

Article

Dual-Branch Seasonal Error Elimination Change Detection Framework Using Target Image Feature Fusion Generator

Hongming Zhu , Jipeng Zhang, Zeju Wang, Xinyu Liu, Qin Liu and Bowen Du * 

School of Computer Science and Technology, Tongji University, Shanghai 201804, China; zhu_hongming@tongji.edu.cn (H.Z.); 2333081@tongji.edu.cn (J.Z.); 2231550@tongji.edu.cn (Z.W.); lxy146146@gmail.com (X.L.); qin.liu@tongji.edu.cn (Q.L.)

* Correspondence: bowendu@tongji.edu.cn

Abstract: In change detection tasks, seasonal variations in spectral characteristics and surface cover can negatively impact performance when comparing image pairs from different seasons. Many existing change detection methods do not specifically address the performance degradation caused by seasonal errors. To tackle this issue, the Dual-Branch Seasonal Error Elimination Change Detection Framework using Target Image Feature Fusion Generator (DBSEE-CDF) is introduced. Specifically, the approach utilizes the Target Image Feature Fusion Generator (TIFFG), which incorporates spatial and channel attention mechanisms to extract features from target images and integrates them with deep features from input images using cross-attention. To avoid a severe loss of visual fidelity caused by the significant differences in texture and color features between snow-covered and snow-free images, as well as the different requirements for style transformation, different generators for snow-covered winter images and snow-free winter images are employed to produce intermediate images that eliminate seasonal errors before conducting change detection tasks. The experimental results demonstrate significant improvements in both quantitative and qualitative assessments of change detection tasks compared to directly performing change detection with various models, highlighting the effectiveness of the proposed DBSEE-CDF.

Keywords: change detection; seasonal error; computer vision



Academic Editor: Tania Stathaki

Received: 15 December 2024

Revised: 22 January 2025

Accepted: 30 January 2025

Published: 3 February 2025

Citation: Zhu, H.; Zhang, J.; Wang, Z.; Liu, X.; Liu, Q.; Du, B. Dual-Branch Seasonal Error Elimination Change Detection Framework Using Target Image Feature Fusion Generator. *Remote Sens.* **2025**, *17*, 523. <https://doi.org/10.3390/rs17030523>

Copyright: © 2025 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

Remote sensing images are images that capture the Earth's surface information, collected through distant sensors mounted on platforms such as aircraft, satellites, or unmanned aerial vehicles (UAVs) [1]. These images offer a comprehensive view of the Earth's surface, enabling the extraction of valuable geographic and environmental data. These data include, but are not limited to, land cover types, land use patterns, topography, atmospheric conditions, and oceanic conditions. They provide diverse geographic information. These images play an important role in environmental monitoring, agriculture, urban planning, resource management, etc. [2]. The change detection of remote sensing images refers to the process of detecting changes in the same geographical area by analyzing a set of images captured at different time periods [3]. This task plays a crucial role in various applications, as it helps to track the evolution of landscapes, environmental conditions, and human activities over time. The current change detection tasks can be categorized into two main types: pixel-level change detection and semantic-level change detection. This paper primarily focuses on pixel-level change detection, which involves identifying changes at the individual pixel level, thereby providing a detailed, fine-grained view of the

differences between images [4]. Pixel-level change detection is particularly important in a variety of fields, offering critical insights into both natural and human-induced changes in the environment [5].

However, the current form of remote sensing image change detection still suffers from significant errors, which may originate from various factors, including different cloud cover conditions, lighting conditions, object occlusion and shadows, as well as noise caused by atmospheric and terrain changes at different time points in remote sensing images [6]. Among these errors, remote sensing images from different seasons may incur larger errors in change detection tasks due to seasonal variations in spectral characteristics, land cover, and vegetation. For example, when faced with images covered by a large amount of snow, a change detection model may be interfered with by the snow-covered areas, mistakenly identifying the change from snow-free to snow-covered land as a real change. However, since the type and usage of the land in this area have not changed, the change detection model is expected to exclude such misjudgments caused by seasonal changes. In change detection tasks, these seasonal errors can lead to misjudgments of seasonal changes and misidentifications of genuine changes by change detection models, significantly impairing the performance of change detection tasks.

However, most existing change detection methods have not been optimized specifically for seasonal errors [7–19]. Existing methods for eliminating seasonal errors are mostly based on image-to-image transformation techniques [20–27]. However, these methods lack guidance from the target image visual fidelity information (visual fidelity refers to the authenticity of the synthesized image in both spatial and spectral dimensions, including the realistic representation of ground features such as color, texture, and semantics [28]), which may lead to the incorrect transformation of the original image during the conversion process. That is, in an attempt to change a region to resemble the corresponding season, the type, usage, texture, and color of the land may undergo incorrect changes. For example, in snow-covered winter images, roads are typically black due to vehicles passing over the snow. During seasonal error elimination, this may be mistakenly transformed into an asphalt road or a rural dirt road, which could result in a deviation from the original appearance of the road. This phenomenon can seriously interfere with change detection tasks. Additionally, it was found in the experiment that due to the significant differences in surface coverage, color, texture, etc., between snow-covered winter images and snow-free winter images, using a single transformer for seasonal error elimination can lead to severe loss of visual fidelity. For example, when using CycleGAN (Cycle-Consistent Generative Adversarial Networks) to eliminate seasonal errors of snow-covered winter images and snow-free winter images, the overall lighter color tone of snow-covered winter images causes an issue. After training the generator with a dataset that contains snow-covered winter images, using the same generator to transform snow-free winter images results in darker overall color tones in the generated images, with blurred ground feature details. This leads to a decline in the accuracy of subsequent change detection tasks.

To address the performance degradation in change detection caused by seasonal errors in remote sensing images, a method aimed at eliminating seasonal errors in remote sensing images is proposed. For remote sensing images of the same area at different times, the method first classifies them and uses two generators trained on two different datasets to perform seasonal error elimination on snow-covered winter images and snow-free winter images. In each generator, a hybrid attention module is used, integrating channel attention and spatial attention to extract features from images not in winter. Then, it performs multiple down-sampling processes on the winter images to obtain their deep features, and uses a cross-attention mechanism to fuse these features with the features of images not in winter. After several up-sampling steps, it obtains the final generated

images, forming a new structure: Target Image Feature Fusion Generator (TIFFG). The TIFFG is embedded into the CycleGAN network as a generator and two TIFFGs are trained on two different datasets, thereby obtaining two TIFFGs for transforming different land cover images. Compared to traditional change detection models, this method optimizes for the performance degradation in change detection caused by seasonal errors in remote sensing images taken in winter. Compared to existing seasonal error elimination methods, the method incorporates features of images not taken in winter into TIFFG, reducing the incorrect transformation of the original image in the process of seasonal error elimination by other image-to-image transformation models. Our code is available at <https://github.com/zjp-zjp-zjp/TIFFG> (accessed on 26 November 2024).

In addition, it should be clarified that the seasonal error elimination method proposed in this paper is aimed at eliminating errors between winter and non-winter images. Specifically, it uses TIFFG to convert the texture and color of certain regions in the winter image to match the corresponding regions in the non-winter image, thereby eliminating seasonal errors. The use of a dual-branch structure to separately process snow-covered and snow-free image pairs is primarily to improve the accuracy of the model's transformation, as experiments show that snow cover has a significant impact on the spectral characteristics of the image and the seasonal error elimination task.

In summary, our innovations and contributions are as follows:

- A hybrid attention module formed by combining spatial attention and channel attention is used to extract information from remote sensing images and fuse it with the deep features of the original image obtained through down-sampling, and perform up-sampling to obtain the final generated image. This approach combines information from remote sensing images not in winter to reduce seasonal errors between winter and non-winter remote sensing images and prevents the incorrect transformation of the original image, thus improving change detection performance.
- The Dual-Branch Seasonal Error Elimination Change Detection Framework using Target Image Feature Fusion Generator is proposed with respect to snow-covered winter images and snow-free winter images with substantial differences in surface coverage, color, and texture, which performs seasonal error elimination separately for snow-covered data pairs in winter and snow-free data pairs in winter to solve the problem of a severe loss of visual fidelity.
- Extensive experiments demonstrate that the proposed model improves the performance of change detection tasks compared to methods without seasonal error elimination, achieving an average increase of 7% in the F1-score for change detection tasks. Additionally, it outperforms other image-to-image methods used for seasonal error elimination.

The remainder of this paper is organized as follows. Section 2 introduces the mainstream methods currently used for seasonal error elimination and change detection. Section 3 first describes the overall process of the proposed seasonal error elimination method and provides the corresponding pseudocode. It then introduces the technical approaches of the dual-branch structure, the proposed TIFFG, the CycleGAN network used to train TIFFG, the seasonal classification model, and the change detection model used. Section 4 discusses the methods for acquiring training datasets for each module in the model, model training methods, experimental setup, as well as the results and visualizations of comparative experiments and ablation studies. Section 5 analyzes the model's performance based on the comparative experiments and ablation studies, the necessity of key components in the model, and the model's benefits and shortcomings. Section 6 summarizes the findings, the overall approach of the proposed method, its advantages and disadvantages, and directions for future work.

2. Related Works

2.1. Change Detection for Remote Sensing

Traditional change detection methods typically rely on mathematical transformations or statistical analysis techniques. Most current change detection models are based on deep learning and neural network methods, including Unet, attention mechanism and other feature extraction methods. In addition, there are many valuable methods in other areas of remote sensing image processing (such as anomaly detection). Due to the high similarity of the data and tasks they focus on with change detection, the methods they employ also offer valuable insights for change detection.

