remote sensing

MDPI

Article

Attention-Guided Generative Adversarial Network for False Alarm-Resistant Change Detection in Remote Sensing Orthophotos

Yuxuan Hu [1], Zheng Ji [1,*], Wei Liu [2] and Yichao Li [1]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; 2023302081281@whu.edu.cn (Y.H.); 2021302131172@whu.edu.cn (Y.L.)
[2] Xi'an Institute of Surveying and Mapping, Xi'an 710054, China; liuwei45_2000@126.com
* Correspondence: jz07@whu.edu.cn

Highlights

What are the main findings?

- An Attention-GAN framework is proposed to detect real orthophoto changes while suppressing false alarms caused by residual geometric and radiometric inconsistencies.
- Perturbation-negative samples are used as unchanged examples, and thus the model learns that small nuisance differences should not be labeled as land cover change.

What are the implications of the main findings?

- The approach provides a reproducible training and evaluation protocol for false alarm resistance in orthophoto change detection.
- It is suited to operational mapping scenarios where candidate change maps must remain useful despite imperfect co-registration and illumination consistency.

Abstract

Orthophoto change detection is used to find real land cover changes in urban monitoring, disaster assessment, and environmental management. In practice, however, multi-temporal orthophotos are rarely identical in geometry and radiometry even after standard preprocessing. Small residual misregistration, local building displacement, shadow movement, and illumination differences can produce edge-like responses that look like change but do not correspond to any land cover transition. These false alarms increase manual checking costs and reduce the reliability of change maps. This study addresses that practical problem by proposing an attention-guided conditional adversarial framework, named Attention-GAN, for false alarm-resistant orthophoto change detection. The aim is not to detect small perturbations as changes but to detect real land cover changes while suppressing responses to nuisance variations that should be treated as unchanged. The framework integrates a multi-scale spatial attention module, a channel attention module, and a PatchGAN discriminator. It also introduces perturbation-negative training pairs, where controlled geometric and radiometric perturbations are applied to unchanged image pairs and assigned all-zero change masks. Experiments on LEVIR-CD, WHU-CD, and DSIFN-CD show competitive or moderately higher accuracy than the selected representative baselines, with F1 scores of 91.2%, 92.45%, and 93.18%, respectively. In the ablation experiment, the false change rate on perturbation-negative validation pairs is reduced to 4.9%. Repeated-run statistics and ablation results indicate that the proposed training strategy mainly improves robustness by reducing false alarms under the evaluated perturbation range. The results support the use of controlled nuisance perturbations as a reproducible way to train and evaluate false

Check for updates

Academic Editor: John Trinder

Received: 5 June 2026
Revised: 3 July 2026
Accepted: 7 July 2026
Published: 8 July 2026

Copyright: © 2026 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.

alarm resistance, while broader validation under real multi-view, seasonal, and cross-sensor distortions remains necessary.

Keywords: remote sensing imagery; change detection; generative adversarial networks; attention mechanism; false alarm reduction; nuisance robustness

1. Introduction

The rapid development of remote sensing technology has produced dense archives of multi-temporal imagery for monitoring land-surface dynamics [1,2]. Orthophotos are particularly useful for change detection because they are geometrically corrected, georeferenced, and presented at a near-uniform scale. Comparing orthophotos acquired at different times is therefore a common way to locate building construction, land cover transitions, urban expansion, and disaster damage, supporting urban planning, emergency assessment, and environmental management [3,4].

In operational orthophoto change detection, the desired output is not a map of every pixel-level difference between two images. The desired output is a map of real land cover change. This distinction is important because multi-temporal orthophotos are rarely perfectly matched, even after standard preprocessing [5,6]. Differences in illumination, solar altitude angle, atmospheric condition, sensor response, and acquisition season can change image radiometry. Terrain relief, building displacement, off-nadir viewing, residual registration error, orthorectification uncertainty, and shadow movement can also create local geometric inconsistencies [7,8]. These factors are usually nuisance variations rather than target changes, but they often appear along building edges, roads, and object boundaries, where many change detectors are most sensitive [9–11].

This problem matters in practical mapping. In many monitoring tasks, unchanged areas occupy most of the image, and only a small fraction contains real change. A detector that responds strongly to residual misalignment or brightness differences may obtain visually plausible masks but still generate many false alarms. These false alarms increase manual verification effort and can hide the true change signal in downstream statistics. Therefore, the goal of this study is not to detect small perturbations themselves. The goal is to make the change detector less sensitive to small geometric and radiometric differences that should be ignored while preserving sensitivity to real land cover change.

Traditional remote sensing change detection methods include image differencing, image ratioing, change vector analysis, and principal component analysis [12]. These methods are efficient but usually rely on handcrafted features or manually selected thresholds, and thus their performance can degrade under illumination variation, registration residuals, and heterogeneous land cover changes. Deep learning methods, especially convolutional neural network (CNN)-based models, have improved feature representation for change detection. Representative models include FC-EF, FC-Siam-Diff, and FC-Siam-Conc. Nevertheless, many existing methods still focus mainly on standard benchmark accuracy and do not explicitly train or evaluate the ability to reject false responses caused by nuisance variation [13–15].

Generative adversarial networks (GANs) [16] have been widely used for image generation, image translation, and representation learning [17,18]. In change detection, adversarial supervision can encourage predicted change maps to be more spatially coherent. However, adversarial training alone does not tell the model which image differences should be ignored. A change detector also needs negative evidence, namely examples in which two images are different in appearance but should still be labeled as unchanged. This motivates the use of perturbation-negative samples in this study.

To address the above issues, this study proposes an attention-based generative adversarial network (Attention-GAN) for false alarm-resistant orthophoto change detection. The model combines a Siamese CNN backbone, a multi-scale spatial attention module, a channel attention module, and a PatchGAN discriminator. In addition, controlled geometric and radiometric perturbations are applied to construct perturbation-negative image pairs with all-zero change masks. These samples do not define a new semantic class. They are used to teach the model that residual misalignment and radiometric differences are nuisance variables when no land cover change is present. We emphasize that the individual modules are established components; the contribution lies in their task-oriented integration with a reproducible negative-sample training and evaluation protocol for reducing false alarms. Extensive experiments on multiple public datasets show competitive or moderately improved accuracy compared with the selected representative approaches, together with lower false change rates under the evaluated perturbation settings.

The remainder of this study is organized as follows. Section 2 reviews the related work. Section 3 presents the problem formulation and the proposed network architecture. Section 4 describes the datasets, test areas, implementation details, and comparison methods. Section 5 reports the experimental results and analysis. Section 6 discusses the main findings and limitations of the proposed method. Finally, Section 7 concludes the paper and outlines future research directions.

2. Related Work

2.1. Traditional and Deep Learning-Based Change Detection

Remote sensing change detection has moved from pixel- and object-based analysis to deep learning models [12]. Classical approaches such as image differencing, ratioing, change vector analysis (CVA), and principal component analysis (PCA) are efficient, but they are sensitive to radiometric inconsistency, geometric residuals, and threshold selection [5,6]. CNN-based methods, including FC-EF, FC-Siam-Diff, and FC-Siam-Conc, learn paired-image features and predict change maps from end to end [14,19,20]. Later studies introduced spatiotemporal attention, dense Siamese connections, and transformer architectures such as STANet, SNUNet, BIT, and ChangeFormer to improve feature interaction and context modeling [21–24]. Recent lightweight and hard sample-oriented methods, including TinyCD and HSONet, further refine high-resolution change detection [8,25]. Foundation models and self-supervised geospatial representation learning are also becoming relevant because they provide transferable visual features [26,27]. Most of these methods are evaluated on real-change masks, while their behavior on unchanged pairs with residual shifts or illumination differences is measured less often [9–11].

