framework that generates change maps based on feature differences between two images. Therefore, these methods are highly dependent on high contrast between the two images, in contrast to our model, which is a single stream framework that can map directly the two input images into a building change map. This approach leads to more efficient feature extraction and change detection, as it eliminates the need for separate processing and alignment of the two images used in other methods. Moreover, the generator part of the Adv-CDNet utilized U-Net architecture to generate a change map from input images and skip connection to fuse shallow and deep feature representations. These allow the proposed model to be able to recognize complex features that are difficult to extract using the aforementioned methods. This is especially helpful in recognizing changes in some complicated scenarios. However, when we compare its performance for the two datasets (LEVIR-CD and WHU-CD), a notable discrepancy emerges. Specifically, the F1-Score exhibits a substantial 20% difference between the two datasets, underscoring the adverse impact of class imbalance in WHU-CD on the model training. As illustrated in Figure 8, our model exhibits superior performance on LEVIR-CD, with more effective detection of building changes compared to WHU-CD. Especially in cases of very subtle changes, the model struggles to detect building alterations in WHU-CD.

Including the attention module in our Adv-CDNet has shown improvements for LEVIR-CD in the OA, IoU, Re, and $F1$, as shown at the bottom of Table 1. We can observe that our model outperformed the BIT network in terms of OA, IoU, and $F1$. Moreover, it also outperformed STA-Net, which uses an attention mechanism, with 2.75%, 1.09%, and 2.46% in terms of OA, IoU, and $F1$, respectively. Meanwhile, there was a trade-off in precision and recall where our change detection model provided a 12.75% improvement in precision, while STA-Net outpaced our network by 12.64%. We can also note the similar qualitative and quantitative performance between our Adv-CDNet with attention and the HANet model in terms of OA, IoU, and $F1$.This similarity is likely because both models incorporated the same attention mechanism. Similarly, the incorporation of an attention module into our model for WHU-CD training has yielded significant performance enhancements. We observe significant improvements in IoU (9.4%), precision (6.47%), recall (6.8%), and F1-Score (6.84%). These improvements are very high compared to those achieved on the LEVIR-CD dataset, where the gains are 1.25% for IoU, 2.02% for recall, and 0.94% for F1-Score. This highlights the substantial influence of the attention module, particularly when dealing with severely imbalanced datasets such as WHU-CD in our case.

Notably, the introduction of the attention module amplifies its impact on WHU-CD more than LEVIR-CD, underscoring its efficacy in addressing dataset imbalances, as seen in this challenging dataset. The enhanced performance observed with the attention module stems from its ability to establish relationships among individual channels and recalibrate feature responses on a per-channel basis. This functionality enables the model to concentrate its training efforts on more pertinent features, fostering improved deep representations. Essentially, it facilitates the creation of channel connections and the recalibration of channel-wise feature responses. In practical terms, this empowers the network to boost performance by amplifying the response to semantic changes while constraining the impact of non-changes.

4.4. Data Augmentation

In this section, we examine our generation model's performance through two key aspects: visual evaluation of the generated images and their impact when integrated into the original change-detection dataset for training our Adv-CDNet. Initially, we employ visual interpretation to compare outcomes produced by our generator model with those from the copy–paste method. Subsequently, we delve into assessing the qualitative results stemming from training our model on augmented datasets utilizing both our augmentation method and the conventional approach. The detailed findings are presented in Figure 9 and Table 2.

As shown in Figure 9, our findings reveal a stark contrast between the quality of generated images produced by our generator compared to those generated through the copy–paste method [16] for both WHU and LEVIR datasets. Notably, images generated using our model exhibit a striking resemblance to reality, complete with intricate details such as the inclusion of shadows on buildings. In contrast, the copy–paste method fails to account for the surrounding environment, resulting in buildings appearing detached and unnatural.