2.1.1. Traditional Methods

In change detection research, traditional methods typically rely on mathematical transformations or statistical analysis techniques to extract spectral features from multitemporal remote sensing data in order to identify change areas in the image.

Bovolo et al. [7] addressed unsupervised change detection by proposing an appropriate framework for the formal definition and theoretical study of the Change Vector Analysis (CVA) technique. This framework, based on the representation of CVA in polar coordinates, introduces a set of formal definitions in the polar domain to better describe the information in spectral change vectors, and theoretically analyzes the distributions of changed and unchanged pixels in the polar domain. Wu et al. [8] proposed a novel Slow Feature Analysis (SFA) algorithm for change detection. SFA extracts the most temporally invariant components from multitemporal images and transforms the data into a new feature space. In this feature space, the differences in unchanged pixels are suppressed, allowing for the better separation of changed pixels. Celik [9] proposed an unsupervised change detection technique for multitemporal satellite images using Principal Component Analysis (PCA) and k-means clustering. The difference image is partitioned into non-overlapping $h \times h$ blocks. PCA is applied to the $h \times h$ non-overlapping block set to extract S orthonormal eigenvectors, creating the eigenvector space. Change detection is achieved by using k-means clustering to partition the feature vector space into two clusters ($k = 2$), and then each pixel is assigned to one of the two clusters based on the minimum Euclidean distance between the pixel's feature vector and the mean feature vector of the clusters.

These methods typically assume that the background image remains stable across different time points, with changes occurring only in the target area. However, in cases of significant seasonal variation, the background image may also undergo substantial changes in texture and color. As a result, these methods are prone to misinterpreting seasonal variations as changes, leading to incorrect change detection results.

2.1.2. Unet-Based Methods

In change detection (CD) tasks, the U-Net architecture is widely used. It connects the shallow features in the encoder with the corresponding layers in the decoder through skip connections, which helps preserve the details and spatial information of the image. This is particularly important in change detection tasks, as changes are often manifested at the pixel level. The skip connections enable the model to accurately capture local detailed changes.

Codegoni et al. [13] proposed TinyCD. Due to industrial needs, the design goal of this model is to be faster and more compact than current state-of-the-art change detection models. TinyCD adopts a Siamese U-Net architecture, utilizing low-level features in a globally temporal and locally spatial manner. Additionally, it employs a novel strategy to mix features in the space-time domain, both to merge embeddings obtained from the Siamese backbones and, combined with an MLP module, to form a novel space-semantic attention mechanism called the Mix and Attention Mask Block (MAMB). Fang et al. [19]

proposed Siamese Nested Unet (SNUNet-CD). SNUNet-CD alleviates the loss of localization information in the deep layers of the neural network through compact information transmission between the encoder and decoder, and between decoders. Additionally, they propose an Ensemble Channel Attention Module (ECAM) for deep supervision. Through ECAM, the most representative features from different semantic levels can be refined and used for the final classification.

These methods primarily perform change detection by learning the spatial features (such as pixel distribution, texture, shape, and color) of images. However, seasonal errors may cause significant differences in the spatial features of the same area at different time points. As a result, these methods may misinterpret seasonal variations as changes in the target area, leading to incorrect change detection results.

2.1.3. Attention Mechanism-Based Methods

With the widespread application of Transformer models, they have become the dominant method in the field of feature extraction, and many change detection models are also incorporating their attention mechanisms. The attention mechanism allows the model to automatically learn and focus on important regions or features in the input image. In change detection, changes are often concentrated in certain local areas of the image (e.g., objects, edges, etc.). By weighting the features of different regions, the attention mechanism can effectively enhance the focus on these areas, thereby improving the model's ability to recognize subtle changes.

Bandara et al. [15] proposed Change Former, utilizing the transformer architecture to effectively capture both global and local context in images; Change Former employs self-attention mechanisms to better identify subtle changes. This makes it particularly well-suited for tasks such as environmental monitoring or time-series image analysis. Han et al. [16] also proposed Hierarchical Attention Network (HANet-CD). By incorporating different types of attention modules, such as channel attention and spatial attention, HANet better captures key information and context in images, thereby improving segmentation accuracy and change detection capability. Chen et al. [18] proposed a Bitemporal Image Transformer (BIT) to efficiently and effectively model contexts within the spatial-temporal domain. They represent the bitemporal image as a few tokens and use a transformer encoder to model contexts in the compact token-based space-time. The learned context-rich tokens are then fed back into the pixel space to refine the original features via a transformer decoder. They incorporate BIT into a deep feature differencing-based CD framework.

Attention mechanisms can help models to focus on key areas within images, but in the presence of seasonal errors, models may treat seasonal variations as significant features. Since seasonal changes impact background image colors and textures, models may incorrectly perceive these changes as actual target variations, leading to inaccurate change detection results.

2.1.4. Feature Extraction-Based Methods

Xing et al. [11] proposed the Lightweight change detection network (LightCDNet). This is a novel lightweight change detection model that effectively preserves input information. LightCDNet consists of an early fusion backbone network and a pyramid decoder for end-to-end change detection. The core component of LightCDNet is the Deep Supervised Fusion Module (DSFM), which guides the early fusion of primary features to improve performance. Fang et al. [12] proposed Changer. They propose a novel general change detection architecture, MetaChanger, which includes a series of alternative interaction layers in the feature extractor. To better align bi-temporal features, they introduce a Flow Dual-Alignment Fusion (FDAF) module that allows interactive alignment and feature fu-

sion. Han et al. [14] proposed the Change Guiding Network (CGNet-CD), which integrates both global and local context information from images. By considering both local details and the global background of the image, CGNet more accurately identifies change regions within the image. Zheng et al. [17] proposed ChangeStar. Initially, they proposed a new change detection method—Single-Temporal Supervised Learning (STAR)—which takes a novel approach by using object changes in unpaired images as supervisory signals. STAR allows the training of a high-accuracy change detector using only unpaired labeled images and can be generalized to real-world bitemporal images. To evaluate the effectiveness of STAR, they designed a simple yet effective change detector called ChangeStar, which can reuse any deep semantic segmentation architecture through the ChangeMixin module.

However, the primary method these change detection models use to address seasonal errors still focus on improving the model's generalization, rather than specifically eliminating seasonal errors.

2.1.5. Methods from Anomaly Detection Fields

Hyperspectral image anomaly detection (Anomaly Detection, AD) is a technique used to automatically identify anomalous targets in remote sensing images that differ from the background. Since anomaly detection, like change detection, deals with remote sensing images and involves identifying regions in the image that differ from the regular patterns, the latest methods used in this field provide valuable reference for change detection research.

Wang et al. [29] proposed a self-supervised blind-block network (BockNet), which introduces a guard window. BockNet creates a blind-block (guard window) at the center of the network's receptive field, preventing it from seeing the information inside the guard window when reconstructing the central pixel. This process seamlessly integrates a sliding dual-window model into BockNet, where the inner window is the guard window and the outer window is the receptive field outside the guard window. Cheng et al. [30] proposed a novel deep feature aggregation network (DFAN) and developed a new paradigm for HAD to represent multiple background patterns. The DFAN adopts an adaptive aggregation model (AAM), which combines the orthogonal spectral attention module (OSAM) with the background anomaly category statistics module. This enables the effective utilization of both spectral and spatial information to capture the distribution of the background and anomalies. To better optimize the proposed DFAN, a novel multiple aggregation separation loss (MASL) is designed, which is based on the intrasimilarity and interdifference between the background and anomalies. Wang et al. [31] proposed a hierarchical integration framework for hyperspectral image anomaly detection and denoising (HyADD). The joint anomaly detection and denoising processes are integrated and mutually stimulate each other in each iteration. Inspired by spatial-spectral gradient domain constraints, HyADD provides effective feedback to the denoising process by utilizing the conduction of an anti-noise dictionary and the low-rankness in the subspace domain, with features differing from the background in anomaly detection.

2.2. Seasonal Error Removal

The existing seasonal error correction methods are mainly based on image-to-image translation techniques in image generation, with structures including Generative Adversarial Networks (GANs), Diffusion Models, and Variational Autoencoders (VAEs); there are also some unique algorithms based on other graph theory principles, such as the shortest path assumption.

2.2.1. Generative Adversarial Network-Based Methods

The method of using GAN for image-to-image image conversion is based on the principle of training a generator network to convert the input source image into an output image

similar to the target image domain while training a discriminator network to distinguish between real target images and images generated by the generator. Through adversarial training between the two, the generation ability of the generator is continuously optimized, and high-quality image conversion is finally achieved.

Zhu et al. [20] proposed CycleGAN (Cycle-Consistent Generative Adversarial Networks), which learns a mapping, $G: X \rightarrow Y$, without the need for paired examples, such that the distribution of images generated from $G(X)$ cannot be distinguished from the distribution of Y using adversarial loss. Since this mapping is highly unconstrained, it is coupled with an inverse mapping $F: Y \rightarrow X$, and a cycle-consistency loss is introduced to enforce $F(G(X)) \approx X$. Ko et al. [23] proposed SuperstarGAN, an improved version of StarGAN [32]. They adopted the idea first introduced in Controllable GAN (ControlGAN) [33] by using data augmentation techniques to train an independent classifier, addressing the overfitting problem in the classification of StarGAN's structure. Since the generator, combined with a well-trained classifier, can express small features belonging to the target domain, SuperstarGAN enables image-to-image translation across large-scale domains.

However, existing GAN-based methods do not incorporate visual fidelity information from the target image and only perform image style conversion based on the original image, which leads to misjudgment and loss of visual fidelity information during the conversion process.