2.2. GANs and Adversarial Training for Change Detection

Generative adversarial networks (GANs) [16] and conditional GANs (cGANs) [17] have been used in remote sensing for image-to-image translation, domain adaptation, and semantic segmentation [18]. In change detection, the generator produces a change map from a bi-temporal input, and the discriminator judges whether the map is realistic, encouraging the generator to align with the distribution of real change patterns [28]. Adversarial training can improve boundary sharpness and reduce over-smoothing compared with pure supervised segmentation losses. Existing GAN-based change detection methods, however, often emphasize the realism of positive change masks. They provide less direct supervision for the opposite requirement; unchanged pairs should not produce fragmented or edge-like false positives when the two dates differ only because of imaging conditions. PatchGAN discriminators [18] are commonly adopted to enforce local realism; we follow this design and combine it with attention modules and perturbation-negative samples.

2.3. Attention Mechanisms in Vision and Remote Sensing

Attention modules reweight spatial locations or feature channels according to their relevance. Spatial attention can strengthen change boundaries, while channel attention emphasizes semantic responses that are useful for separating changed and unchanged regions. The Squeeze-and-Excitation (SE) module [29] and related designs such as CBAM [30] have been widely used in classification and segmentation [31,32]. In remote sensing change detection, attention has been combined with Siamese and transformer backbones to improve multi-scale feature interaction on building and land cover datasets [31–34]. In this study, MSAM is applied after bi-temporal feature fusion, and SE recalibrates the fused channels. They are used as established attention components within a PatchGAN-supervised framework rather than as standalone methodological novelties.

3. Methodology

3.1. Problem Definition

The objective of the change detection task is to identify regions that have undergone real land cover changes in a pair of bi-temporal remote sensing images $I_1, I_2 \in \mathbb{R}^{H\times W\times C}$, where H, W, and C denote the image's height, width, and number of channels, respectively. The objective is to learn a mapping from the input image pair (I_1, I_2) to a binary change map $Y \in \{0,1\}^{H\times W}$, in which $Y_{ij} = 1$ represents a changed pixel and $Y_{ij} = 0$ represents an unchanged pixel.

In this formulation, "unchanged" does not mean that the two image patches are numerically identical. It means that no target land cover transition is present. Small residual shifts, illumination differences, and other acquisition-related discrepancies are treated as nuisance variation. A useful model should therefore satisfy two requirements at the same time: it should respond to real object-level change, and it should remain conservative when the observed differences can be explained by nuisance variation.

To this end, we formulate the task within a conditional generative adversarial network (cGAN) framework. The generator G learns to predict the change map

$$\hat{Y} = G(I_1, I_2), \tag{1}$$

conditioned on the bi-temporal images, whereas the discriminator D aims to distinguish the real change map Y from the generated one $\hat{Y}$ under the same condition. By optimizing the two networks in an adversarial manner, the proposed framework encourages the generator to produce more realistic and structurally coherent change maps, thereby improving the quality of the change detection results.

3.2. Network Architecture Overview

In this study, we used a conditional adversarial network with multi-scale spatial attention and channel attention. The generator was constructed on a dual-tower Siamese CNN [14,22]; the two towers extract features from I_1 and I_2 separately with shared weights, and the fused features are then processed by the Multi-Scale Spatial Attention Module (MSAM) and the Squeeze-and-Excitation (SE) channel attention module [29,31,32]. This design combines temporal feature comparison with spatial and channel feature reweighting. The decoder then generates the final change probability map.

The discriminator employs the PatchGAN structure, which makes local judgments about image patches rather than the entire image. This local patch-based discrimination helps the model capture fine details in the change map. The generator and discriminator were optimized adversarially to improve the quality of the generated change map.

The proposed model differs from STANet and SNUNet mainly in its training objective and supervision structure. STANet focuses on spatiotemporal attention for correspondence modeling, and SNUNet relies on dense Siamese feature fusion. The Attention-GAN combines multi-scale spatial attention, channel recalibration, PatchGAN supervision, and perturbation-negative samples so that unchanged but visually inconsistent pairs are explicitly represented during training. The contribution is therefore the task-oriented integration and evaluation protocol, not the invention of the MSAM, SE, or PatchGAN.

3.3. Generator Network

3.3.1. Dual-Tower Siamese CNN Backbone

The generator adopts a dual-tower Siamese CNN as the backbone for change detection [14,22]. In this architecture, the two towers process the bi-temporal images I_1 and I_2 separately while sharing the same network parameters. This weight-sharing strategy ensures that both images are encoded in a common feature space, enabling consistent comparison of features at corresponding spatial locations.

Compared with early-fusion strategies that concatenate I_1 and I_2 before a single encoder, the dual-tower design explicitly models temporal correspondence and typically yields better performance in change detection tasks [14,20].

Each tower employs the same CNN encoder E (e.g., ResNet blocks [35] or VGG-style convolutional layers) to extract hierarchical features over L stages. The feature maps at stage l are defined as follows:

$$F_1^l = E^l(I_1), \qquad F_2^l = E^l(I_2), \quad l = 1, 2, \ldots, L, \tag{2}$$

where E^l denotes the shared encoder up to stage l. The feature maps F_1^l and F_2^l have dimensions $H^l \times W^l \times C^l$.

The two branches operate in parallel, and their outputs are fused through channel-wise concatenation:

$$F^l = [F_1^l; F_2^l], \tag{3}$$

where $F^l \in \mathbb{R}^{H^l \times W^l \times 2C^l}$. This fusion strategy preserves complementary spatial and semantic information from both temporal observations and provides the basis for subsequent attention-based feature refinement.

3.3.2. Multi-Scale Spatial Attention Module (MSAM)

To capture change patterns at different spatial scales, we introduce the MSAM. Given the fused feature map F^l, parallel convolution operations with different kernel sizes are applied:

$$F_k^l = \delta(\mathrm{Conv}_k(F^l)), \quad k \in \{3, 5, 7\}, \tag{4}$$

where $\delta(\cdot)$ denotes the ReLU activation function.The multi-scale features are then aggregated:

$$F_{\mathrm{sum}}^l = F_3^l + F_5^l + F_7^l. \tag{5}$$

Next, a spatial attention map is generated:

$$A_s^l = \sigma(\mathrm{Conv}_{1\times1}(F_{\mathrm{sum}}^l)), \tag{6}$$

where $\sigma(\cdot)$ denotes the sigmoid activation. The final spatially enhanced feature is obtained by

$$\tilde{F}^l = F_{\mathrm{sum}}^l \odot A_s^l. \tag{7}$$

This module enables the network to emphasize spatial regions corresponding to potential changes while suppressing irrelevant responses.

3.3.3. Squeeze-and-Excitation (SE) Module

To further enhance feature discrimination, we incorporated the Squeeze-and-Excitation (SE) module to model channel-wise dependencies. Given the spatially refined feature $\tilde{F}^l$, global average pooling is applied first:

$$z_c = \frac{1}{H^l W^l} \sum_{i=1}^{H^l} \sum_{j=1}^{W^l} \tilde{F}^l_{c,i,j}. \tag{8}$$

Then, two fully connected layers capture channel interactions:

$$s = \sigma(W_2\, \delta(W_1 z)), \tag{9}$$

where $\delta(\cdot)$ denotes the ReLU and $\sigma(\cdot)$ denotes sigmoid activation. The channel-refined feature is

$$F^l_{\text{se}} = \tilde{F}^l \odot s, \tag{10}$$

where s is broadcast along spatial dimensions. This mechanism adaptively highlights informative channels while suppressing redundant features.