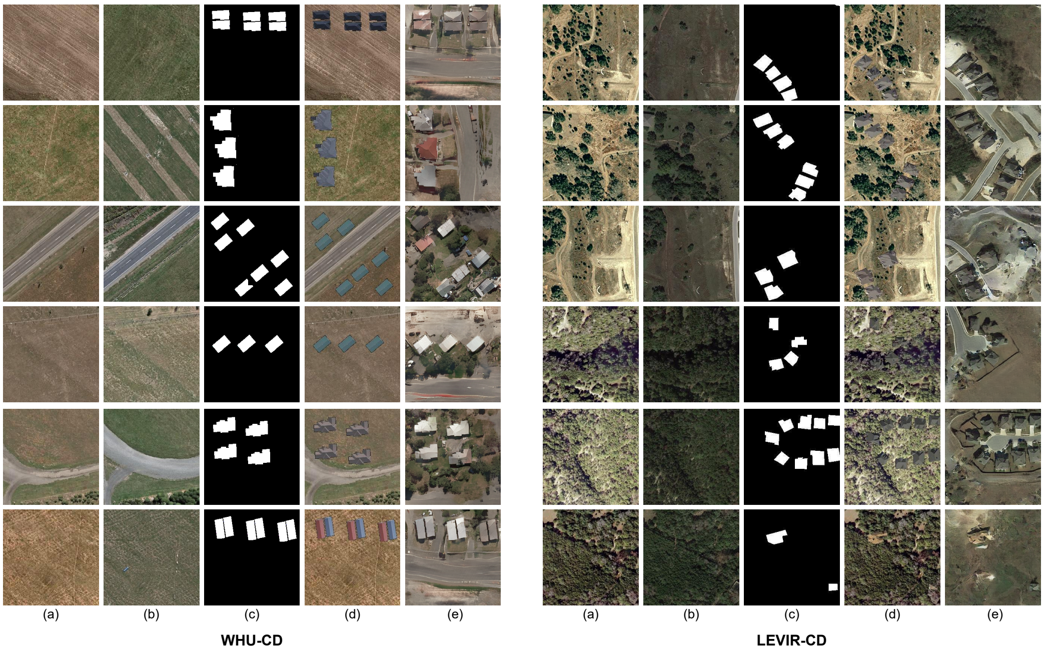


Figure 9. Comparative visualizations of WHU-CD and LEVIR-CD image generation. (**a**) Image at T_0, (**b**) Original image at T_1, (**c**) Created building label, (**d**) Created image at T_1 using Copy-Paste, and (**e**) Generated image at T_1 using our building generator.

Furthermore, our generator showcases its versatility by extending its capability to create realistic representations of roads and trees surrounding the buildings, enhancing the overall contextual fidelity of the generated scenes. It is worth noting that the performance of our model can be further improved by expanding the training dataset to include a larger number of images and by training for an extended number of epochs. These refinements hold the potential to elevate the model's ability to capture even finer nuances of the urban environment in generated imagery.

Table 2. Assessing model performance on WHU-CD and LEVIR-CD: our data augmentation strategy vs. traditional methods. The highest classification accuracy is marked in bold.

Methods	WHU-CD					LEVIR-CD				
	OA	IoU	Pre	Re	$F1$	OA	IoU	Pre	Re	$F1$
Adv-CDNet without data augmentation	98.20	50.43	77.74	56.22	65.26	98.62	74.85	**90.95**	80.88	85.62
Adv-CDNet with traditional augmentation	98.23	53.36	84.05	59.37	69.58	98.11	74.91	89.98	81.63	85.60
Adv-CDNet with our augmentation	**98.61**	**71.87**	**85.55**	**78.29**	**81.76**	**98.64**	**75.24**	90.24	**82.75**	**86.33**

Table 2 presents a comprehensive evaluation of our Adv-CDNet model's performance across three distinct scenarios. First, we examined its performance without any data augmentation, meaning the model was trained on the original datasets. Second, we applied traditional data augmentation techniques, including copy–paste [16] rotation, reflection, and color saturation [11] to increase the number of changed images and create balanced datasets. Finally, we assessed its performance using our proprietary data augmentation approach for balancing the original data. Notably, both data augmentation approaches improved model performance on WHU-CD. However, our method demonstrated substantial superiority, with improvements of 21.44% in IoU, 7.81% in Pre, 22.67% in Re, and 16.50% in $F1$ compared to just 2.93%, 6.31%, 3.15%, and 4.32% with traditional methods, respectively. These results highlight the remarkable efficacy of our data augmentation technique, which has a more pronounced positive impact on change detection model performance compared to conventional methods. In contrast, data augmentation resulted in a lower performance improvement of the model when trained on the augmented LEVIR-CD dataset compared to WHU-CD. This discrepancy is due to the difference in the original distribution of changed and unchanged samples in both datasets. Specifically, only 25% of change samples are present in WHU-CD compared to 44.6% in LEVIR-CD. Consequently, the number of change samples in LEVIR-CD was not augmented as dramatically as in WHU-CD, thereby explaining the greater impact of data augmentation on model performance when trained on WHU-CD compared to LEVIR-CD.