2.2.2. Variational Autoencoders-Based Methods

The principle of using Variational Autoencoders (VAEs) for image-to-image image conversion lies in training an encoder–decoder architecture that maps input images to a latent space and then reconstructs them into output images similar to the target image domain, with the latent space allowing for the generation of diverse outputs while maintaining a regularized distribution, enabling high-quality and varied image transformations.

Zhang et al. [27] proposed VAE-CoGAN. They proposed a method for unpaired image-to-image translation based on the coupled generative adversarial network CoGAN, which combines a variational autoencoder (VAE) with a generative adversarial network (GAN) under a conditioned generative process. Unlike the basic CoGAN, they proposed a shared latent space and a variational autoencoder VAE in the framework. Park et al. [21] proposed the swapping autoencoder, which encodes images using two independent components. It enforces that any combination of swaps maps to a realistic image and forces one component to encode the co-occurrence statistics of patches appearing in different parts of the image.

VAE learns the latent space representation of an image to generate a reconstructed version of the image. However, in the presence of seasonal errors, VAE may confuse seasonal changes with real changes and reconstruct real changes as part of the seasonal changes. This could lead to the loss of real change information.

2.2.3. Diffusion Model-Based Methods

The principle of the method of image-to-image image conversion using a diffusion model is to gradually convert the input image into a noisy image through a forward process, and then gradually denoise and guide the generation of an output image similar to the target image domain based on learned noise distribution characteristics through a reverse process. This process utilizes the randomness of Markov chains and optimizes model parameters to generate images that not only conform to the characteristics of the target image but also maintain diversity.

Tumanyan et al. [22] proposed pnp-diffusion (Plug-and-play diffusion), which leverages the combination of noise images (i.e., low-quality images generated by the diffusion

model) and the target text prompt to control and regulate the image generation process. This approach relies on a conditional diffusion process, where the model learns to gradually guide the noise image towards an image that aligns with the target prompt during the diffusion process. Parmar et al. [25] proposed pix2pix-zero, a method for image-to-image translation that preserves the original image's content without manual prompting. They first automatically discover editing directions in the text embedding space that reflect the desired edits. To retain the content structure, they propose cross-attention guidance, which aims to preserve the cross-attention maps of the input image during the diffusion process. Finally, to enable interactive editing, they distill the diffusion model into a fast conditional generative adversarial network (GAN). Zhao et al. [24] proposed EGSDE (Energy-Guided Stochastic Differential Equations), which employs an energy function pretrained on both the source and target domains to guide the inference process of a pretrained SDE for realistic and faithful unpaired I2I. Building upon two feature extractors, they carefully design the energy function such that it encourages the transferred image to preserve the domain-independent features and discard domain-specific ones. Further, they provide an alternative explanation of the EGSDE as a product of experts, where each of the three experts (corresponding to the SDE and two feature extractors) solely contributes to faithfulness or realism.

The image generation process of diffusion models typically focuses on the global features of the image, and may not be able to distinguish the local texture and color differences caused by seasonal changes with sufficient precision, thus affecting the effectiveness of seasonal error removal.

2.2.4. Graph Theory-Based Methods

In addition, there are also some models that perform image-to-image translation based on graph theory principles.

Xie et al. [26] proposed Santa. They focus on the paths connecting images in the first domain to images in the second domain and propose the shortest path assumption to address the ill-posed joint distribution learning problem. They first introduce a shared generation assumption to synthesize images along the path. To find an appropriate mapping, they aim to minimize the path length, which is formulated as the expectation of the norm of the Jacobian matrix with respect to the domain variable. To reduce the computational cost of the high-dimensional Jacobian matrix, they further propose using the finite difference method to approximate the Jacobian matrix and penalize its squared norm.

Graph Theory-based methods optimize images by modeling the relationships between nodes, which can sometimes lead to excessive smoothing, making it difficult to preserve the real change information in the image and thus affecting the precise elimination of seasonal errors.

3. Methods

In order to maintain the accuracy of land cover types and uses while aligning the texture and color of certain areas in the winter image as closely as possible with the corresponding areas in the non-winter image, the TIFFG is proposed, which captures the land cover and use information from certain areas of the winter image, as well as the texture and color information from the corresponding areas in the non-winter image, and integrates both into the transformation process to prevent incorrect transformations of the original image.

To avoid the severe loss of visual fidelity caused by the significant differences in texture and color features between snow-covered and snow-free images, as well as the different requirements for style transformation, the proposed method first classifies pairs of

remote sensing images and processes and transforms snow-covered and snow-free winter images through a dual-branch structure with different TIFFGs. This allows the seasonal error elimination task to use more suitable transformers for style transformation when dealing with images of different land cover, thereby improving the transformation effect.

In the proposed method, for a pair of cross-season remote sensing images, it first classifies them into winter and non-winter images. The winter images are further divided into snow-covered and snow-free winter images. Then, using a dual-branch structure, depending on the snow coverage of the winter image in the image pair, different TIFFGs are applied to transform the textures and colors of certain regions in the winter image to match the corresponding regions in the non-winter image, thereby eliminating seasonal errors. Finally, the transformed image is input together with the non-winter image from the image pair into a change detection model to complete the final change detection task. Algorithm 1 is the pseudocode for the overall process of the method proposed.

Algorithm 1: Dual-Branch Seasonal Error Elimination Change Detection Framework using Target Image Feature Fusion Generator

Data: A pair of cross-season remote sensing images X and Y
Result: Change map Out between X and Y

- 1 Train a ResNet network *Classifier* to classify seasonal and snow cover conditions in remote sensing images.;
- 2 Train two generators $TIFFG_{snow-covered}$ and $TIFFG_{snow-free}$ to eliminate seasonal errors.;
- 3 Train a change detection model CD to obtain the change map between image pairs.;
- 4 Obtain the seasonal classification result of X : $R1 = Classifier(X)$.;
- 5 Obtain the seasonal classification result of Y : $R2 = Classifier(Y)$.;
- 6 **if** $R1 == notwinter$ **then**
 - 7 **if** $R2 == snow - covered$ **then**
 - 8 $R2' = TIFFG_{snow-covered}(R2)$
 - 9 **else**
 - 10 $R2' = TIFFG_{snow-free}(R2)$
 - 11 **end**
 - 12 $Out = CD(R1, R2')$;
 - 13 **return** Out
- 14 **else**
 - 15 **if** $R1 == snow - covered$ **then**
 - 16 $R1' = TIFFG_{snow-covered}(R1)$
 - 17 **else**
 - 18 $R1' = TIFFG_{snow-free}(R1)$
 - 19 **end**
 - 20 $Out = CD(R1', R2)$;
 - 21 **return** Out
- 22 **end**

3.1. Dual-Branch Seasonal Error Elimination Change Detection Framework Using Target Image Feature Fusion Generator

Through experiments, it was found that snow coverage has a significant impact on the seasonal error elimination task. Specifically, in the transformation from snow-covered winter images to non-winter images, the generator needs to address the strong reflection properties of snow and eliminate its texture. The reflection and color of snow are typically

much more complex than the other elements in snow-free winter images (such as bare soil, dead grass, etc.), meaning the generator must adapt to complex texture mapping and lighting changes. On the other hand, the transformation from snow-free winter images to non-winter images is relatively simple because the background already resembles the non-winter scene. The generator only needs to handle seasonal changes (such as color variation and vegetation growth).

Using a dual-branch generator, with two separate TIFFGs trained on snow-covered images and their corresponding non-winter images, as well as snow-free images and their non-winter corresponding non-winter images, can effectively meet the style transformation requirements for both and solve the issue of significant surface coverage, color, and texture differences between snow-covered and snow-free images. In practical experiments, it was found that using a single generator for transformation led to overly dark generated images in the seasonal error elimination from snow-free images to non-winter images. This could be due to the stronger reflection intensity of snow-covered images, which results in higher overall image brightness. As a result, the trained model would also severely reduce the image brightness when transforming snow-free images.

For this reason, DBSEE-CDF is proposed. This network handles change detection in remote sensing images for both snow-covered and snow-free scenes separately. The structure is illustrated in Figure 1.

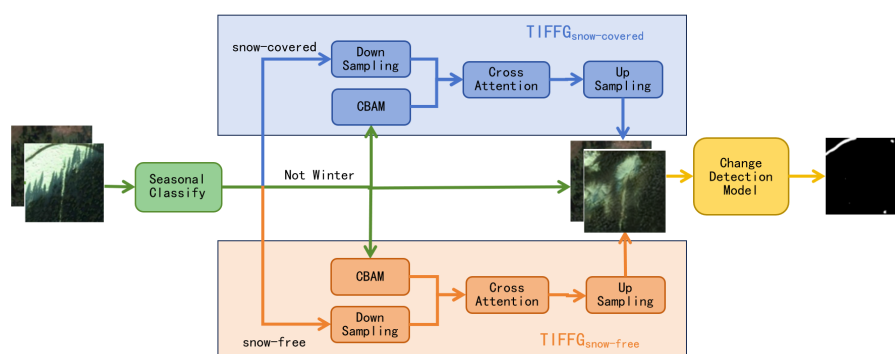


Figure 1. The structure of the proposed DBSEE-CDF. First, it classifies the winter images into snow-covered and snow-free categories. It then uses the TIFFG trained on the corresponding dataset to transform these images, generating intermediate images. These intermediate images are then used for change detection tasks with images not in winter.

Specifically, when a pair of remote sensing images is input into the network, they undergo a structured processing flow to classify and transform the images based on their seasonal characteristics.