3.3.4. Decoder

The decoder progressively restores the spatial resolution of encoded features to generate the final change map. Let F^l_d denote the decoder feature at stage l. The decoding process can be written as follows:

$$F^{l-1}_d = \phi(\text{Up}(F^l_d) \oplus F^{l-1}_{\text{se}}), \tag{11}$$

where $\text{Up}(\cdot)$ denotes upsampling, $\oplus$ represents feature concatenation, and $\phi(\cdot)$ denotes convolution followed by nonlinear activation.

Finally, a 1×1 convolution followed by sigmoid activation produces the change probability map:

$$\hat{Y} = \sigma(\text{Conv}_{1\times1}(F^1_d)). \tag{12}$$

Each pixel value in $\hat{Y}$ represents the probability that the corresponding location belongs to the changed class.

3.4. Discriminator Network

The discriminator adopts a PatchGAN structure, which performs local patch-based discrimination rather than judging the entire image. In the conditional adversarial setting, the discriminator receives the bi-temporal input pair together with either the reference change map or the predicted change map. It then outputs a patch-wise probability map:

$$D(I_1, I_2, Y) \in \mathbb{R}^{H_d \times W_d}, \tag{13}$$

where each element represents the probability that the corresponding local patch belongs to the real distribution. This local discrimination strategy encourages the generator to produce spatially consistent and realistic change patterns, which is particularly beneficial for fine-grained change detection tasks.

3.5. Loss Functions

The generator is trained with a composite loss consisting of adversarial loss, segmentation loss, and perceptual loss. The three terms target different aspects of the prediction: local realism, pixel-level accuracy, and structural consistency.

3.5.1. Adversarial Loss

The adversarial objective is written separately for the discriminator and generator. The discriminator is trained to distinguish the reference change map from the generated change map under the same image pair condition:

$$\mathcal{L}_D = -\mathbb{E}[\log D(I_1, I_2, Y^{\mathrm{gt}})] - \mathbb{E}[\log(1 - D(I_1, I_2, \hat{Y}))]. \tag{14}$$

The generator is trained to make the predicted change map difficult for the discriminator to distinguish from a reference map:

$$\mathcal{L}_{\mathrm{adv}}^{G} = -\mathbb{E}[\log D(I_1, I_2, \hat{Y})]. \tag{15}$$

3.5.2. Segmentation Loss

Segmentation loss measures the pixel-wise difference between the generated change map and the real change map. We use the binary cross-entropy (BCE) loss to calculate the error between the predicted and true change maps:

$$\mathcal{L}_{\mathrm{seg}} = -\frac{1}{N}\sum_{i,j}\Big[Y_{ij}^{\mathrm{gt}} \log \hat{Y}_{ij} + (1 - Y_{ij}^{\mathrm{gt}}) \log(1 - \hat{Y}_{ij})\Big], \tag{16}$$

where Y^{gt} is the ground-truth change map and $\hat{Y}$ is the predicted change map.

3.5.3. Perceptual Loss

To maintain structural consistency, we use the perceptual loss [36] to measure the difference between the generated and real change maps in a feature space. In the implementation, both the predicted probability map and the binary reference map are replicated to three channels and fed into an ImageNet-pretrained VGG-16 network. Features from layer `relu3_3` are used to compute the perceptual loss, while VGG parameters are frozen during training. We selected `relu3_3` because it provides a compromise between local boundary information and mid-level structural representation; lower layers were more sensitive to pixel noise, whereas deeper layers were less responsive to small building changes. Perceptual loss encourages the predicted change map to remain close to the reference map in this feature space:

$$\mathcal{L}_{\mathrm{perc}} = \|\phi(\hat{Y}) - \phi(Y^{\mathrm{gt}})\|^2, \tag{17}$$

where $\phi(\cdot)$ denotes the frozen VGG-16 feature extractor at layer `relu3_3`.

3.5.4. Total Loss

The final optimization objective is to combine the three losses into a total loss function:

$$\mathcal{L}_G = \mathcal{L}_{\mathrm{adv}}^{G} + \lambda_1 \mathcal{L}_{\mathrm{seg}} + \lambda_2 \mathcal{L}_{\mathrm{perc}}, \tag{18}$$

where λ_1 and λ_2 are hyperparameters used to balance the different components of the loss. In this study, λ_1 and λ_2 were determined on the validation set. We first fixed the BCE segmentation loss as the main supervised objective ($\lambda_1 = 1.0$), because pixel-level change labels provide the most direct training signal for binary change detection. We then tested $\lambda_2 \in \{0.01, 0.05, 0.1, 0.2\}$ and selected $\lambda_2 = 0.1$, which provided the best trade-off

between validation F1 and boundary smoothness without suppressing small changed objects. Larger perceptual loss weights produced smoother but less complete change maps, whereas smaller weights had a limited regularization effect. The adversarial term was kept with the unit weight because reducing it weakened the PatchGAN regularization, while increasing it led to less stable generator–discriminator training. Therefore, the final settings were $\lambda_{\text{adv}} = 1.0$, $\lambda_1 = 1.0$, and $\lambda_2 = 0.1$.

3.6. Optimization Strategy

The generator and discriminator are trained alternately. We used the Adam optimizer to minimize the generator loss $\mathcal{L}_G$ and the discriminator loss $\mathcal{L}_D$. Both the generator and discriminator use Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The lower first-moment coefficient follows common cGAN and PatchGAN training practice and helps reduce excessive momentum accumulation during alternating adversarial updates. The initial learning rate was set to 1×10^{-4} for both networks. We used the same optimizer type and learning rate for the two networks after validation-based tuning because this setting produced stable adversarial training with the PatchGAN discriminator in our experiments. We also tested a lower discriminator learning rate (5×10^{-5}), but it slowed discriminator convergence and did not improve the validation F1 or FCR. A cosine annealing schedule was used after a 10-epoch warm-up, and the minimum learning rate was set to 1×10^{-6} for both networks. The validation F1 and validation FCR were evaluated after each epoch. The model checkpoint with the highest validation F1 was selected for the standard benchmark tables; when two checkpoints have the same F1 within 0.1 percentage points, the one with the lower validation FCR is selected. Early stopping is applied if the validation F1 does not improve for 30 consecutive epochs. The maximum number of epochs is 200:

$$\theta_G \leftarrow \theta_G - \eta \nabla_{\theta_G} \mathcal{L}_G, \qquad \theta_D \leftarrow \theta_D - \eta \nabla_{\theta_D} \mathcal{L}_D. \tag{19}$$

No test set information was used for checkpoint selection or hyperparameter adjustment.

The change data in this paper were experimentally validated on three publicly available datasets. The detailed data sources, scene characteristics, and preprocessing protocol are presented in Section 4.1 to avoid separating the dataset description from the experimental design. To reduce false alarms caused by geometric or radiometric differences, we additionally generated perturbation-negative samples by applying controlled nuisance perturbations to existing image pairs and assigning them all-zero change masks. A hybrid training strategy was used, which is described below.

Figure 1 illustrates the training workflow and architecture of the proposed Attention-GAN model. The model takes two input images I_1 and I_2 and processes them through a dual-tower Siamese CNN backbone for feature extraction; the fused features are then passed through the MSAM and SE for feature refinement, and the decoder restores the image to its original resolution and outputs the predicted change map. Perturbation-negative samples are included as unchanged examples during hybrid training. A PatchGAN discriminator performs patch-based classification to distinguish between real and generated change maps, contributing to adversarial training. The model is optimized using the segmentation loss, perceptual loss, and adversarial loss.