To assess the impact of our data augmentation method on Deep-Learning (DL) models trained on imbalanced and balanced datasets, we conducted experiments using State-of-the-Art (SOTA) methods such as BIT, STA-Net, and HANet alongside our ADV-CDNet with attention. Each model was trained with and without data augmentation. The results from Table 3 demonstrate performance enhancements across all models when trained on augmented datasets for both WHU-CD and LEVIR-CD. Notably, the improvements are more pronounced in WHU-CD compared to LEVIR-CD. For instance, in terms of F1-Score, the enhancements for WHU-CD are substantial, with increases of approximately 8.67%, 9.63%, 10.95%, and 11.17% for BIT, STA-Net, HANet, and ADV-CDNet with attention, respectively. In contrast, the improvements for LEVIR-CD are more modest, with gains of approximately 0.65%, 2.05%, 0.29%, and 0.15% for the same models, respectively. These findings underscore the significant impact of data augmentation on model performance, particularly when dealing with highly imbalanced original datasets.

Table 3. Assessing the Impact of the Data Augmentation (DA) method on the performance of BIT, STA-Net, HANet, and our Adv-CDNet with attention when trained with and without DA.

Methods	WHU-CD					LEVIR-CD				
	OA	*IoU*	*Pre*	*Re*	*F1*	*OA*	*IoU*	*Pre*	*Re*	*F1*
BIT	96.96	57.49	79.58	65.49	71.85	97.81	75.65	88.84	83.05	85.85
BIT with DA	98.42	70.20	82.35	78.76	80.52	98.22	76.35	89.94	83.33	86.50
STA-Net	93.18	56.63	61.83	82.57	70.71	95.92	75.01	77.79	**95.56**	84.10
STA-Net with DA	96.11	69.92	75.98	**85.20**	80.33	96.56	76.11	78.53	95.41	86.15
HANet	97.35	58.34	82.14	62.96	71.28	98.63	75.85	87.45	85.12	86.27
HANet with DA	98.63	71.72	83.65	80.86	82.23	98.65	76.37	90.54	82.92	86.56
Adv-CDNet + Attention	98.53	59.83	84.21	63.02	72.10	98.67	76.10	90.54	82.92	86.56
Adv-CDNet + Attention with DA	**98.99**	**72.95**	**86.94**	79.89	**83.27**	**98.70**	**76.74**	**91.12**	82.71	**86.71**

To further elucidate the influence of dataset composition on the change detection model, we conducted an extended series of experiments, systematically varying the ratio of changed to unchanged images within our datasets. As shown in Table 4, initially WHU-CD comprised 25% changed and 75% unchanged images, while LEVIR-CD contained 44.4% changed and 55.6% unchanged images. We incrementally augmented the proportion of changed images using our data augmentation approach to create an equivalent number to the unchanged images. Our findings revealed noticeable enhancements in model performance with both configurations, with and without the attention module, for both datasets. Continuing this trend, we progressively adjusted the datasets until reaching 75% changed images and 25% unchanged images. Interestingly, this particular configuration exhibited significant performance boosts for WHU-CD in both model variants. Notably, the model performed excellently without the attention module, demonstrating that a well-prepared dataset can render this extra module unnecessary. In contrast, this configuration did not improve model performance for LEVIR-CD compared to the previous one. However, when we explored using only changed images and deleting all unchanged ones, we observed stark declines in model performance compared to the prior configuration for both datasets. This underscores the importance of including unchanged images during training.

Table 4. Effects of changed image count vs. unchanged images in WHU-CD and LEVIR-CD datasets on our model performance.