First, both images in the pair are classified using a seasonal classification model, with ResNet50 [34] serving as the architecture for the model. Since ResNet itself has a sufficiently high image classification accuracy, we did not alter its network structure and instead began fine-tuning from the pre-trained ResNet50 model. This classification model outputs one of three possible categories for each image: snow-covered image in winter, snow-free image in winter, or image not in winter.

Once the classification is performed, the network examines the specific combination of image types in the pair. If the pair consists of one snow-covered image in winter and one image not in winter, the snow-covered image in winter is forwarded to the TIFFG specialized for snow-covered images, denoted as $TIFFG_{snow-covered}$. This generator is responsible for transforming the snow-covered winter image into an image that closely resembles the non-winter conditions, thus eliminating the seasonal error and ensuring consistency across the pair. Similarly, if the pair consists of one snow-free image in winter and one

image not in winter, the snow-free image in winter is passed to the TIFFG specialized for snow-free images, denoted as $TIFFG_{snow-free}$. This generator performs the necessary transformation to convert the snow-free winter image into an image more representative of non-winter conditions.

After the transformation process, the transformed image is then combined with the corresponding image from the non-winter set. This combined pair is then input into the final change detection model, which is designed to analyze and detect changes between the two images.

By separating the processing of snow-covered and snow-free winter images, the network ensures that seasonal errors are effectively mitigated and change detection is performed more accurately in both snow-covered and snow-free scenes. This structured approach allows for a more robust handling of seasonal variations in the dataset and improves the overall performance of the change detection task.

The training processes for the seasonal classification model, the two generators, and the change detection model will be described in Section 4. Particularly, for the training of the change detection model, to adapt it to the images generated by TIFFG, the transformed dataset will be used instead of the original dataset.

3.2. Target Image Feature Fusion Generator

Motivated by hybrid attention module [35], TIFFG is proposed, which integrates spatial attention and channel attention with a UNet-based generator. These components are combined through a cross-attention module, as illustrated in Figure 2.

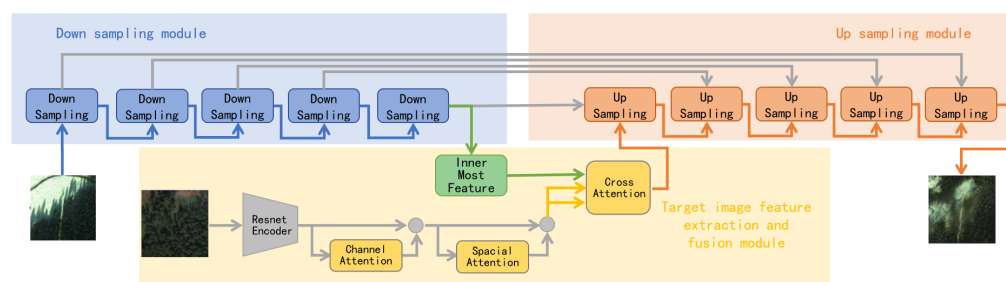


Figure 2. The pipeline of our proposed TIFFG, which consists of a (i) down-sampling module (top left), (ii) a target image feature extraction and fusion module (bottom) and (iii) an up-sampling module (top right). The down-sampling module is responsible for extracting the deep features of the input image to be transformed. The target image feature extraction and fusion module is responsible for extracting features from the target image and integrating them with the deep features of the input image to be transformed. The up-sampling module is responsible for generating intermediate images for change detection based on blended features.

For a set of remote sensing images captured at different times, assuming the winter image is denoted as $Input_A \in R^{B \times C \times H \times W}$ and the non-winter image as $Input_B \in R^{B \times C \times H \times W}$, $Input_A$ undergoes five down-sampling steps to obtain its deep features $F_A^{deep} \in R^{B \times C' \times \frac{H}{32} \times \frac{W}{32}}$. As for $Input_B$, it first goes through a ResNet50 encoder to be encoded, resulting in shallow features $F_B^{shallow} \in R^{B \times C' \times \frac{H}{32} \times \frac{W}{32}}$. Subsequently, these shallow features are fed into a hybrid attention module to generate the deep features of $Input_B$. Specifically, for the shallow features of $Input_B$, they are first processed by a channel attention module to extract channel features $F_B^{channel} \in R^{B \times C' \times 1 \times 1}$, and then the channel features $F_B^{channel}$ and the shallow features $F_B^{shallow}$ are multiplied using element-wise multiplication to obtain intermediate features $F_B^{mid} \in R^{B \times C' \times \frac{H}{32} \times \frac{W}{32}}$. Introducing channel attention in seasonal error elimination can help the model to focus more precisely on channels that are heavily influenced by seasonal changes. This enhancement allows the model to better adapt to

variations in data distribution across different seasons and to capture essential feature patterns more effectively during specific seasons. As a result, it reduces the negative impact of seasonal errors on change detection results. The detailed processes are as follows:

$$F_B^{channel} = \sigma(MLP(AvgPool(F_B^{shallow}))) + MLP(MaxPool(F_B^{shallow})), \quad (1)$$

$$F_B^{mid} = F_B^{channel} \otimes F_B^{shallow}, \quad (2)$$

where σ represents sigmoid function. MLP is multilayer perceptron. $AvgPool$ and $MaxPool$ are the average pooling layer and max pooling layer. Using a combination of max pooling and average pooling in channel attention allows for the integrated utilization of different pooling methods to enhance the model's perception of channel importance. Average pooling, coupled with a sigmoid function, maps the average response of each channel to a normalized weight value, while max pooling directly selects the strongest feature response without needing additional probability transformations; hence, it does not use a sigmoid function.

For the middle features F_B^{mid} , they are first processed by a spatial attention module to extract spatial features $F_B^{spacial} \in R^{B \times 1 \times \frac{H}{32} \times \frac{W}{32}}$, and then the spatial features $F_B^{spacial}$ and the middle features F_B^{mid} are multiplied using element-wise multiplication to obtain deep features $F_B^{deep} \in R^{B \times C' \times \frac{H}{32} \times \frac{W}{32}}$:

$$F_B^{spacial} = \sigma(Conv([AvgPool(F_B^{mid}); MaxPool(F_B^{mid})])), \quad (3)$$

$$F_B^{deep} = F_B^{spacial} \otimes F_B^{mid}. \quad (4)$$

Introducing spatial attention in seasonal error elimination can help the model to more effectively capture and model spatial relationships between different locations in images. It aids the model in focusing on significant spatial positions, such as corners of buildings or specific segments of roads, enhancing its ability to generalize across different geographical locations or shooting conditions. This improvement allows the model to better distinguish between seasonal variations and actual changes. The advantage of a combination of max pooling and average pooling in spatial attention functions is similar to channel attention.

Finally, the deep features of $Input_A$ and the deep features of $Input_B$ are input into the cross-attention module to obtain the final blended features:

$$F_{A'}^{deep} = softmax(\frac{(F_A^{deep} W_q)(F_B^{deep} W_k)^T}{\sqrt{d_k}})(F_B^{deep} W_v). \quad (5)$$

where W_q, W_k, W_v are the query, key, value weight matrices in the cross-attention module. Using cross-attention to fuse deep feature matrices allows each element in one feature matrix to attend to relevant elements in another feature matrix. This enhances the model's capability to integrate complementary information from both matrices, enabling selective focus on important features across different matrices. It promotes the effective communication and alignment of information between the two matrices, leveraging the strengths of each feature matrix while mitigating their weaknesses.

After obtaining the blended features $F_{A'}^{deep}$, they are passed through five up-sampling layers to reconstruct the final generated image.

3.3. TIFFG Training Method

TIFFG is trained using CycleGAN training strategy, as shown in Figure 3. In experiments, it is required for the images generated by TIFFG to closely resemble non-winter images. In such scenarios, employing GANs is a conventional approach. Among these,

by introducing cycle consistency loss and enabling transformation of unpaired data, CycleGAN effectively captures style features from images and prevents the significant loss of essential image content during the conversion process. Given that the image pairs in this study were captured at different times, resulting in considerable variations in image content, the preservation of image content information by CycleGAN holds significant importance for seasonal transformation tasks and change detection tasks.

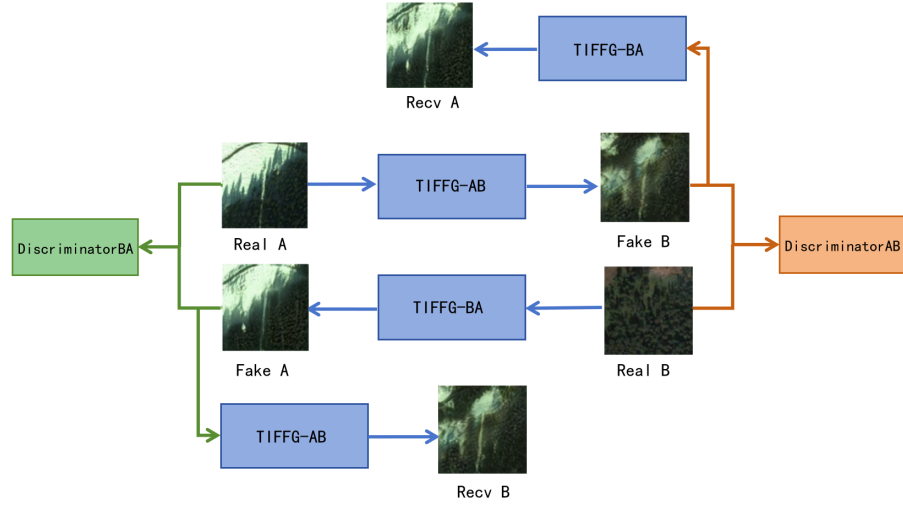


Figure 3. Network for training TIFFG based on CycleGAN.