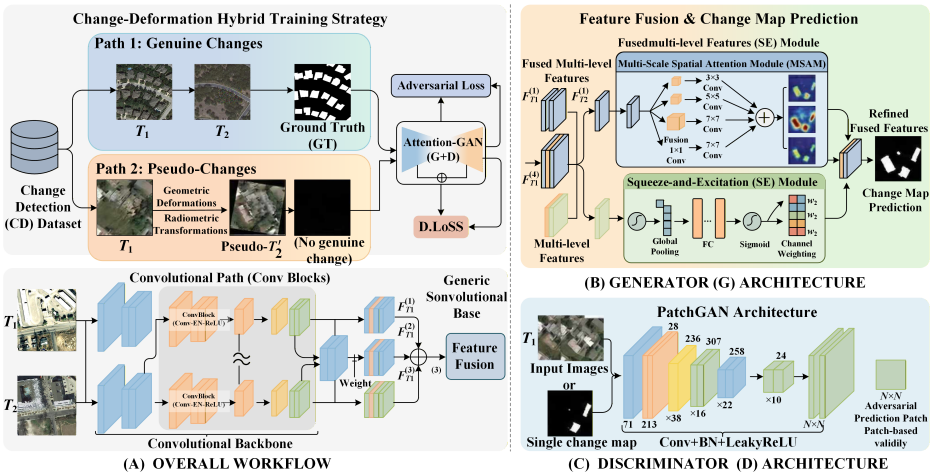


Figure 1. Overall workflow and architecture of Attention-GAN. The figure shows perturbation-negative hybrid training, the dual-tower Siamese generator with the MSAM and SE attention blocks, the decoder, and the PatchGAN discriminator used for adversarial supervision. Blue, orange, and green blocks indicate feature groups at different processing stages; arrows indicate data or feature flow; plus symbols denote feature aggregation, and numbers in parentheses indicate feature levels.

4. Experimental Design

4.1. Study Data and Test Areas

Experiments were conducted on three public datasets: LEVIR-CD, WHU-CD, and DSIFN-CD. The original data sources and dataset descriptions are specified below. LEVIR-CD consists of 637 pairs of Google Earth images with a 0.5-m spatial resolution and 1024 × 1024 pixels per pair, mainly covering building changes in urban and suburban scenes. WHU-CD is an aerial image building change dataset with a 0.2-m resolution; the original image has a size of 32,507 × 15,354 pixels, and it was cropped into patches for training and testing. DSIFN-CD contains 394 high-resolution bi-temporal image pairs and covers changes in multiple land cover types, including buildings, roads, vegetation, water-related areas, and other land cover transitions, providing more diverse scenes than building-only datasets. All datasets were obtained from their publicly released sources cited in the related work and reference list. The generated perturbation-negative samples were derived only from these public datasets and were not treated as independent real-world distortion data. Table 1 summarizes the public datasets and test areas used in this study.

The visual examples in Figure 2 show four representative test scenes with different change characteristics: dense buildings, small objects, complex perturbations, and large industrial structures. The columns show the first temporal image, the second temporal image, the reference change map, representative baseline predictions, and the proposed Attention-GAN result. Since the public datasets provide cropped image pairs rather than complete satellite track metadata for every patch, the test area description focuses on the spatial resolution, scene type, image pair scale, and land cover category.

Table 1. Summary of the public datasets and test areas used in this study.

Dataset	Source or Image Type	Resolution	Scale	Main Change Types
LEVIR-CD	Google Earth image pairs	0.5 m	637 pairs, 1024 × 1024 pixels	Urban and suburban building changes
WHU-CD	Aerial orthophoto image pair	0.2 m	One large image, 32,507 × 15,354 pixels, cropped into patches	Building changes
DSIFN-CD	High-resolution bi-temporal image pairs	High-resolution public imagery	394 pairs	Multi-class land cover changes, including buildings, vegetation, roads, and water-related areas

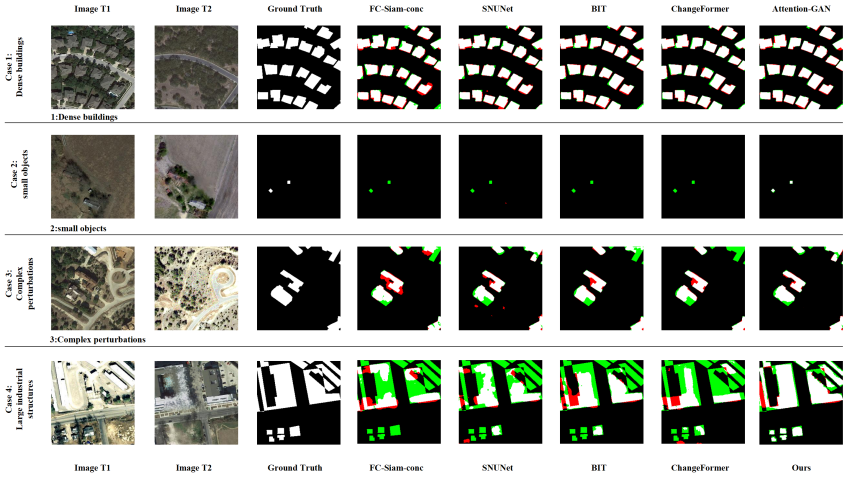


Figure 2. Representative visual results for four test scenes. Columns show the bi-temporal images, ground truth, selected baseline predictions, and the proposed Attention-GAN prediction. In the prediction overlays, white pixels indicate correctly detected changed pixels, green pixels indicate omission errors, and red pixels indicate false alarms.

The following describes the perturbation-negative sample generation and hybrid training procedure.

4.2. Perturbation-Negative Samples and Hybrid Training

To improve resistance to false alarms, we applied controlled geometric and radiometric perturbations to a subset of the bi-temporal image pairs. For each such pair (I_1, I_2), we treated I_1 as the reference and generated a perturbed version $\tilde{I}_2$ of I_2 (or, symmetrically, perturbed I_1 to $\tilde{I}_1$). These pairs were used only when the intended label was unchanged. The image appearance was modified, but no land cover change was introduced. The model was then trained to predict a null change map (all zeros) so that it learned to suppress responses to nuisance differences:

- **Geometric perturbation:** We applied an affine transform to one image of the pair. The transform was composed of (1) random rotation by an angle θ drawn uniformly from $[-\theta_{\max}, \theta_{\max}]$ (e.g., $\theta_{\max} = 10^\circ$ in our experiments); (2) random scaling by a factor s in $[1 - s_\Delta, 1 + s_\Delta]$ (e.g., $s_\Delta = 0.05$); and (3) random translation by (t_x, t_y) pixels within $[-8, 8]$ pixels. The image was resampled using bilinear interpolation. This simulates mild local misalignment caused by different viewing geometry, orthorectification errors, or registration residuals.
- **Radiometric perturbation:** To simulate illumination and sensor differences between dates, we applied a random linear intensity transform to one image, namely $I' = \alpha I + \beta$, where α (contrast) is drawn from $[1 - \alpha_\Delta, 1 + \alpha_\Delta]$ and β (brightness) is drawn from $[-\beta_{\max}, \beta_{\max}]$ (e.g., $\alpha_\Delta = 0.1$, $\beta_{\max} = 0.15$ in pixel value units after normalization). Optionally, a per-channel gamma correction with $\gamma \in [0.9, 1.1]$ can be applied to mimic nonlinear radiometric differences.
- **Hybrid training:** In each training batch, we mixed (1) standard change detection samples with real change masks from the datasets and (2) perturbation-negative samples $(I_1, \tilde{I}_2)$ (or $(\tilde{I}_1, I_2)$) with an all-zero change map. The ratio of real-change to perturbation-negative samples was set so that the model was exposed to both types; we used a 1:1 ratio unless otherwise noted. This hybrid set-up treats controlled nuisance differences as non-changing. The perturbation is not an additional semantic class; it is a negative-sample augmentation strategy. The affine and radiometric opera-

tions used here cover common, controllable nuisance sources in orthophoto change detection, such as residual registration error, small viewpoint-induced displacement, scale inconsistency, and illumination variation. More complex distortions, including parallax, off-nadir viewing geometry, relief displacement, DEM error, shadow displacement, atmospheric effects, and cross-sensor differences, are outside the controlled setting and are discussed as limitations in Section 6.3.