Dataset	Changed/Unchanged Maps	Method	*OA*	*IoU*	*Pre*	*Re*	*F*1
WHU-CD	25%/75% (original data)	Adv-CDNet	98.20	50.43	77.74	56.22	65.26
		Adv-CDNet + Attention	98.53	59.83	84.21	63.02	72.10
	50%/50%	Adv-CDNet	98.61	71.87	85.55	78.29	81.76
		Adv-CDNet + Attention	98.99	72.95	86.94	79.89	83.27
	75%/25%	Adv-CDNet	**99.35**	**90.53**	**91.21**	**99.19**	**95.03**
		Adv-CDNet + Attention	99.17	88.03	90.3	97.22	93.63
	100%/0%	Adv-CDNet	98.84	84.14	85.21	98.53	91.34
		Adv-CDNet + Attention	98.95	85.49	86.18	99.07	92.17
LEVIR-CD	44.4%/55.6% (original data)	Adv-CDNet	98.62	74.85	90.95	80.88	85.62
		Adv-CDNet + Attention	98.67	76.10	90.54	82.92	86.56
	50%/50%	Adv-CDNet	98.64	75.24	90.24	82.75	86.33
		Adv-CDNet + Attention	**98.70**	**76.74**	**91.12**	82.71	**86.71**
	75%/25%	Adv-CDNet	98.59	74.56	89.95	81.33	85.42
		Adv-CDNet + Attention	98.48	73.62	86.24	**83.41**	84.80
	100%/0%	Adv-CDNet	98.02	67.44	80.33	80.78	80.56
		Adv-CDNet + Attention	98.08	68.53	80.38	82.30	81.33

In summary, our experiments underscore the significance of specific image ratios for optimal performance in change detection. For WHU-CD, the optimal configuration involves 75% changed and 25% unchanged images, while for LEVIR-CD a balanced distribution of 50% changed and 50% unchanged images yields the best results. These ratios are crucial in maximizing the effectiveness of our change detection model. These findings can be attributed to the characteristics of the original datasets. WHU-CD contains a relatively small number of changed buildings (3281) compared to the more extensive set in LEVIR-CD (31,000). Furthermore, the change maps in LEVIR-CD exhibit a more uniform distribution of building changes, in contrast to WHU-CD, where a majority of change maps depict smaller alterations. This discrepancy underscores the necessity for an augmented dataset when training the change detection model on WHU-CD, emphasizing the importance of introducing an excess of building change samples to capture the nuanced variations present in the data.

Figure 10 provides valuable insights into model performance for LEVIR and WHU CD across three illustrative examples. In the first two WHU-CD instances, we observe decreased false positives (FP) and false negatives (FN) when increasing the number of changed images in the dataset. This positively influences the model's precision and recall, respectively. The inclusion of an attention mechanism demonstrates effectiveness with the original dataset composition. However, the second example indicates that the attention mechanism has minimal effect when an adequate number of changed images are available. For LEVIR-CD, the first two examples show that the best change detection result occurs when the model is trained with an attention mechanism on balanced data. Moving to the third example, for both datasets it becomes evident that the improvement in model performance due to an increased number of changed images reaches a limit. This is characterized by emerging false positive pixels, subsequently reducing precision and model performances, as discussed in Table 4. Overall, these findings underscore the dynamic relationships between dataset composition, attention mechanisms, and model performance. They provide nuanced insights into predicted image interpretation.