For a pair of input images a_{real} and b_{real} , in one direction, CycleGAN first utilizes the TIFFG-AB (G_{AB}) to convert a_{real} into b_{fake} , and then uses the TIFFG-BA (G_{BA}) to convert b_{fake} into a_{recv} . In the opposite direction, it performs the same process to obtain a_{fake} and b_{recv} . To generate identity loss, it feeds b_{real} into G_{AB} , resulting in a_{idt} , and feeds a_{real} into G_{BA} , resulting in b_{idt} . Then, the b_{real} and b_{fake} images are fed into the discriminator D_{AB} to determine which one is real and which one is fake. In a Generative Adversarial Network (GAN), the discriminator plays a crucial role by providing a key gradient signal to the generator. This forces the generator to improve its generated images to deceive the discriminator as much as possible, making it unable to accurately distinguish generated images from real ones. Ultimately, this process enables the generator to produce more realistic images. The output of the GAN discriminator can be interpreted as the probability that the input image is real. Next, it calculates the adversarial loss:

$$L_{AB} = \mathbb{E}_{b_{real} \sim p(b_{real})} [\log D_{AB}(b_{real})] + \mathbb{E}_{b_{fake} \sim p(b_{fake})} [\log(1 - D_{AB}(b_{fake}))], \quad (6)$$

$$L_{BA} = \mathbb{E}_{a_{real} \sim p(a_{real})} [\log D_{BA}(a_{real})] + \mathbb{E}_{a_{fake} \sim p(a_{fake})} [\log(1 - D_{BA}(a_{fake}))]. \quad (7)$$

After that, it computes the consistency loss:

$$L_{cycA} = \mathbb{E}_{a_{real} \sim p(a_{real})} [\|a_{recv} - a_{real}\|_1], \quad (8)$$

$$L_{cycB} = \mathbb{E}_{b_{real} \sim p(b_{real})} [\|b_{recv} - b_{real}\|_1]. \quad (9)$$

Then, it is necessary to calculate the identity loss:

$$L_{idtA} = \mathbb{E}_{b_{real} \sim p(b_{real})} [\|a_{idt} - b_{real}\|_1], \quad (10)$$

$$L_{idtB} = \mathbb{E}_{a_{real} \sim p(a_{real})} [\|b_{idt} - a_{real}\|_1]. \quad (11)$$

Identity loss encourages the generator to map the input image back to itself, thereby promoting the preservation of the input image's structure and content. This addresses the issue where generated images might lose some important features of the original input.

Finally, CycleGAN calculates the total loss for the entire network and updates the weights accordingly:

$$L = L_{AB} + L_{BA} + \lambda_a(L_{cycA} + \lambda_{idt}L_{idtB}) + \lambda_b(L_{cycB} + \lambda_{idt}L_{idtA}), \quad (12)$$

where $\lambda_a, \lambda_b, \lambda_{idt}$ are hyperparameters that defines the respective impact levels of L_{cycA} , L_{cycB} , and L_{idtA}, L_{idtB} on training.

4. Experiment

4.1. Overall Training Process

The training process for this framework involves multiple stages, including the training of a seasonal classification model, two TIFFGs, and change detection models. Each stage is designed to systematically address seasonal errors and enhance the performance of change detection tasks.

The first step involves preparing a season-classified image dataset derived from the original high-resolution images. This dataset is labeled to categorize images into three distinct classes: snow-covered winter images, snow-free winter images, and images not in winter. The labeled dataset is then used to train the seasonal classification model, enabling it to accurately identify and classify images based on their seasonal characteristics.

Next, using the seasonal classification model, images from the original dataset classified as snow-covered winter images and their corresponding images not in winter are selected to form one dataset, while images classified as snow-free winter images and their corresponding images not in winter are selected to form another dataset. Each of these datasets is then used to train a separate TIFFG. The purpose of these TIFFGs is to transform winter images—either snow-covered or snow-free—into images representative of non-winter conditions. This targeted transformation process is essential for mitigating the impact of seasonal variations on subsequent change detection tasks.

The trained seasonal classification model and TIFFGs are combined to process the original dataset. By applying these models, seasonal errors are systematically eliminated, resulting in a seasonally corrected dataset where the effects of winter conditions are minimized. This corrected dataset serves as the foundation for training and testing the change detection models, ensuring that they are optimized to focus on genuine changes in the environment without being influenced by seasonal anomalies.

4.2. Dataset

The ChangeDetectionDataset [36] is utilized as an experimental dataset. This dataset consists of 11 pairs of high-resolution remote sensing images, each of varying sizes, and is complemented by a substantial set of 10,000 image pairs for the training set, 2998 image pairs for the validation set, and 3000 image pairs for the test set. These image pairs are generated from the 11 high-resolution remote sensing images and represent different time periods and seasonal conditions, providing a diverse and rich foundation for change detection tasks.

Each image pair in the dataset corresponds to a before-and-after scenario from different seasons or time periods, with the corresponding ground truth clearly depicting the changes that occurred between the two images. Each image in the training set, validation set, and test set is sized at 256×256 pixels.

Given its diverse range of images from different seasons and time points, as well as its detailed ground truth annotations, the ChangeDetectionDataset is highly suitable for training and validating our experimental model. The combination of high-resolution imagery, large-scale data, and seasonal variation makes this dataset an ideal choice for evaluating the effectiveness of the proposed seasonal error elimination and change detection framework.

4.3. Seasonal Classification Dataset Acquisition and Model Training

Due to the mixed seasonal nature of the original dataset, which contains both summer and winter images in sets A and B, it is essential to train a seasonal classification model to properly organize the dataset. The original 11 pairs of high-resolution images are relatively easy to classify in terms of their seasonal characteristics, as the seasonal differences are quite distinct. By re-cropping these 11 pairs of high-resolution images, a large number of individual samples can be generated. Each cropped sample's season is determined by referencing the corresponding high-resolution image.

These cropped samples are then classified according to their seasonal characteristics and stored in the appropriate categories. As a result, the dataset is divided into three distinct groups: 1119 samples of images not in winter, 495 samples of snow-covered images from winter, and 624 samples of snow-free images from winter. This carefully organized dataset is subsequently used to train the seasonal classification model, allowing the model to learn to differentiate between different seasonal conditions and to classify future images accordingly. This seasonal classification model forms the foundation for subsequent processing steps, such as seasonal error elimination and change detection tasks, ensuring that the images are accurately classified based on their seasonal context.

4.4. TIFFG Dataset Acquisition and TIFFG Models Training

Utilizing the seasonal classification model mentioned in Section 4.3, the original dataset is organized and classified as follows: first, all winter images in the original dataset are placed in set A, while non-winter images are placed in set B. Then, the entire dataset is divided to ultimately form three datasets:

- The comprehensive dataset $dataset_{total}$, which contains all images from the original dataset, with a basic guarantee that set A consists entirely of images in winter and set B consists entirely of images not in winter.
- The snow-covered dataset $dataset_{snow-covered}$, which contains a subset of images from $dataset_{total}$, with a basic guarantee that set A consists entirely of snow-covered images, while set B consists entirely of images not in winter.
- The snow-free dataset $dataset_{snow-free}$, which contains a subset of images from $dataset_{total}$, with a basic guarantee that set A consists entirely of snow-free images, while set B consists entirely of images not in winter.

Using $dataset_{snow-covered}$ and $dataset_{snow-free}$ in combination with the CycleGAN, two TIFFGs are trained specifically for transforming the images in set A. These TIFFGs are designed to handle the transformation of winter-related images in set A, with one TIFFG tasked with transforming snow-covered images and the other with transforming snow-free images. These two trained TIFFGs serve as the two generators in the DBSEE-CDF. By leveraging the strengths of these generators, the framework aims to effectively address and eliminate seasonal errors that might otherwise interfere with change detection tasks.

In addition to the primary TIFFGs, for the purpose of ablation experiments, two more TIFFGs are trained without the target image feature extraction and fusion modules. These TIFFGs are trained using $dataset_{snow-covered}$ and $dataset_{snow-free}$, employing the same CycleGAN network architecture. These ablation experiments are crucial for evaluating the

impact of the target image feature extraction and fusion module on the performance of the seasonal error elimination process.

To further benchmark and compare the seasonal error elimination method proposed in this paper with other existing approaches, several well-known image-to-image translation models are trained, including the original CycleGAN, Swapping Autoencoder, SuperstarGAN, EGSDE, and Pnp-diffusion. These models are trained using the $dataset_{total}$ to assess their relative performance in seasonal error elimination and change detection tasks. This comparative analysis helps to validate the effectiveness of the proposed method and highlights any potential advantages it may offer over other state-of-the-art techniques.

4.5. Change Detection Models Training

First, for each sample in $dataset_{total}$, the classification model discussed in Section 4.3 is applied to classify the images, and the corresponding TIFFG is used to perform the transformation. The transformed images are then stored in a new set, A' . Afterward, A' is merged with the original set B and the ground truth to create the modified dataset $dataset'_{total}$.

Next, for each sample in $dataset_{total}$, the same classification model is used to classify the images; however, this time, the TIFFGs that do not include the target image feature extraction and fusion module are employed for transformation. The resulting transformed images are stored in a new set, A'' . This set A'' is then merged with the original set B and the ground truth to form the dataset $dataset''_{total}$.

