Article

Edge-Bound Change Detection in Multisource Remote Sensing Images

Zhijuan Su [1], Gang Wan [1,*], Wenhua Zhang [2,*], Zhanji Wei [1], Yitian Wu [1], Jia Liu [1], Yutong Jia [1], Dianwei Cong [1] and Lihuan Yuan [1]

[1] School of Space Information, Space Engineering University, Beijing 101407, China; hgfgh@stu.cuz.edu.cn (Z.S.); weizhanji@163.com (Z.W.); ytwu@whu.edu.cn (Y.W.); xssf@stu.cuz.edu.cn (J.L.); hgdjyt@163.com (Y.J.); congdianwei@sina.com (D.C.); yuanlihuan10@alumni.nudt.edu.cn (L.Y.)

[2] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

[*] Correspondence: casper_51@163.com (G.W.); whzhang@njust.edu.cn (W.Z.)

Abstract: Detecting changes in multisource heterogeneous images is a great challenge for unsupervised change detection methods. Image-translation-based methods, which transform two images to be homogeneous for comparison, have become a mainstream approach. However, most of them primarily rely on information from unchanged regions, resulting in networks that cannot fully capture the connection between two heterogeneous representations. Moreover, the lack of a priori information and sufficient training data makes the training vulnerable to the interference of changed pixels. In this paper, we propose an edge-oriented generative adversarial network (EO-GAN) for change detection that indirectly translates images using edge information, which serves as a core and stable link between heterogeneous representations. The EO-GAN is composed of an edge extraction network and a reconstructive network. During the training process, we ensure that the edges extracted from heterogeneous images are as similar as possible through supplemented data based on superpixel segmentation. Experimental results on both heterogeneous and homogeneous datasets demonstrate the effectiveness of our proposed method.

Keywords: heterogeneous image change detection; generative adversarial network; edge orientation; image translation

1. Introduction

Change detection (CD) is the inference task of recognizing variations between two images of the same region obtained at different times [1–10]. It is used in a wide variety of applications, such as urban planning, land management, agricultural survey, and natural disaster monitoring [11–13].

Plenty of methods for change detection problems have been proposed. Additionally, deep learning has been extensively used. Farahani et al. [14] proposed a domain adaptation method based on an autoencoder, which fuses features of synthetic aperture radar (SAR) and optical images to achieve better accuracy by measuring the complementary information. Ma et al. [15] proposed an approach for SAR image change detection based on multigrained cascade forest (gcForest) and multiscale fusion. Different sizes of image blocks are fed into gcForest, greatly improving the accuracy of detection. Qu et al. [16] proposed a dual-domain network (DDNet). The spatial and frequency domains are combined to improve classification performance. They developed a multiregion convolution module in the spatial domain to improve the input image patches and used DCT transformation and a gating mechanism to acquire frequency information in the frequency domain. Many approaches introduce a generative adversarial network for its excellent feature representation abilities. Zhao et al. [17] proposed to exploit invariant feature representations by the use of a GAN combined with a metric learning strategy and introduced a seasonal transition term to

exclude pseudo changes. In [18], Hou et al. designed a GAN with a dual-branch architecture as a generator to explore the distribution.

The above methods are for homogeneous images; i.e., multitemporal images are acquired via the same type of sensors. However, homogeneous images are not always available in many applications, such as disaster evaluation, where the available homogeneous data may be fragmented or not exhaustive for urgent events. As a consequence, there are more requirements for change detection in heterogeneous images that are acquired by different types of sensors, such as synthetic aperture radar (SAR) and optical images [19]. However, it is also a great challenge for change detection methods to deal with heterogeneous images, especially for unsupervised ones. It is not feasible for a direct comparison like most of the unsupervised methods. Optical sensors record the intensity of ground objects in the visible and infrared parts of the electromagnetic spectrum. They cover the earth widely, but the image quality is vulnerable to influence by the atmosphere and illumination conditions. Meanwhile, SAR sensors measure radar backscatter. They can penetrate the clouds and are immune to the effects of sunlight conditions because they collect information about ground objects in a microwave frequency band, whereas the speckle noise in SAR images is intractable [20,21]. Although different sensors give inconsistent feature representations of the same ground object, they measure unique physical qualities and the information acquired by different sensors can be complementary.