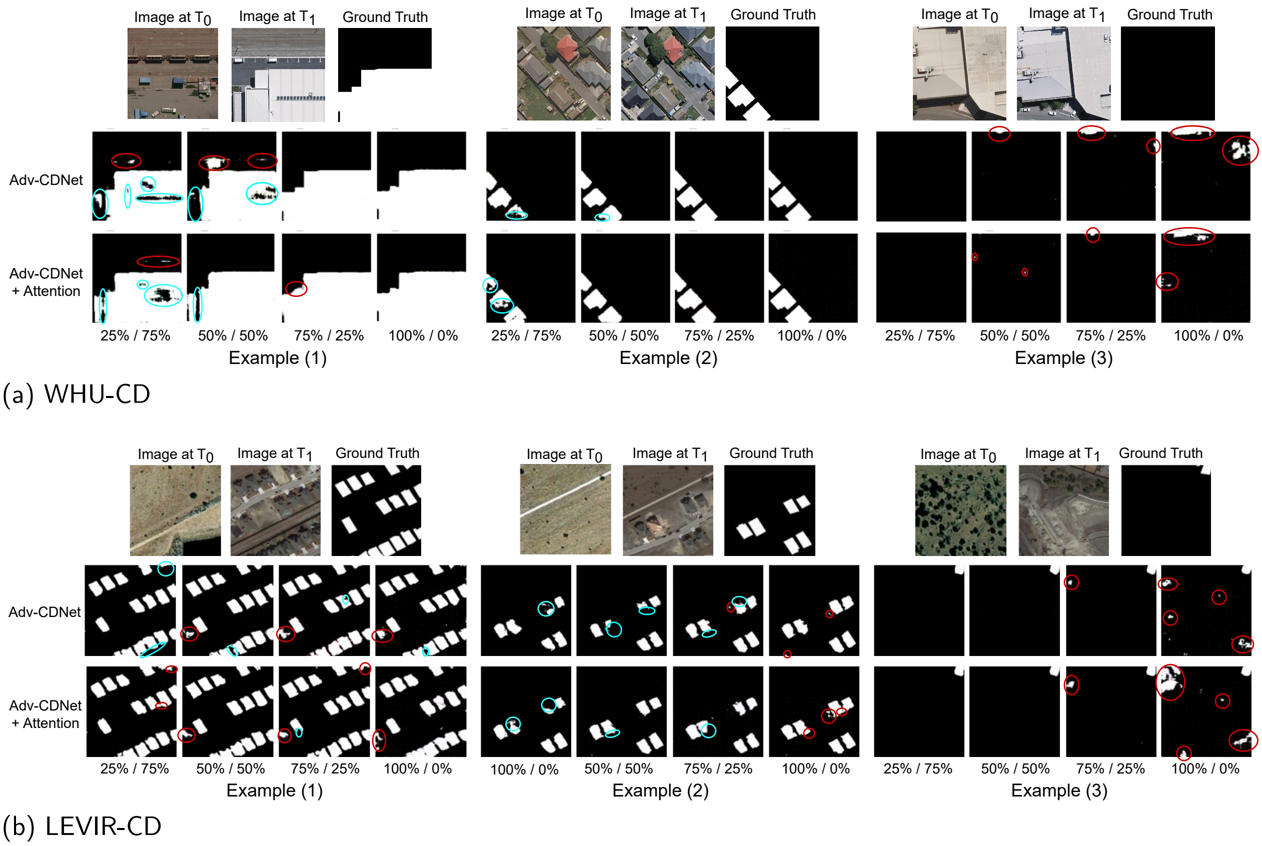


Figure 10. Qualitative results of the impact of increasing the number of changed samples in the training dataset on the prformance of Adv-CDNet with and without attention. The results of testing on the WHU-CD and LEVIR-CD test sets are illustrated in three examples for each data. For easier comparison, some of the relevant detection errors have been marked with red circles for false positives (FP) and blue circles for false negatives (FN).

5. Discussion

The Adv-CDNet has demonstrated competitive performance in detecting changes in remote sensing images. However, it faces challenges in detecting very subtle changes, particularly in building alterations, when dealing with highly imbalanced data such as the WHU-CD dataset, as discussed in the previous section. To enhance change detection in such data, the incorporation of an attention module has been proposed. The attention module is designed to selectively focus on the most informative features by assigning higher weights to the features that are most relevant to the change detection task. This is achieved through the use of channel connections and channel weight recalibration, which refine intricate features and are then used to modulate the generator's output. By selectively focusing on the most informative features, the attention module helps the generator produce more accurate change maps, even in cases of highly imbalanced data.

In addition to the attention module, another approach to enhance the performance of Adv-CDNet in imbalanced data is the creation of extensive labeled datasets and the generation of remote sensing images to augment the change detection dataset. Labeled datasets play a pivotal role in supervised learning tasks, enabling models to learn patterns and make accurate predictions. In the context of building CD, having a diverse and extensive labeled dataset is essential for training models to detect building changes accurately across various environmental conditions. Moreover, the use of generator models based on GAN for data augmentation has significantly enhanced the performance of Adv-CDNet by generating synthetic data samples that can supplement the original dataset. GANs are particularly useful for addressing data scarcity issues and improving model generalization by creating additional training examples. By leveraging GANs for data augmentation, the robustness and accuracy of CD model has been enhanced, especially in scenarios where the dataset is highly imbalanced.

The obtained findings illuminate a delicate trade-off between leveraging attention modules to optimize model performance and enhancing dataset effectiveness through data augmentation techniques. Each method carries its distinct advantages and limitations. employing attention modules to enhance change detection model performance emerges as particularly advantageous when dealing with imbalanced datasets. In these instances, the attention mechanism can significantly bolster the model's ability to discern changes. However, its impact becomes marginal when the dataset achieves a balanced state, primarily adding complexity and latency without yielding substantial performance improvements compared to non-attention models. This nuanced understanding underscores the significance of dataset balance and the judicious application of attention mechanisms in the context of change detection models.