To compare the seasonal error elimination method proposed in this paper with other existing methods, several alternative models are applied. For each sample in $dataset_{total}$, the original versions of CycleGAN, Swapping Autoencoder, SuperstarGAN, EGSDE, and Pnp-diffusion, as described in Section 4.4, are used for transformation. This results in the creation of sets A^3 , A^4 , A^5 , A^6 , and A^7 . Each of these sets is then merged with the original set B to form the datasets $dataset^3_{total}$, $dataset^4_{total}$, $dataset^5_{total}$, $dataset^6_{total}$, and $dataset^7_{total}$.

Different change detection models are then used, including CGNet-CD [14], Change Former [15], and HANet-CD [16], to train on $dataset_{total}$, $dataset'_{total}$, $dataset''_{total}$, $dataset^3_{total}$, $dataset^4_{total}$, $dataset^5_{total}$, $dataset^6_{total}$, and $dataset^7_{total}$, resulting in corresponding change detection models $cdmodel$, $cdmodel'$, $cdmodel''$, $cdmodel^3$, $cdmodel^4$, $cdmodel^5$, $cdmodel^6$, and $cdmodel^7$.

The performance of $cdmodel$ is evaluated on the three datasets mentioned in Section 4.4 to establish the baseline performance of the change detection model. Next, the two TIFFGs and $cdmodel'$ are integrated into the DBSEE-CDF and tested on the same three datasets to assess the performance improvements achieved by the seasonal error elimination techniques proposed in this work.

For ablation experiments, the target image feature extraction and fusion modules is removed from the two TIFFGs and embed with $cdmodel''$ into the DBSEE-CDF. These ablation tests are then conducted on the same datasets to compare performance with the full version of the framework.

Finally, the original CycleGAN, Swapping Autoencoder, SuperstarGAN, EGSDE, and Pnp-diffusion models, as trained in Section 4.4, are connected to $cdmodel^3$, $cdmodel^4$, $cdmodel^5$, $cdmodel^6$, and $cdmodel^7$, respectively, for a comparative analysis of the seasonal error elimination techniques and their impact on change detection performance. This comprehensive comparison provides insights into the relative effectiveness of the proposed approach in handling seasonal errors compared to existing methods.

4.6. Hyperparameters and Runtime Environment

For the training of the seasonal classification model, we use 50 iterations, the SGD optimizer with a learning rate of 0.001, and momentum of 0.9 to accelerate the training process. The batch size for the data is set to 4.

For the training of the image-to-image seasonal error elimination model, we use 200 iterations, an initial learning rate of 0.0002, the Adam optimizer, a first-order momentum decay rate of 0.5, and a second-order momentum decay rate of 0.999. In the first 100 iterations, the learning rate remains constant, and from iteration 101 onward, the learning rate gradually decays to 0. The batch size for the data is set to 4.

For the training of the change detection model, we use 100 iterations, an initial learning rate of 0.0001, the Adam optimizer, and the learning rate remains constant for the first 50 iterations. From iteration 51 onward, the learning rate gradually decays with a decay factor of 0.1. The batch size for the data is set to 4.

All training is conducted on NVIDIA GeForce RTX 3090 GPUs with 24GB of VRAM.

4.7. Experiment Setup

Extensive experimental validations are conducted to assess the effectiveness of the proposed framework and the necessity of its key components.

To validate the generalizability of the framework proposed (i.e., that using the proposed model for seasonal error elimination can improve the accuracy of change detection tasks regardless of the variation detection method), several change detection models are trained and tested, including CGNet-CD, ChangeFormer, and HANet-CD, using a dataset that underwent seasonal error elimination with the proposed framework. The results are compared with those where no seasonal error elimination is performed.

To demonstrate that the proposed model achieves better performance on the seasonal error elimination task compared to existing image-to-image translation models, and provides a greater improvement in the change detection task, several image-to-image translation models for seasonal error elimination (including CycleGAN, Swapping Autoencoder, SuperstarGAN, Pnp-diffusion, and EGSDE) are applied. Change detection models are then trained and tested on the dataset with seasonal errors eliminated, and the results are compared with those obtained using the seasonal error elimination framework proposed.

To verify the necessity of the target image feature extraction module and the dual-branch structure, and their effect on improving the seasonal error elimination task, we separately removed target image feature extraction module (N_{double}) and the dual-branch structure ($N_{single-TIFFG}$) and performed seasonal error elimination on the dataset. The change detection models are then trained and tested using the dataset with seasonal errors eliminated and the results are then compared with those obtained using the seasonal error elimination framework proposed.

4.8. Comparison

For each change detection model, the comparison experiment began by testing the $cdmodel$ described in Section 4.5 on three datasets: $dataset_{snow-covered}$, $dataset_{snow-free}$, and $dataset_{total}$. These tests are performed to establish baseline performance metrics for the model when applied to these datasets under standard conditions, without incorporating any seasonal error elimination techniques. This initial testing allowed us to obtain the model's effectiveness and identify areas for improvement.

Next, the modified $cdmodel'$ along with the two TIFFGs discussed in Section 4.5 are integrated into the DBSEE-CDF. The updated framework is then tested on the same three datasets— $dataset_{snow-covered}$, $dataset_{snow-free}$, and $dataset_{total}$ —to evaluate the performance improvements brought about by the incorporation of seasonal error elimination. This

allowed us to assess how well the proposed framework addressed seasonal errors and whether it led to better change detection accuracy.

Finally, the evaluation is extended by connecting the various image-to-image translation models described in Section 4.4 to *cdmodel*³, *cdmodel*⁴, *cdmodel*⁵, *cdmodel*⁶, and *cdmodel*⁷. These models are also tested on *dataset_{snow-covered}*, *dataset_{snow-free}*, and *dataset_{total}* to provide a comparative analysis. This step allowed us to benchmark the performance of the proposed DBSEE-CDF against a variety of other image-to-image translation methods, further validating the effectiveness of the proposed approach.

The F1-score results of all the above experiments are shown in Table 1.

Five pairs of images are sampled with corresponding change labels from *dataset_{snow-covered}* and *dataset_{snow-free}*, respectively. Using CGNet-CD, Changeformer, and HANet-CD, we first applied the proposed dual-branch seasonal error elimination method before conducting change detection. Then, we directly performed change detection. The final change detection results for both approaches are shown in Figure 4, illustrating that the proposed dual-branch seasonal error elimination method effectively enhances the performance of change detection tasks for these three change detection models.

Table 1. F1-score of CGNet-CD [14], Change Former [15], and HANet-CD [16] change detection models on three datasets: *dataset_{snow-covered}*, *dataset_{snow-free}*, and *dataset_{total}*, under the conditions of using the proposed dual-branch seasonal error elimination method, not using the dual-branch seasonal error elimination method, and using CycleGAN, SuperstarGAN, Pnp-diffusion, and EGSDE to eliminate seasonal errors before change detection task. For each of the change detection models and each of the three datasets, the bolded portions represent the highest F1-score of the change detection tasks among all mentioned transfer methods. The F1-score results of the change detection task after seasonal error elimination using the proposed method can be found in the rows where their stages are *dual TIFFG transfer*.

CD-Model	Stage	<i>dataset_{snow-covered}</i>	<i>dataset_{snow-free}</i>	<i>dataset_{total}</i>
CGNet-CD [14]	no transfer	0.6596	0.6901	0.6726
	SWAT transfer	0.6738	0.6924	0.6817
	CycleGAN transfer	0.7149	0.7473	0.7287
	SuperstarGAN transfer	0.6174	0.6668	0.6385
	Pnp-diffusion transfer	0.6679	0.7018	0.6823
	EGSDE transfer	0.6700	0.7687	0.7121
	dual TIFFG transfer	0.7164	0.7845	0.7455
Change-Former [15]	no transfer	0.6363	0.6514	0.6468
	SWAT transfer	0.6387	0.6787	0.6558
	CycleGAN transfer	0.6661	0.6938	0.6779
	SuperstarGAN transfer	0.6663	0.6937	0.6780
	Pnp-diffusion transfer	0.6549	0.7023	0.6751
	EGSDE transfer	0.6671	0.7393	0.6979
	dual TIFFG transfer	0.6822	0.7494	0.7108

Table 1. Cont.

CD-Model	Stage	$dataset_{snow-covered}$	$dataset_{snow-free}$	$dataset_{total}$
HANet-CD [16]	no transfer	0.6390	0.6573	0.6468
	SWAT transfer	0.6459	0.6763	0.6589
	CycleGAN transfer	0.6893	0.7458	0.7134
	SuperstarGAN transfer	0.6634	0.7821	0.7140
	Pnp-diffusion transfer	0.6713	0.6646	0.6684
	EGSDE transfer	0.6857	0.6670	0.6777
	dual TIFFG transfer	0.6954	0.7532	0.7200