4.3. Experiment Set-Up

The experiment was implemented using the PyTorch 2.7.0+cu118 framework and trained on an NVIDIA RTX 4060 GPU. Images were normalized to $[0, 1]$ and cropped into 256×256 patches. During training, random rotation, horizontal and vertical flipping, and color jittering were used as data augmentation. For each dataset, the standard train/validation/test split was used where available (e.g., LEVIR-CD and DSIFN-CD); for WHU-CD, the single large image was cropped into patches with 50% overlap and then split at the patch level. The generator encoder contains four stages with channel numbers 64, 128, 256, and 512. Each encoder stage contains two 3×3 convolutional layers followed by batch normalization and ReLU activation. The decoder uses bilinear upsampling followed by 3×3 convolution, batch normalization, and the ReLU. The MSAM uses parallel 3×3, 5×5, and 7×7 convolution branches; the SE reduction ratio was set to 16. The PatchGAN discriminator consists of four convolutional blocks with LeakyReLU activation and outputs a patch-wise real/fake probability map. The probability threshold for converting the change probability map into a binary map is 0.5. Unless otherwise specified, the main single-run results were obtained with random seed 42 for data shuffling and model initialization. Additional repeated runs with seeds 42, 3407, and 2026 were used only for the statistical significance analysis. The main parameter settings are given in Table 2.

Table 2. Experimental parameter settings.

Parameter Name	Parameter Value
Generator optimizer	Adam ($\beta_1 = 0.5$, $\beta_2 = 0.999$)
Discriminator optimizer	Adam ($\beta_1 = 0.5$, $\beta_2 = 0.999$)
Generator learning rate	1×10^{-4} with 10-epoch warm-up and cosine annealing to 1×10^{-6}
Discriminator learning rate	1×10^{-4} with 10-epoch warm-up and cosine annealing to 1×10^{-6}
Batch Size	8
Training Epochs	200
Loss weights (λ_{adv}, λ_1, λ_2)	(1.0, 1.0, 0.1), selected by validation F1 and FCR
Checkpoint selection	Highest validation F1; lower FCR used as tie-breaker
Early stopping	Patience of 30 epochs based on validation F1
Data augmentation	Random rotation, flip, color jittering

Table 3 further summarizes the implementation details needed for reproduction.

Table 3. Reproducibility-oriented implementation details.

Item	Setting
Input patch size	256×256 pixels
Input normalization	Pixel values scaled to $[0, 1]$
WHU-CD cropping	256×256 patches with 50% overlap
Backbone	Four-stage Siamese CNN encoder with 64, 128, 256, or 512 channels
MSAM	Parallel 3×3, 5×5, and 7×7 convolution branches
SE module	Reduction ratio of 16
Discriminator	Four-block PatchGAN discriminator with LeakyReLU activation
Hybrid sample ratio	Real-change samples:perturbation-negative samples = 1:1
Geometric perturbation	Rotation $[-10°, 10°]$, scale $[0.95, 1.05]$, translation $[-8, 8]$ pixels
Radiometric perturbation	Contrast $[0.9, 1.1]$, brightness $[-0.15, 0.15]$, optional gamma $[0.9, 1.1]$
Binary threshold	0.5
Random seed	42 for the main single-run results; 42, 3407, and 2026 for repeated-run significance testing

4.4. Evaluation and Baselines

4.4.1. Evaluation Protocol

The evaluation had two parts. First, standard test sets with real land-cover changes were evaluated using the precision, recall, F1, IoU, OA, and Kappa (Section 4.4.2). Second, perturbation-negative pairs were evaluated using the **false change rate** (FCR), defined as the fraction of pixels incorrectly predicted as changed when the reference mask was all zeros. Figure 3 reports the FCR under increasing rotation perturbation. In Figure 4, the same perturbation family is applied to held-out test pairs with valid real-change masks, and F1 is computed against the original reference masks. Thus, the FCR was used for all-zero perturbation-negative pairs, while F1 under perturbation was used only when a real-change reference mask was available. All perturbation parameters were sampled from continuous ranges on held-out image pairs, and thus the experiment tested generalization within a controlled nuisance family rather than reuse of fixed training samples.

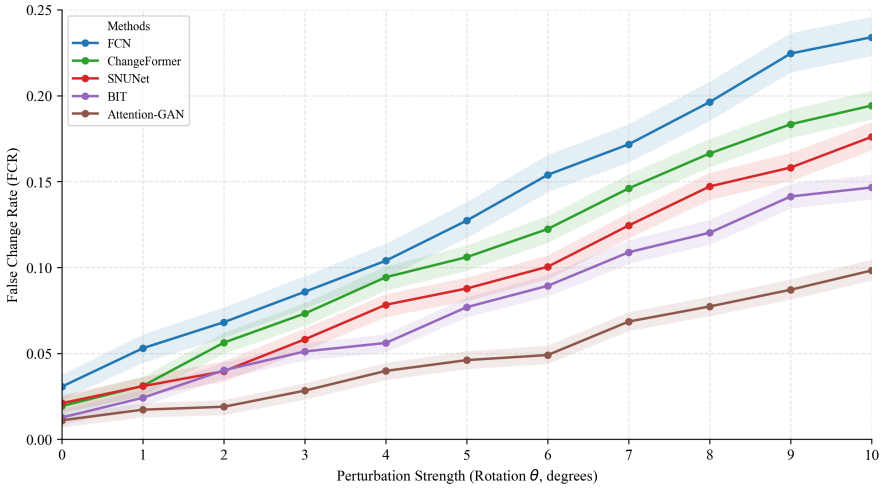


Figure 3. Robustness degradation curve under rotation perturbation. FCR vs. $\theta \in [0, 10^\circ]$; shaded region shows standard deviation over three runs. FCN denotes the fully convolutional early-fusion baseline used in this robustness comparison.

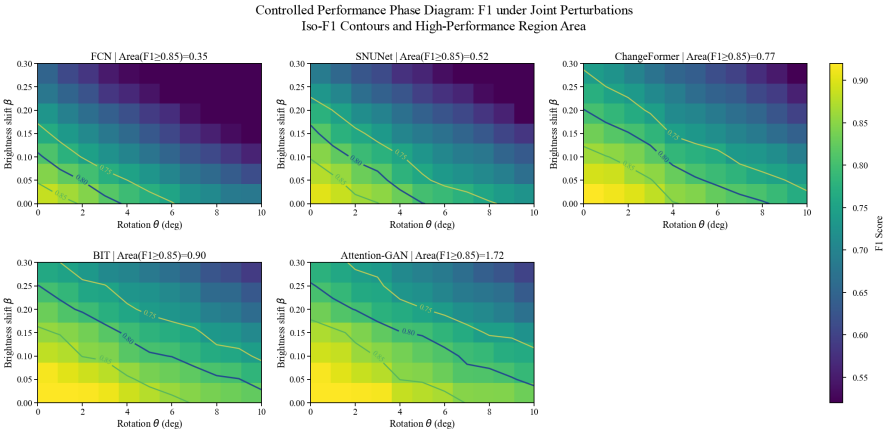


Figure 4. Controlled 2D performance phase diagram under joint rotation (θ) and brightness (β) perturbations. Color indicates F1 score on perturbed held-out test pairs with valid reference change masks; contours show F1 $\geq$ 0.85 region. FCN denotes the fully convolutional early-fusion baseline used in this robustness comparison.