On the other hand, data augmentation, particularly through GANs, proves invaluable in augmenting the dataset with effective training samples. By introducing diverse building changes to images, this technique enhances the diversity of the dataset, augmenting both its quality and quantity within the realm of change detection. These samples closely align with real-world scenarios, facilitating training on a higher number of high-quality images. Furthermore, the training loss plot of the change detection model in Figure 11 shows that the use of our data augmentation method led to faster model stability compared to other approaches for addressing data imbalance. Specifically, Adv-CDNet with our data augmentation strategy yielded stability in fewer epochs than using an attention module or traditional augmentation techniques. This more rapid stabilization also helps address overfitting issues. However, it is essential to note that GAN-based augmentation requires its own set of training samples to fine-tune the generative model, a task facilitated in our WHU-CD dataset through building labels in conjunction with before and after images. Nonetheless, datasets lacking such supplementary information may necessitate a dedicated effort to construct an appropriate training dataset for the generative model. For example, in the case of LEVIR-CD, a set of images with their building label is meticulously extracted from the primary CD dataset. Moreover, the generator model needs further improvement to generate more realistic RS images.

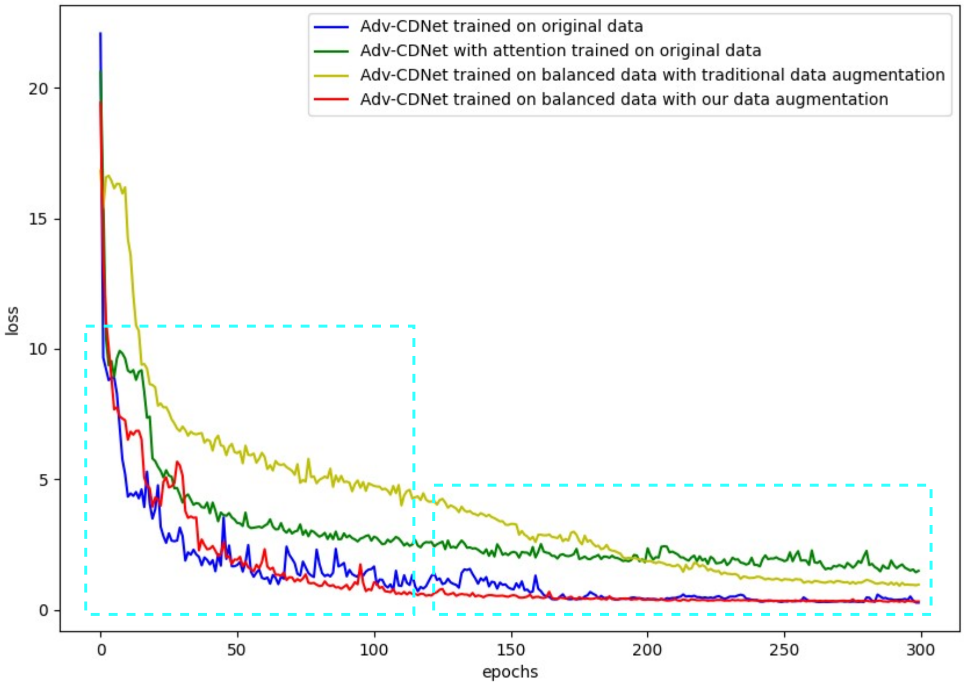


Figure 11. Training loss of different training configurations of Adv-CDNet model on WHU-CD dataset.

6. Conclusions

This paper proposed a data augmentation strategy and attention mechanism to enhance the performance of change detection-based adversarial learning frameworks when training on imbalanced data. The proposed data augmentation strategy aims to synthesize effective building CD samples to mitigate the data imbalance. This approach relies on a GAN-based technique to generate realistic building images, guided by created building label inputs, facilitating the generation of diverse images containing building changes. To evaluate the effectiveness of the proposed data augmentation method, a change detection model rooted in adversarial learning called Adv-CDNet has been proposed. This model is primarily based on the Pix2Pix architecture. Furthermore, we enhanced this model by incorporating a channel attention module to amplify its performance. The combination of Adv-CDNet with the attention module produced exceptional results, outperforming several state-of-the-art CD methods on the imbalanced WHU-CD and LEVIR-CD datasets. Our rigorous experimentation has underscored the efficacy of our proposed data augmentation methodology, highlighting its potential in addressing class imbalance.