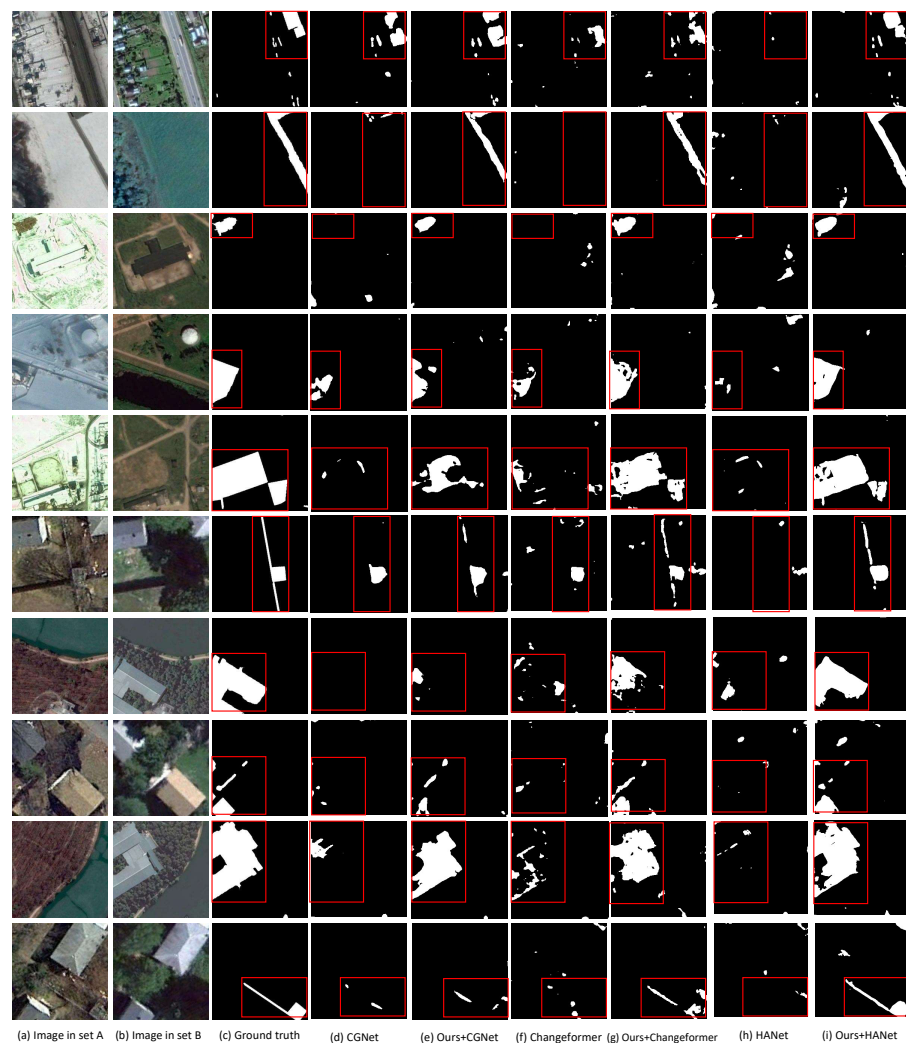


Figure 4. Change detection result comparison. The top five pairs of images are from $dataset_{snow-covered}$, and the bottom five pairs of images are from $dataset_{snow-free}$. (a) Image in set A. (b) Image in set B. (c) Ground truth change map. (d) Change map using CGNet-CD without seasonal error elimination. (e) Change map using CGNet-CD with seasonal error elimination. (f) Change map using Changeformer without seasonal error elimination. (g) Change map using Changeformer with seasonal error elimination. (h) Change map using HANet-CD without seasonal error elimination. (i) Change map using HANet-CD with seasonal error elimination.

4.9. Ablation Study

To validate the effectiveness and necessity of both the target image feature extraction and fusion module, as well as the dual-branch structure, a series of ablation experiments are conducted. These experiments aimed to isolate and evaluate the individual contributions of these components to the overall performance of the framework.

First, to assess the impact of the target image feature extraction and fusion module, two TIFFGs are integrated into the DBSEE-CDF, resulting in the network configuration N_{ours} . This configuration includes the full functionality of both TIFFGs, which process the seasonal features of the images to enhance the accuracy of change detection.

Next, an experiment is conducted where the target image feature extraction and fusion modules are removed from the two TIFFGs. The remaining TIFFGs are then embedded into the DBSEE-CDF, forming the network configuration N_{double} . This experiment allowed us to observe how the removal of these modules impacted the framework's ability to address seasonal errors and perform change detection tasks effectively.

To evaluate the necessity of the dual-branch structure, a single TIFFG is connected to the change detection models, omitting the seasonal classification module. This formed a single-branch seasonal error elimination change detection network, denoted as $N_{single-TIFFG}$. By using a single TIFFG without the full dual-branch architecture, it could be assessed whether the additional branch structure provided significant improvements in eliminating seasonal errors and enhancing change detection performance.

These three network configurations— N_{ours} , N_{double} , and $N_{single-TIFFG}$ —are trained according to the methods described in Sections 4.4 and 4.5. This training process resulted in distinct generators and corresponding change detection models for each configuration. After training, each network is tested on $dataset_{snow-covered}$, $dataset_{snow-free}$, and $dataset_{total}$ using a variety of change detection models to evaluate their performance.

The results of these ablation experiments, which provide insights into the individual contributions of each module and structure, are presented in Table 2.

Table 2. F1-score of CGNet-CD [14], Change Former [15], and HANet-CD [16] change detection models on three datasets in ablation experiments. $N_{single-TIFFG}$ represents the elimination of the dual-branch structure, using only one TIFFG for seasonal error elimination. N_{double} represents performing seasonal error elimination after eliminating the target image feature extraction and fusion module in the proposed method. N_{ours} represents using the proposed method for seasonal error elimination. For each of the change detection models and each of the three datasets, the bolded portions represent the highest F1-score of change detection task among all mentioned transfer methods. The F1-score results of the change detection task after seasonal error elimination using the proposed method can be found in the rows where their stages are N_{ours} .

Model	Stage	$dataset_{snow-covered}$	$dataset_{snow-free}$	$dataset_{total}$
CGNet-CD [14]	$N_{single-TIFFG}$	0.6991	0.7624	0.7261
	N_{double}	0.7061	0.7574	0.7280
	N_{ours}	0.7164	0.7845	0.7455
Change Former [15]	$N_{single-TIFFG}$	0.6613	0.7304	0.6908
	N_{double}	0.6700	0.7272	0.6944
	N_{ours}	0.6822	0.7494	0.7108
HANet-CD [16]	$N_{single-TIFFG}$	0.6784	0.7269	0.6991
	N_{double}	0.6768	0.7232	0.6966
	N_{ours}	0.6954	0.7532	0.7200

5. Discussion

5.1. Comparative Experiment Analysis

It can be observed from Table 1 that after applying the DBSEE-CDF, all the change detection models used in this study showed improvements in the change detection task. Among them, CGNet-CD performed the best on $dataset_{snow-free}$, $dataset_{snow-covered}$, and $dataset_{total}$. After applying the proposed method, its F1-score improved by 6%, 9%, and 7% on $dataset_{snow-covered}$, $dataset_{snow-free}$ and $dataset_{total}$, respectively. In addition, the F1-score of Change Former improved by 4%, 9%, and 6% on the $dataset_{snow-covered}$, $dataset_{snow-free}$, and $dataset_{total}$ datasets, respectively, while the F1-score of HANet-CD also improved by 5%, 9%, and 7% on the $dataset_{snow-covered}$, $dataset_{snow-free}$, and $dataset_{total}$ datasets, respectively. In addition, compared to other image-to-image translation models, the method proposed in this paper achieves a greater improvement in the F1-score for the elimination of seasonal errors. These results demonstrate that the method proposed in this study effectively eliminates seasonality errors, thereby mitigating their negative impact on the change detection task. By addressing these errors, the method enhances the accuracy and reliability of change detection, leading to notable improvements in task performance. Furthermore, our innovative DBSEE-CDF surpasses existing seasonal error elimination networks, showcasing superior capabilities in handling seasonality-related challenges. This performance advantage underscores the robustness and efficiency of the proposed framework in addressing the complex issue of seasonal errors, making it a valuable contribution to the field of change detection.

From Figure 4, it is evident that performing change detection tasks without applying seasonal error elimination is significantly hindered by the presence of seasonal errors. These errors negatively affect the model's ability to accurately recognize changes in critical features such as roads and buildings, resulting in notable inaccuracies and performance losses. In contrast, the DBSEE-CDF introduced in this paper, which incorporates the innovative TIFFG, demonstrates a remarkable ability to mitigate the adverse effects of seasonal errors. By effectively addressing these issues, the proposed framework significantly enhances the model's capability to detect and recognize changes in roads and buildings with greater precision, thereby improving the overall performance and reliability of change detection tasks. Additionally, by adopting a dual-branch approach, where the TIFFG trained on different datasets processes snow-covered winter images and snow-free winter images separately, our model effectively eliminates seasonal errors for both types of surface coverage data while avoiding severe loss of visual fidelity. Furthermore, by incorporating guidance from target image information in the TIFFG, the model demonstrates improved transformation accuracy. For instance, as shown in rows 1 and 5 of Figure 4, the winter images feature black roads, but in non-winter images, the first row depicts asphalt roads, while the second row shows rural dirt roads. By leveraging the texture and color information from non-winter images, the model partially avoids the incorrect transformation of the original image, and thus enhances the accuracy of change detection tasks.

5.2. Ablation Experiment Analysis

As shown in Figure 4, the proposed framework significantly outperforms approaches that do not incorporate target image feature extraction and fusion or lack the dual-branch structure. By integrating these innovative components, the framework achieves a marked improvement in the performance of change detection tasks, consistently attaining the highest F1-score across various change detection models and the comprehensive dataset $dataset_{total}$. Specifically, for each change detection model, using the proposed dual-branch TIFFG N_{ours} for seasonal error elimination achieves a higher F1-score in both snow-covered and snow-free scenes compared to using a single TIFFG $N_{single-TIFFG}$, which demonstrates

the necessity of the dual-branch structure. Additionally, using the proposed dual-branch TIFFG N_{ours} for seasonal error elimination achieves a higher F1-score than using a converter that removes the target image feature extraction and fusion module from the dual-branch TIFFG N_{double} , in both snow-covered and snow-free scenes. This proves the necessity of target image feature extraction and fusion. This validation underscores both the effectiveness and the necessity of these design choices in achieving superior results in change detection applications.