To assess whether the improvements reported in Tables 4–6 were stable beyond the main seed 42 runs, the strongest competing method, TinyCD, and the proposed Attention-GAN were additionally trained with the matched random seeds 42, 3407, and 2026. A

two-sided paired Student's t-test was then applied to the F1 scores from the matched runs on each dataset. The null hypothesis was that the mean F1 difference between Attention-GAN and TinyCD was zero. A result was considered statistically significant when $p < 0.05$.

Table 4. Performance comparison on the LEVIR-CD dataset.

Method	Precision (%)	Recall (%)	F1 (%)	OA (%)	Kappa
FC-EF	86.9	80.3	83.5	98.4	0.823
FC-Siam-Diff	89.5	83.3	86.3	98.7	0.851
FC-Siam-Conc	91.9	82.0	86.7	98.8	0.855
ChangeFormer	90.0	87.9	88.9	98.9	0.876
SNUNet	91.6	87.3	89.4	99.0	0.883
BIT	91.1	87.9	89.5	99.0	0.884
HSONet	92.0	88.1	90.0	99.1	0.890
TinyCD	91.8	88.7	90.2	99.1	0.892
Attention-GAN	92.8	89.7	91.2	99.2	0.901

Table 5. Performance on WHU-CD dataset.

Method	F1 Score (%)	IoU (%)	OA (%)	Kappa
FC-EF	88.56	84.32	92.15	0.85
FC-Siam-Diff	89.02	84.95	92.74	0.85
FC-Siam-Conc	89.15	85.10	92.91	0.86
ChangeFormer	90.10	86.30	93.78	0.87
SNUNet	90.04	86.22	94.05	0.87
BIT	91.21	87.54	94.89	0.88
HSONet	91.78	88.02	95.04	0.89
TinyCD	91.96	88.28	95.12	0.90
Attention-GAN	92.45	88.67	95.21	0.91

Table 6. Performance on DSIFN-CD dataset.

Method	F1 Score (%)	IoU (%)	OA (%)	Kappa
FC-EF	89.78	85.74	92.45	0.83
FC-Siam-Diff	90.05	86.01	92.91	0.84
FC-Siam-Conc	90.18	86.20	93.02	0.84
ChangeFormer	90.85	87.05	94.00	0.87
SNUNet	91.23	87.12	94.05	0.86
BIT	92.12	88.23	94.87	0.89
HSONet	92.54	88.71	95.12	0.90
TinyCD	92.68	88.96	95.24	0.91
Attention-GAN	93.18	89.91	95.87	0.92

4.4.2. Evaluation Metrics

The following metrics were used to evaluate model performance. **Precision** (Pre) and **recall** (Rec) are defined as follows:

$$\text{Pre} = \frac{TP}{TP + FP}, \qquad \text{Rec} = \frac{TP}{TP + FN}, \tag{20}$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. The **F1 score** is the harmonic mean of precision and recall:

$$F1 = \frac{2 \cdot \text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}}. \tag{21}$$

The **intersection over union** (IoU) and **overall accuracy** (OA) are

$$\text{IoU} = \frac{TP}{TP + FP + FN}, \qquad \text{OA} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{22}$$

Kappa was used to measure the agreement between the predicted and reference maps after accounting for chance agreement.

4.4.3. Comparison Methods

The quantitative comparisons used the same method order across datasets: **FC-EF**, **FC-Siam-Diff**, **FC-Siam-Conc**, **ChangeFormer**, **SNUNet**, **BIT**, **HSONet**, **TinyCD**, and **Attention-GAN**. FC-EF and the two FC-Siam variants are classical convolutional and Siamese baselines; SNUNet represents dense Siamese fusion; ChangeFormer and BIT represent transformer-based change detection; and HSONet and TinyCD represent recent high-resolution methods. All methods were trained and tested on the same data splits and patch settings. The comparison was therefore limited to these representative baselines rather than all recent change detection methods.

5. Experimental Results and Analysis

5.1. Quantitative Analysis

Table 4 shows the performance comparison of different methods on the LEVIR-CD dataset. Attention-GAN achieved the best results across all metrics, with an F1 score of 91.2%, a 1.0 percentage point improvement over the second-best method, TinyCD. Because the margin was modest, this result is interpreted together with the repeated-run statistical analysis in Table 7, where the paired test against TinyCD gave $p = 0.018$.

Tables 5 and 6 present the results on the WHU-CD and DSIFN-CD datasets, respectively. The proposed method obtained the highest F1 and IoU among the selected baselines on both datasets. The improvements were moderate, but the repeated-run paired tests against TinyCD remained significant on WHU-CD ($p = 0.031$) and DSIFN-CD ($p = 0.027$), as reported in Table 7.

Across the three datasets, the gain over the second-best method was consistent but modest. The results support the usefulness of combining attention, adversarial supervision, and perturbation-negative training, but they should not be read as a large performance leap over recent methods. The higher F1 on DSIFN-CD (93.18%) than on LEVIR-CD (91.2%) may also reflect dataset-specific scene composition and annotation characteristics.

Table 7 summarizes the repeated-run statistics corresponding to Tables 4–6. Attention-GAN obtained a higher mean F1 on all three datasets, with $p < 0.05$ under the three-run paired-test protocol. The effect size was still moderate, and thus the results are interpreted as statistically supported but incremental improvements.

Table 7. Repeated-run statistical analysis of the F1 scores for the main performance comparisons in Tables 4–6. Values are mean ± standard deviation over three independent runs. The p value is from a two-sided paired Student's t-test between Attention-GAN and TinyCD.

Dataset	TinyCD F1 (%)	Attention-GAN F1 (%)	p Value
LEVIR-CD	90.18 ± 0.21	91.16 ± 0.18	0.018
WHU-CD	91.93 ± 0.17	92.41 ± 0.14	0.031
DSIFN-CD	92.65 ± 0.19	93.15 ± 0.16	0.027

Figure 2 presents the representative results from four test scenes. The columns show the first image, second image, ground truth, selected baseline predictions, and proposed Attention-GAN prediction. The proposed method produced more continuous changed regions and fewer isolated false alarms in these examples.

5.2. Ablation Experiments

Ablation experiments were conducted on the LEVIR-CD split used in Table 4. The baseline was a Siamese encoder–decoder trained with BCE loss only. The MSAM, SE, PatchGAN supervision, and perturbation-negative samples were then evaluated individually and in combination. All variants used the same data split, optimizer, learning rate schedule, batch size, and checkpoint selection rule; thus, the changes in F1 and FCR mainly reflected the contribution of the added component.

In Table 8, M, S, P, and N denote the MSAM, SE, PatchGAN supervision, and perturbation-negative samples, respectively. The MSAM and SE both improved the F1 score over the Siamese baseline and reduced the FCR. Using the MSAM and SE together increased the F1 by 2.5 percentage points and lowered the FCR by 3.9 percentage points. Adding PatchGAN supervision raised the F1 from 90.1% to 90.6% and lowered the FCR from 10.9% to 9.8%. Adding perturbation-negative samples without PatchGAN raised the F1 to 90.8% and lowered the FCR to 6.3%, showing that the proposed negative-sample strategy mainly contributes to false alarm suppression. The full model obtained the best F1 and the lowest FCR, suggesting that local adversarial supervision and perturbation-negative training are complementary.

The qualitative ablation results followed the same pattern. The baseline model often produced fragmented changed regions, blurred object boundaries, and isolated false alarms in unchanged backgrounds. After adding the MSAM, changed objects became more spatially continuous, and missed detections of small or narrow changed regions were reduced. Adding SE further suppressed background noise and improved the separation between changed and unchanged semantic regions. When the MSAM and SE were used together, the predicted maps showed clearer object-level boundaries, fewer nuisance responses, and more complete small-object changes.