Existing unsupervised change detection methods for heterogeneous images can be divided into feature-transformation-based and image-translation-based ones. Feature-transformation-based ones compare multitemporal images in a common feature space via feature transformation operators. For example, the symmetric convolutional coupling network (SCCN) proposed in [22] generates the common feature space via a network with a convolution layer and several coupling layers, and an objective function is defined to train the network in an unsupervised manner. Image-translation-based ones convert one image of multitemporal images from one domain to another domain in order to make the multitemporal images comparable. For example, Niu et al. [23] proposed a framework that includes a translation network based on a generative adversarial network (GAN) for translating an optical image into that similar to a radar image. Liu et al. [24] proposed a method based on homogeneous pixel transformation (HTP). HTP transfers one image into another image's space with some unchanged pixels being selected as supervised knowledge. Li et al. [25] developed a spatially self-paced convolutional network (SSPCN). They obtain pseudo labels using a classification-based method and provide each sample a weight about easiness. The network learns simple samples first, and then progresses to more complex and detailed samples. Jiang et al. [26] proposed a model termed deep homogeneous feature fusion (DHFF) by introducing the idea of image style transfer (IST), which separates semantic content and style to prevent a semantic content from being corrupted.

These aforementioned approaches are mostly based on the idea of homogeneous transformation, that is, transforming two heterogeneous images into a more consistent feature space or translating one image into the style of another. Then they can be compared directly. However, they also have some drawbacks, with the main problem focusing on the learning of the mapping relationship between two feature spaces. Some samples need to be selected for the learning. Generally, they are unchanged pixels. However, in an unsupervised manner, the unchanged pixels are usually approximated. For example, a preclassification method is used in [25], pseudo unchanged labels are defined in [22], and all samples are used in [23] with the hypothesis that changed pixels are much less than unchanged pixels. Actually, it is quite tricky to directly explore the relationship between two images located in the original observation space, while for edge information, first, it is easy to extract, and second, although it may be affected by the difference of image properties or disturbed by factors such as noise, its representation is still mainly determined by the content of the ground objects in the image, which makes the edge maps of the two images located in a more consistent space.

As a consequence, in this paper, we explore the relationship between heterogeneous images via edge information and propose the edge-oriented GAN (EO-GAN) to translate one image into the representation style of another image. First, edge information is easy to extract, and second, although it may be affected by the difference of image properties or disturbed by factors such as noise, its representation is still mainly determined by the content of the ground objects in the image, which makes the edge maps of the two images not influenced by the representation capability of heterogeneous sensors. The EO-GAN is composed of an edge extraction network and a reconstruction network. The extraction network consists of several residual blocks that learn to extract edges from preprocessed images, with pseudo labels provided by the Canny operator. Then, a GAN is built to learn to reconstruct the optical image from the edges. Moreover, we use a super-pixel-segmentation-based approach to preprocess the input image by adding artificial changes to force the network to learn to capture the connection between edge changes and actual content changes. This will help to reconstruct the SAR image's unique content of the change region from its edges in the end. At the same time, a series of preprocessing operations are implemented to make the edges of the optical image used for training more consistent with the edges of the final input SAR image. Then the network can learn the consistent edges between SAR and optical images.

The contributions of this paper are summarized as follows: (1) We propose a new unsupervised change detection framework called EO-GAN for heterogeneous images by translating the heterogeneous images into homogeneous ones via edge information. (2) We design a network that consists of an edge extraction network and a reconstruction network to learn the consistent edges between heterogeneous images and reconstruct the image of homogeneous features. (3) Superpixel segmentation and other preprocessing methods are used to avoid a learning discrepancy of edges. Experiments demonstrate the effectiveness of image translation and change detection.

The remainder of this paper is organized as follows: Section 2 discusses the theoretical foundation and related work. Section 3 details the proposed method and its implementation details. The experimental results on five datasets are presented in Section 4. Section 5 provides the conclusion of the paper.

2. Related Work and Preliminaries

2.1. Edge Detection

There are various classical edge detection operators [27,28] in the traditional image processing field, which detect abrupt changes in gray level, color, texture, etc., by measuring the first-order derivative or second-order derivative. Meanwhile, a large number of deep-learning-based edge detection methods have been proposed in recent years. He et al. [29] proposed a bidirectional cascade network (BDCN) utilizing several parallel-dilated convolution-to-yield multiscale features, improving the accuracy of edge detection for objects at different scales. Xie et al. [30] proposed a convolutional-network-based edge detection system that uses a skip-layer architecture to fuse multiscale feature maps. In this paper, we use a network to capture the edges in order to reduce the influence of noise.