5.3. Benefits and Shortcomings

By adding the target image feature extraction and fusion module, the proposed method guides the seasonal error elimination model's transformation using the texture and color features of the target image, thereby solving the problem of incorrect transformation of the original image. The proposed method also addresses the problem of significant coverage, color and texture differences and style transfer requirements between snow-covered and snow-free winter images by using a dual-branch structure for seasonal error elimination separately for both types of images. Experimental results show that, compared to other methods, the proposed approach successfully improves the performance of the change detection task, with a greater improvement than existing image transformation methods. In addition, the proposed method improves the performance of the change detection task across multiple change detection models, which demonstrates the versatility of the proposed approach.

However, the proposed method still has certain limitations. First, the two TIFFGs in the dual-branch structure, as well as the seasonal classification model and change detection model, all require separate training, and the training process is relatively complex, which increases the difficulty of using the model. Second, since the proposed TIFFG incorporates the target image feature extraction and fusion module and needs to be embedded into the CycleGAN network for training, this increases the training overhead. Finally, the proposed method requires a large amount of data for training, which is a disadvantage in remote sensing image scenarios with high resolution and low data availability.

6. Conclusions

In summary, we proposed the DBSEE-CDF method.

In our study, it was found that in seasonal error elimination, since the downstream task is change detection, it is crucial to preserve the land cover types and uses in certain areas of the winter image while aligning the texture and color as closely as possible with the corresponding areas in the non-winter image. Therefore, it is essential to capture the land cover and use information from the winter image's certain areas and the texture and color information from the corresponding areas of the non-winter image, and to integrate both into the transformation process. To achieve this, TIFFG is proposed, which performs down-sampling on the original images to obtain deep features. Using a ResNet encoder and a mixed attention module, the method generates deep features for the target images and fuses them through a cross-attention module. The fused features are then up-sampled five times to generate the final images. The experimental results show that the proposed network effectively enhances the performance of various change detection models and mitigates the impact of seasonal errors on remote sensing image change detection tasks.

We also discovered that during seasonal error elimination, the snow-covered areas have lighter colors and reflect more sunlight, leading to significant differences in texture and color features between snow-covered and snow-free regions in remote sensing images. This results in a severe loss of visual fidelity caused by the significant differences in texture and color features as well as the different requirements for style transformation. To address this issue, the proposed method first classifies pairs of remote sensing images and processes and

transforms snow-covered and snow-free winter images through a dual-branch structure with different TIFFGs. The transformed images are then input along with non-winter images into the change detection model to generate the change detection results.

However, the proposed method still has certain limitations such as complex training processes, a high difficulty of use, and significant training time overhead. In the future, we will simplify the model structure and training process to reduce the difficulty of using the model and the training overhead. Additionally, to better capture the land cover types and uses in certain areas of the winter image and integrate them into the transformation process, we plan to incorporate semantic information from the winter images into the model to further enhance its performance.

Author Contributions: Conceptualization, H.Z., and J.Z.; methodology, H.Z. and J.Z.; software, J.Z.; validation, J.Z.; formal analysis, H.Z., J.Z., X.L., and Q.L.; investigation, J.Z. and Z.W.; resources, H.Z. and Q.L.; data curation, J.Z.; writing—original draft preparation, J.Z. and X.L.; writing—review and editing, H.Z., B.D. and Q.L.; visualization, J.Z.; supervision, H.Z.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Research and Application of Key Technologies for Enhancing the Environmental Quality of Livable Cities (Grant No. 2023YFC3805305), the Science and Technology Commission of Shanghai Municipality (No. 23511103100), the Shanghai Pujiang Program (No. 23PJ1412700,) and the National Natural Science Foundation of China (No. 62406227).

Data Availability Statement: The CDdataset used in the article is publicly available at https://drive.google.com/file/d/1GX656JqqOyBi_Ef0w65kDGVto-nHrNs9 (accessed on 8 March 2024).

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

References

1. Mirzakarimova, G.M. Remote sensing data: International experiences and applications. *Actual Probl. Sci. Educ. Face Mod. Challenges* **2023**, *14*, 1.
2. Jiang, H.; Peng, M.; Zhong, Y.; Xie, H.; Hao, Z.; Lin, J.; Ma, X.; Hu, X. A Survey on Deep Learning-Based Change Detection from High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 1552. [\[CrossRef\]](#)
3. Bai, T.; Wang, L.; Yin, D.; Sun, K.; Chen, Y.; Li, W.; Li, D. Deep learning for change detection in remote sensing: A review. *Geo-Spat. Inf. Sci.* **2023**, *26*, 262–288. [\[CrossRef\]](#)
4. Liu, C.; Chen, K.; Qi, Z.; Liu, Z.; Zhang, H.; Zou, Z.; Shi, Z. Pixel-level change detection pseudo-label learning for remote sensing change captioning. In Proceedings of the IGARSS 2024—2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 8405–8408.
5. Stilla, U.; Xu, Y. Change detection of urban objects using 3D point clouds: A review. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 228–255. [\[CrossRef\]](#)
6. Ding, L.; Zhu, K.; Peng, D.; Tang, H.; Yang, K.; Bruzzone, L. Adapting segment anything model for change detection in VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5611711. [\[CrossRef\]](#)
7. Bovolo, F.; Bruzzone, L. A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Trans. Geosci. Remote Sens.* **2006**, *45*, 218–236. [\[CrossRef\]](#)
8. Wu, C.; Du, B.; Zhang, L. Slow Feature Analysis for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2858–2874. [\[CrossRef\]](#)
9. Celik, T. Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and *k*-Means Clustering. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 772–776. [\[CrossRef\]](#)
10. Fan, T.; Wang, G.; Li, Y.; Wang, H. MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. *IEEE Access* **2020**, *8*, 179656–179665. [\[CrossRef\]](#)
11. Xing, Y.; Jiang, J.; Xiang, J.; Yan, E.; Song, Y.; Mo, D. Lightcdnet: Lightweight change detection network based on vhr images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 2504105. [\[CrossRef\]](#)
12. Fang, S.; Li, K.; Li, Z. Changer: Feature interaction is what you need for change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5610111. [\[CrossRef\]](#)
13. Codegoni, A.; Lombardi, G.; Ferrari, A. TINYCD: A (not so) deep learning model for change detection. *Neural Comput. Appl.* **2023**, *35*, 8471–8486. [\[CrossRef\]](#)

14. 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](#)]
15. Bandara, W.G.C.; Patel, V.M. A transformer-based siamese network for change detection. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 207–210.
16. Han, C.; Wu, C.; Guo, H.; Hu, M.; Chen, H. HANet: A Hierarchical Attention Network for Change Detection with Bitemporal Very-High-Resolution Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3867–3878. [[CrossRef](#)]
17. Zheng, Z.; Ma, A.; Zhang, L.; Zhong, Y. Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 11–15 June 2021; pp. 15193–15202.
18. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5607514. [[CrossRef](#)]
19. 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.* **2021**, *19*, 8007805. [[CrossRef](#)]
20. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
21. Park, T.; Zhu, J.Y.; Wang, O.; Lu, J.; Shechtman, E.; Efros, A.; Zhang, R. Swapping autoencoder for deep image manipulation. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7198–7211.
22. Tumanyan, N.; Geyer, M.; Bagon, S.; Dekel, T. Plug-and-play diffusion features for text-driven image-to-image translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1921–1930.
23. Ko, K.; Yeom, T.; Lee, M. Superstargan: Generative adversarial networks for image-to-image translation in large-scale domains. *Neural Netw.* **2023**, *162*, 330–339. [[CrossRef](#)]
24. Zhao, M.; Bao, F.; Li, C.; Zhu, J. Egsde: Unpaired image-to-image translation via energy-guided stochastic differential equations. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 3609–3623.
25. Parmar, G.; Kumar Singh, K.; Zhang, R.; Li, Y.; Lu, J.; Zhu, J.Y. Zero-shot image-to-image translation. In Proceedings of the ACM SIGGRAPH 2023 Conference Proceedings, Los Angeles, CA, USA, 6–10 August 2023; pp. 1–11.
26. Xie, S.; Xu, Y.; Gong, M.; Zhang, K. Unpaired image-to-image translation with shortest path regularization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 10177–10187.
27. Zhang, J.; Lang, X.; Huang, B.; Jiang, X. VAE-CoGAN: Unpaired image-to-image translation for low-level vision. *Signal Image Video Process.* **2023**, *17*, 1019–1026. [[CrossRef](#)]
28. Wang, C.; Chen, B.; Zou, Z.; Shi, Z. Remote sensing image synthesis via semantic embedding generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4702811. [[CrossRef](#)]
29. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Huang, M.; Plaza, A. BockNet: Blind-Block Reconstruction Network with a Guard Window for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5531916. [[CrossRef](#)]
30. Cheng, X.; Huo, Y.; Lin, S.; Dong, Y.; Zhao, S.; Zhang, M.; Wang, H. Deep Feature Aggregation Network for Hyperspectral Anomaly Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5033016. [[CrossRef](#)]
31. Wang, M.; Gao, L.; Ren, L.; Sun, X.; Chanussot, J. Hyperspectral Simultaneous Anomaly Detection and Denoising: Insights From Integrative Perspective. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 13966–13980. [[CrossRef](#)]
32. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.
33. Lee, M.; Seok, J. Controllable generative adversarial network. *IEEE Access* **2019**, *7*, 28158–28169. [[CrossRef](#)]
34. 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; pp. 770–778.
35. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
36. Lebedev, M.; Vizilter, Y.V.; Vygolov, O.; Knyaz, V.A.; Rubis, A.Y. Change detection in remote sensing images using conditional adversarial networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 565–571. [[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.