Throughout this study, we have demonstrated the effectiveness of leveraging Generative Adversarial Networks for both generating new samples and detecting changes, by reconfiguring the baseline model to accommodate these dual tasks. However, it is crucial to acknowledge the existing dataset's limitation, as our labels are predominantly based on building object detection. Future extensions could explore the inclusion of other urban objects for a more comprehensive approach. Moreover, the generated images still require improvements in terms of both the quality of the generated building objects and the surrounding environment. Further research avenues could focus on enhancing the generator model's architecture to attain higher-quality synthetic data with less computational complexity. Lastly, while our change detection model surpasses current state-of-the-art methods, it is acknowledged that room for improvement remains. In this regard, exploring alternative loss functions and spectral–spatial attention mechanisms beyond the baseline Pix2Pix model deserves investigation in future endeavors.

Author Contributions: Conceptualization, Amel Oubara and Falin Wu; methodology, Amel Oubara and Reza Maleki; software, Boyi Ma; validation, Amel Oubara, Abdenour Amamra and Gongliu Yang; formal analysis, Amel Oubara; investigation, Amel Oubara; resources, Falin Wu; data curation, Boyi Ma; writing—original draft preparation, Amel Oubara; writing—review and editing, Falin Wu and Boyi Ma; visualization, Reza Maleki and Abdenour Amamra; supervision, Gongliu Yang; project administration, Falin Wu; funding acquisition, Falin Wu. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The two datasets used in this study can be found at https://study.rsgis.whu.edu.cn/pages/download/building_dataset.html and https://chenhao.in/LEVIR/ for WHU-CD and LEVIR-CD, respectively. Accessed on 6 March 2023.

Acknowledgments: The authors would like to thank Saddam Sarwar for improving the English writing of this paper and the researchers who kindly shared their codes and datasets, particularly the SPADE work [37] (code can be found at https://github.com/nvlabs/spade/ accessed on 3 January 2023) and the open change detection remote sensing datasets: WHU-CD [10] and LIVIR-CD [9].

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

Abbreviations

The following abbreviations are used in this manuscript:

Adv-CDNet	Adversarial Change Detection Network
c-GAN	conditional Generative Adversarial Network
CD	Change Detection
CM	Change Map
CNN	Convolutional Neural Network
DL	Deep Learning
GAN	Generative Adversarial Network
LEVIR-CD	LEVIR Building Change Detection Dataset
RS	Remote Sensing
RSCD	Remote Sensing Change Detection
VHR	Very High-Resolution
WHU-CD	WHU Building Change Detection Dataset

References

1. Singh, A. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]
2. Borana, S.; Yadav, S. Urban land-use susceptibility and sustainability—Case study. In *Water, Land, and Forest Susceptibility and Sustainability, Volume 2*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 261–286. [CrossRef]
3. Oubara, A.; Wu, F.; Amamra, A.; Yang, G. Survey on Remote Sensing Data Augmentation: Advances, Challenges, and Future Perspectives. In *Advances in Computing Systems and Applications, Proceedings of the 5th Conference on Computing Systems and Applications, Algiers, Algeria, 17–18 May 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 95–104. [CrossRef]
4. 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]
5. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 1194–1206. [CrossRef]
6. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
7. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134. [CrossRef]
8. Niu, X.; Gong, M.; Zhan, T.; Yang, Y. A conditional adversarial network for change detection in heterogeneous images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 45–49. [CrossRef]
9. 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]
10. 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]
11. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
12. Lv, N.; Ma, H.; Chen, C.; Pei, Q.; Zhou, Y.; Xiao, F.; Li, J. Remote sensing data augmentation through adversarial training. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9318–9333. [CrossRef]
13. Xiao, Q.; Liu, B.; Li, Z.; Ni, W.; Yang, Z.; Li, L. Progressive data augmentation method for remote sensing ship image classification based on imaging simulation system and neural style transfer. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9176–9186. [CrossRef]
14. Singh, A.; Bruzzone, L. SIGAN: Spectral Index Generative Adversarial Network for Data Augmentation in Multispectral Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6003305. [CrossRef]
15. Liu, W.; Luo, B.; Liu, J. Synthetic Data Augmentation Using Multiscale Attention CycleGAN for Aircraft Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4009205. [CrossRef]
16. Xu, X.; Zhao, B.; Tong, X.; Xie, H.; Feng, Y.; Wang, C.; Xiao, C.; Ke, X.; Du, J. A Data Augmentation Strategy Combining a Modified pix2pix Model and the Copy-Paste Operator for Solid Waste Detection With Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8484–8491. [CrossRef]
17. Seo, M.; Lee, H.; Jeon, Y.; Seo, J. Self-Pair: Synthesizing Changes from Single Source for Object Change Detection in Remote Sensing Imagery. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 6374–6383. [CrossRef]
18. Chen, H.; Li, W.; Shi, Z. Adversarial Instance Augmentation for Building Change Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5603216. [CrossRef]
19. Li, Y.; Chen, H.; Dong, S.; Zhuang, Y.; Li, L. Multi-Temporal SamplePair Generation for Building Change Detection Promotion in Optical Remote Sensing Domain Based on Generative Adversarial Network. *Remote Sens.* **2023**, *15*, 2470. [CrossRef]