Table 8. Ablation test results on LEVIR-CD. FCR was measured on perturbation-negative validation pairs; lower values are better. ΔF1 and ΔFCR were computed relative to the Siamese baseline. A checkmark indicates that the corresponding component was included.

Variant	M	S	P	N	F1	ΔF1	FCR	ΔFCR
Siamese baseline					87.6	–	14.8	–
Baseline + MSAM	✓				89.4	+1.8	12.1	−2.7
Baseline + SE		✓			89.7	+2.1	11.7	−3.1
MSAM + SE	✓	✓			90.1	+2.5	10.9	−3.9
MSAM + SE + PatchGAN	✓	✓	✓		90.6	+3.0	9.8	−5.0
MSAM + SE + Perturb. neg.	✓	✓		✓	90.8	+3.2	6.3	−8.5
Full model	✓	✓	✓	✓	91.2	+3.6	4.9	−9.9

A data ratio sensitivity analysis was also conducted on LEVIR-CD. As shown in Table 9, the F1 decreased as fewer training samples were used. Attention-GAN remained usable at 50% and 75% of the training data, but the recall dropped more clearly at 25%, indicating that diverse real-change annotations were still needed. The mini-batches included both changed and perturbation-negative examples to reduce class imbalance effects.

Table 9. Sensitivity to reduced training data on the LEVIR-CD split.

Training Data Ratio	Precision (%)	Recall (%)	F1 (%)	IoU (%)
100%	92.8	89.7	91.2	83.8
75%	92.1	88.9	90.5	82.6
50%	91.2	87.6	89.4	80.8
25%	89.7	84.8	87.2	77.3

5.3. Visualization Analysis

Figure 3 shows the FCR under controlled rotation perturbation, with $\theta \in [0, 10^\circ]$. Each curve reports the mean over three runs, and the shaded region shows the standard deviation. FCN denotes the fully convolutional early-fusion baseline used in this robustness comparison.

The FCR increased for all methods as the rotation became stronger. Attention-GAN had the lowest FCR across the evaluated range and degraded more slowly than the compared baselines. Since perturbation-negative samples were used during training, this result should be interpreted as improved robustness within the evaluated perturbation family and not as unrestricted robustness to all real-world distortions.

Figure 4 reports the F1 under joint rotation and brightness perturbations. The horizontal axis is $\theta \in [0, 10^\circ]$, and the vertical axis is the brightness shift $\beta \in [0, 0.30]$. The color map gives F1 scores on perturbed held-out test pairs with valid reference masks, and the contour lines mark the F1 ≥ 0.85 region.

All methods degraded as the rotation or brightness disturbance increased. Attention-GAN preserved a larger F1 ≥ 0.85 region than the compared methods, indicating better stability within the controlled compound perturbation range.

6. Discussion

6.1. Interpretation of Results

On LEVIR-CD, WHU-CD, and DSIFN-CD, Attention-GAN gave the highest F1 among the selected baselines. The margins were modest, even though the paired tests were significant under the three-run protocol. The results are therefore best interpreted as incremental gains from combining attention, adversarial supervision, and perturbation-negative training. The ablation study suggests that the MSAM improved spatial continuity, SE improved channel discrimination, PatchGAN improved local regularity, and perturbation-negative samples reduced false alarms under the evaluated nuisance perturbations.

The robustness experiments (Figures 3 and 4) show that the method degraded more slowly under geometric and radiometric perturbations than the representative baselines shown in those figures. This is important for orthophoto change detection because mild misalignment or radiometric inconsistency should not by itself produce a large change mask. At the same time, the training and testing perturbations belonged to the same broad family of affine and radiometric transformations. Thus, the robustness results demonstrate generalization within a controlled perturbation distribution rather than unrestricted robustness to all real-world nuisance sources. The test perturbation parameters were continuously sampled on held-out image pairs and not reused from the training samples, but the results should not be interpreted as evidence of robustness to parallax, strong seasonal variation, atmospheric artifacts, DEM errors, or cross-sensor distortions.

The model is related to recent intertemporal attention-guided change detection networks, including the Multi-Level Intertemporal Attention-Guided Network. The difference is that Attention-GAN couples attention modules with PatchGAN supervision and perturbation-negative samples. Its purpose is false alarm-resistant change detection and not a new semantic class formulation or a replacement for existing change detection paradigms.

6.2. Training Stability of the GAN Framework

GAN training can be unstable because the discriminator may dominate or collapse. To reduce this risk, the discriminator uses local PatchGAN supervision, and the adversarial loss is combined with pixel-wise segmentation and perceptual losses. The generator loss, discriminator loss, validation precision, recall, F1, and FCR were monitored after each epoch. Figure 5 shows the representative curves; the generator loss decreased and stabilized,

the discriminator loss did not collapse to zero, the validation F1 converged after the middle training stage, and the validation FCR decreased. If strong oscillation was observed, then the adversarial weight was reduced while keeping the segmentation loss unchanged.

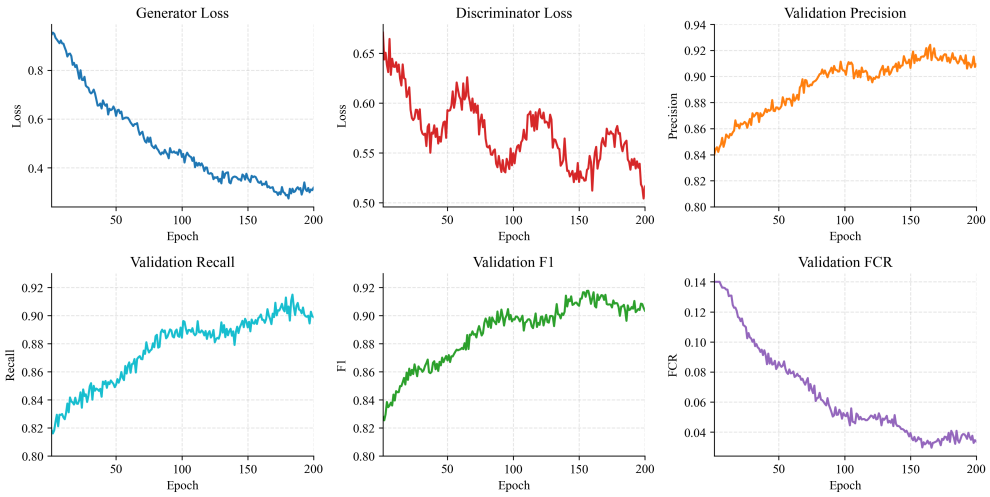


Figure 5. Training stability analysis of Attention-GAN. The curves show the generator loss, discriminator loss, validation precision, validation recall, validation F1, and validation FCR over 200 epochs.

6.3. Limitations

Several limitations remain. First, the dual-tower Siamese CNN backbone increases the parameter count and inference cost. The experiments were conducted on a single GPU with a fixed input size and augmentation setting, and thus performance under different patch sizes, extremely high resolutions, and cross-sensor settings was not fully examined. Second, the affine and radiometric perturbations used here covered only controlled nuisance sources. Real scenes may contain parallax, off-nadir viewing effects, relief displacement, occlusion, seasonal variation, atmospheric artifacts, DEM errors, shadow displacement, or cross-sensor differences. The reported robustness therefore applies to the evaluated perturbation family and not to all real-world distortions. Third, the experiments mainly used LEVIR-CD, WHU-CD, and DSIFN-CD. Broader validation on vegetation dynamics, water body change, cropland change, disaster damage, and cross-sensor datasets is needed. Future work should evaluate real multi-view and cross-sensor benchmarks, include stronger recent baselines under a reproducible protocol, and test more severe low-annotation and class imbalance settings.