2.2. Super Pixel

Superpixel algorithms are used for grouping coherent pixels into new atomic regions that can replace the original pixel grid [31]. It is an increasingly popular image preprocessing technique used in many computer vision applications, such as image segmentation, object recognition, object tracking, classification, and 3D reconstruction [32]. Here, we introduce SLIC [31], a simple and classic superpixel algorithm. SLIC is based on clustering; its only parameter is the number of superpixels k. It randomly initializes cluster centers on a regular grid spaced S pixels apart, where $S = \sqrt{N/k}$, and in the assignment step, for each pixel i in a $2S \times 2S$ region around C_k, computes the distance between C_k and i. In the update step, new cluster centers are computed. The assignment and update steps are repeated iteratively until the error converges. In our paper, superpixels are used as a

preprocessing step to divide the image into atomic blocks based on features before we add distortions, maintaining the integrity of the images' content.

2.3. Image-to-Image Translation Network with Conditional GAN

A generative adversarial network (GAN) [33] was proposed by I. Goodfellow et al. in 2014, which is a quite remarkable work that it novelly constructs two adversarial models: a generator (G) to generate fake data and a discriminator (D) to discriminate whether the data are real or fake. By training them adversatively, a balance can be finally reached that the fake data generated by G are close to the real ones and D's ability is strong enough to recognize real and fake data.

If we provide some extra information y, the GAN can be extended to a conditional version (cGAN [34]); its objective function can be defined as follows:

$$\min_{G} \max_{D} V(D, G) = E_{x \sim p_{data}(x)}[log(D(x|y))]$$
$$+ E_{z \sim p_z(z)}[log(1 - D(z|y))], \tag{1}$$

where y and the noise z are combined and sent to the generator as input. Then the discriminator takes y as a condition and analyzes the probability that a sample came from the training data rather than G.

Based on the cGAN, an image-to-image translation network was proposed in [35] to learn a mapping from one image distribution $p_{data(x)}$ to another distribution $p_{data(y)}$. The generator takes x and the noise z as input. Then in the discriminator, x needs to be concatenated with the input $G(x)$ or y as extra information. It also uses L1 loss, pushing the generated image to be close to the ground truth output. Its objective function can be defined as follows:

$$arg \min_{G} \max_{D} TN(D, G) = \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$
$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_y[logD(y)] + \mathbb{E}_{x,z}[log(1 - D(G(x, z)))]. \tag{2}$$
$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{x,y,z}[||y - G(x, z)||_1].$$

In this paper, we use the cGAN to translate the image of SAR to that of optical. Here, the generator G is used as the reconstruction network and x is the edge map extracted by the edge extraction network. y is the ground truth of the generator, i.e., the optical image. z is the noise map, and here, we use the multiscale pepper noise.

3. Methodology

The flowchart of change detection using the EO-GAN is illustrated in Figure 1. With the multitemporal images I_1 and I_2, which are acquired in times T_1 and T_2, respectively, an edge extraction network is used to extract the edge of the two images. A denoised edge map of I_1 via the two edge maps is then derived. Then the reconstruction network is used to reconstruct the image with the representation style of I_2 from the denoised edge map of I_1. Finally, the reconstructed image and I_2 are compared to generate the difference image.