20. Feng, Y.; Jiang, J.; Xu, H.; Zheng, J. Change detection on remote sensing images using dual-branch multilevel intertemporal network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4401015. [CrossRef]
21. Li, Z.; Tang, C.; Liu, X.; Zhang, W.; Dou, J.; Wang, L.; Zomaya, A.Y. Lightweight Remote Sensing Change Detection With Progressive Feature Aggregation and Supervised Attention. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5602812. [CrossRef]
22. Samadi, F.; Akbarizadeh, G.; Kaabi, H. Change detection in SAR images using deep belief network: A new training approach based on morphological images. *IET Image Process.* **2019**, *13*, 2255–2264. [CrossRef]
23. Ye, Y.; Wang, M.; Zhou, L.; Lei, G.; Fan, J.; Qin, Y. Adjacent-Level Feature Cross-Fusion With 3-D CNN for Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5618214. [CrossRef]
24. He, C.; Zhao, Y.; Dong, J.; Xiang, Y. Use of GAN to Help Networks to Detect Urban Change Accurately. *Remote Sens.* **2022**, *14*, 5448. [CrossRef]
25. Zhang, L.; Yang, F.; Zhang, Y.D.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the 2016 IEEE international conference on image processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712. [CrossRef]
26. Han, C.; Wu, C.; Guo, H.; Hu, M.; Chen, H. HANet: A hierarchical attention network for change detection with bi-temporal very-high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3867–3878. [CrossRef]
27. Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5604816. [CrossRef]
28. Wang, Z.; Zhang, Y.; Luo, L.; Wang, N. CSA-CDGAN: Channel self-attention-based generative adversarial network for change detection of remote sensing images. *Neural Comput. Appl.* **2022**, *34*, 21999–22013. [CrossRef]
29. 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]
30. Zhang, J.; Pan, B.; Zhang, Y.; Liu, Z.; Zheng, X. Building Change Detection in Remote Sensing Images Based on Dual Multi-Scale Attention. *Remote Sens.* **2022**, *14*, 5405. [CrossRef]
31. Ren, H.; Xia, M.; Weng, L.; Hu, K.; Lin, H. Dual-Attention-Guided Multiscale Feature Aggregation Network for Remote Sensing Image Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4899–4916. [CrossRef]
32. Cao, Z.; Wu, M.; Yan, R.; Zhang, F.; Wan, X. Detection of small changed regions in remote sensing imagery using convolutional neural network. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Beijing, China, 18–20 November 2020; Volume 502, p. 012017. [CrossRef]
33. 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]
34. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the 2016 International Conference on Learning Representations (ICLR 2016), San Juan, Puerto Rico, 2–4 May 2016. [CrossRef]
35. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning (ICML'17), Sydney, NSW, Australia, 6–11 August 2017; pp. 214–223.
36. Gauthier, J. Conditional Generative Adversarial Nets for Convolutional Face Generation. 2014. Available online: https://www.semanticscholar.org/paper/Conditional-generative-adversarial-nets-for-face-Gauthier/5cd47e5d004d75fe773a252bde35b56d5d56ce06 (accessed on 6 March 2024).
37. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2337–2346. [CrossRef]
38. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807. [CrossRef]
39. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. In Proceedings of the Sixth International Conference on Learning Representations (ICLR 2018), Vancouver, BC, Canada, 30 April–3 May 2018. [CrossRef]
40. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]
41. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542. [CrossRef]
42. Guo, Q.; Zhang, J.; Zhu, S.; Zhong, C.; Zhang, Y. Deep multiscale Siamese network with parallel convolutional structure and self-attention for change detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5406512. [CrossRef]
43. Daudt, R.C.; 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]
44. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [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.