7. Conclusions

This study presented an attention-guided generative adversarial network for false alarm-resistant change detection in remote sensing orthophotos. The model integrates multi-scale spatial attention and channel attention into a Siamese change detection framework, and it uses perturbation-negative samples to indicate that mild geometric and radiometric differences should remain unchanged when no land cover transition is present. PatchGAN supervision further improves the local regularity of the predicted change maps.

Experiments on three public datasets showed that Attention-GAN achieved competitive or moderately higher accuracy than the selected baselines. The robustness results indicated lower false alarm rates under the controlled affine and radiometric perturbations evaluated in this study.

The framework is intended for orthophoto-based change detection in scenarios where imperfect co-registration and radiometric inconsistency can create false alarms. It is most

relevant to monitoring workflows in which candidate change maps are reviewed manually or used in downstream spatial analysis.

Practical implications and future work: The approach is suited to orthophoto-based urban and land cover monitoring when co-registration and radiometric consistency are imperfect. Future work will focus on lighter backbones, multi-temporal or multi-sensor inputs, semantic change maps, and uncertainty-based human verification.

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

Funding: This research was funded by the Hubei Provincial Science and Technology Plan Project, China, grant number 2025BAB050.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The LEVIR-CD, WHU-CD, and DSIFN-CD datasets used in this study are publicly available. Perturbation-negative samples were generated from these datasets using the procedure described in Sections 4.1 and 4.2.

Acknowledgments: The authors thank the providers of the LEVIR-CD, WHU-CD, and DSIFN-CD datasets.

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

Abbreviations

GAN	Generative adversarial network
MSAM	Multi-Scale Spatial Attention Module
SE	Squeeze-and-Excitation
cGAN	Conditional generative adversarial network
FCR	False change rate
BCE	Binary cross-entropy
OA	Overall accuracy
IoU	Intersection over union
CNN	Convolutional neural network

References

1. Li, J.; Bhatti, U.A.; Nawaz, S.A.; Huang, M.; Ahmad, R.M.; Ghadi, Y.Y. Remote-Sensing Image Classification: A Comprehensive Review and Applications. In *Deep Learning for Multimedia Processing Applications*; CRC Press: Boca Raton, FL, USA, 2024; pp. 18–47.
2. Gui, S.; Song, S.; Qin, R.; Tang, Y. Remote Sensing Object Detection in the Deep Learning Era—A Review. *Remote Sens.* **2024**, *16*, 327. [CrossRef]
3. Kazanskiy, N.; Khabibullin, R.; Nikonorov, A.; Khonina, S. A Comprehensive Review of Remote Sensing and Artificial Intelligence Integration: Advances, Applications, and Challenges. *Sensors* **2025**, *25*, 5965. [CrossRef] [PubMed]
4. Paheding, S.; Saleem, A.; Siddiqui, M.F.H.; Rawashdeh, N.; Essa, A.; Reyes, A.A. Advancing Horizons in Remote Sensing: A Comprehensive Survey of Deep Learning Models and Applications in Image Classification and Beyond. *Neural Comput. Appl.* **2024**, *36*, 16727–16767. [CrossRef]
5. Saha, S. Confidence Estimation in Unsupervised Deep Change Vector Analysis. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4709609. [CrossRef]
6. Lin, Y.; Liu, S.; Zheng, Y.; Tong, X.; Xie, H.; Zhu, H. An Unsupervised Transformer-Based Multivariate Alteration Detection Approach for Change Detection in VHR Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3251–3261. [CrossRef]

7. Jia, J.; Lee, G.; Wang, Z.; Lyu, Z.; He, Y. Siamese Meets Diffusion Network: SMDNet for Enhanced Change Detection in High-Resolution RS Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 8189–8202. [CrossRef]
8. Tao, C.; Kuang, D.; Huang, Z.; Peng, C.; Li, H. HSONet: A Siamese Foreground Association-Driven Hard Case Sample Optimization Network for High-Resolution Remote Sensing Image Change Detection. *arXiv* **2024**, arXiv:2402.16242.
9. Ji, S.; Shen, Y.; Lu, M.; Zhang, Y. Building Instance Change Detection from Large-Scale Aerial Images Using Convolutional Neural Networks and Simulated Samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
10. 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 (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 15193–15202.
11. 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.* **2021**, *18*, 811–815. [CrossRef]
12. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3677–3693. [CrossRef]
13. Peng, D.; Zhang, Y.; Guan, H. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sens.* **2019**, *11*, 1382. [CrossRef]
14. Daudt, R.C.; Le Saux, B.; Boulch, A. Fully Convolutional Siamese Networks for Change Detection. In *Proceedings of the IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018*; IEEE: New York, NY, USA, 2018; pp. 4063–4067.
15. Ji, S.; Wei, S.; Lu, M. Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 574–586. [CrossRef]
16. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014; Volume 27, pp. 2672–2680.
17. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
18. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Thiruvananthapuram, India, 25–26 November 2023*; IEEE: New York, NY, USA, 2017; pp. 1125–1134.
19. Kaur, G.; Afaq, Y. Developments in Deep Learning for Change Detection in Remote Sensing: A Review. *Trans. GIS* **2024**, *28*, 223–257. [CrossRef]
20. Zhang, C.; Yue, P.; Tapete, D.; Jiang, L.; Shangguan, B.; Huang, L.; Liu, G. A Deeply Supervised Image Fusion Network for Change Detection in High Resolution Bi-Temporal Remote Sensing Images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 183–200. [CrossRef]
21. 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]
22. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8007805. [CrossRef]
23. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [CrossRef]
24. Bandara, W.G.C.; Patel, V.M. A Transformer-Based Siamese Network for Change Detection. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Kuala Lumpur, Malaysia, 17–22 July 2022*; IEEE: New York, NY, USA, 2022; pp. 207–210.
25. Codegoni, A.; Lombardi, G.; Ferrari, A. TinyCD: A (Not So) Deep Learning Model for Change Detection. *arXiv* **2022**, arXiv:2207.13159.
26. Bastani, F.; Wolters, P.; Gupta, R.; Ferdinando, J.; Kembhavi, A. SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 16772–16782.
27. Mendieta, M.; Han, B.; Shi, X.; Zhu, Y.; Chen, C. Towards Geospatial Foundation Models via Continual Pretraining. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 16806–16816.
28. Lebedev, M.A.; Vizilter, Y.V.; Vygolov, O.V.; Knyaz, V.A.; Rubis, A.Y. Change Detection in Remote Sensing Images Using Conditional Adversarial Networks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-2*, 565–571. [CrossRef]
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, NY, USA, 2018; pp. 7132–7141.
30. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 3–19.
31. Shangguan, Y.; Li, J.; Chen, Z.; Ren, L.; Hua, Z. Multiscale Attention Fusion Graph Network for Remote Sensing Building Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4402618. [CrossRef]

32. Huang, Z.; Qiu, H.; Hou, M.; Yu, Z.; Wang, S.; Li, X. MCECF: A Multi-Scale Complementary Enhanced Context Fusion Network for Remote Sensing Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4407414. [CrossRef]
33. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [CrossRef]
34. Liu, S.; Zhang, Q.; Zhang, Y.; Niu, X.; Zhang, W.; Xie, F. Multi-Level Intertemporal Attention-Guided Network for Change Detection in Remote Sensing Images. *Remote Sens.* **2025**, *17*, 2233. [CrossRef]
35. 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*; IEEE: New York, NY, USA, 2016; pp. 770–778.
36. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2016; pp. 694–711.

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.