To accurately detect the changes, it is crucial to train the two networks, i.e., the edge extraction network and the reconstruction network. Therefore, we design an adversarial training method based on the cGAN, as shown in Figure 2. To sufficiently train the networks with only the two images, we first construct a training set by data augmentation. As shown in Figure 2, we distort the two images with random patches extracted via superpixel segmentation and image twisting. Then following the cGAN with the edge map as the latent representation, the objective in Equation (2) is first constructed. To extract the edge information in latent representations, the Canny operator is used to extract the edges as the reference labels of the edge extraction network. Next, we detail the edge extraction, image reconstruction, and edge denoising operator.
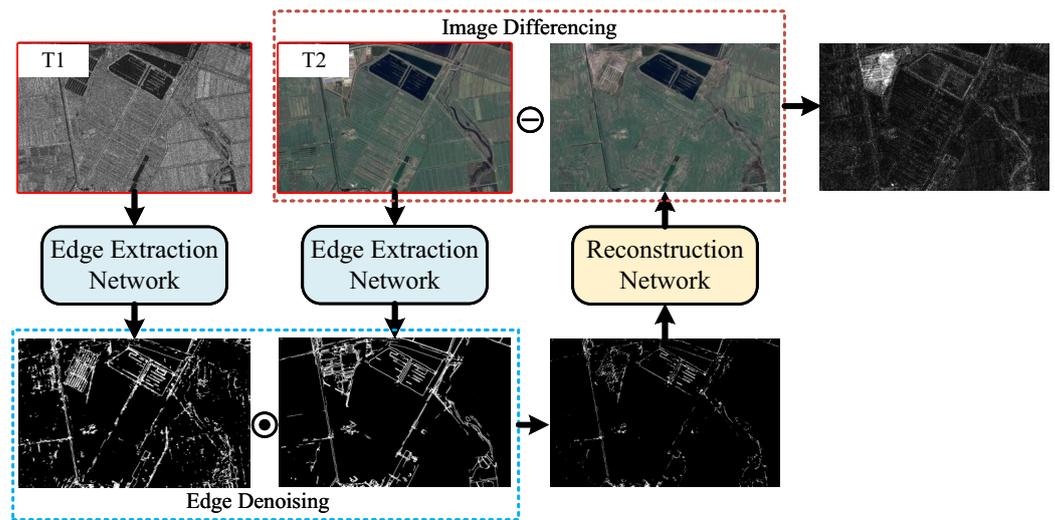
Figure 1. Change detection flowchart of EO-GAN, which is composed of edge extraction network and reconstruction network.
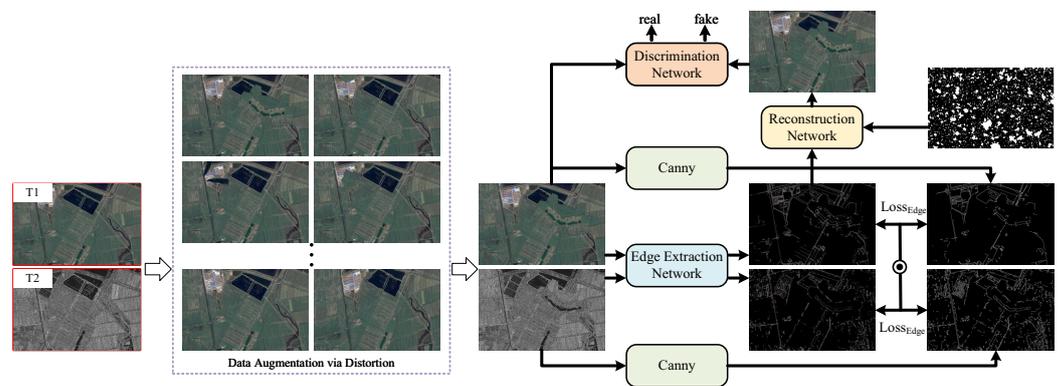

Figure 2. Training process of EO-GAN by using the two multitemporal images.

3.1. Edge Extraction

The basic idea is to extract the common image feature from the two heterogeneous images as the basis for the subsequent reconstruction training. This type of feature needs to contain the major information about ground objects while being insensitive to the differences in image properties. Since the color features and texture features of heterogeneous images obviously differ greatly, we chose to extract the shape features of the images, more specifically, the edge information.

No matter how different the properties of the heterogeneous images are, as long as the objects in a certain area have not changed, the edges extracted from the two images in that area will also have great similarity and overlap partially, and if changes have occurred, then the edges at the corresponding places will definitely be very different.

We first try to obtain the edges of the image using the Canny operator, which was proposed by John F. Canny in 1986 and is now recognized as the optimal edge detection algorithm in the industry. It locates the derivative maximum by the first-order differential of the Gaussian function with direction, and it can achieve a good balance between noise suppression and detection accuracy. However, we find that it is susceptible to noise in complex scenes, especially for SAR images with speckle noise. At the same time, the Canny operator is directional, and its response in some directions is sometimes not obvious, leading to inaccurate results. Therefore, we use a simple network with several residual blocks to extract the edges.

Two heterogeneous images are used as training data, and the edge images generated via the Canny operator are obtained as pseudo labels after denoising. For optical images,

a Gaussian filter is used, and for SAR images, we choose a Lee filter [36], which can significantly suppress the multiplicative speckle noise. We also rotate the input image during the training process to make the network acquire isotropic characteristics like the Laplacian operator in order to capture edges in any direction more accurately. We consider that the pixels of an edge only make up a small fraction of all pixels. We use the batch-balanced contrastive loss [37], which is an improved contrastive loss. It counts the number of positive and negative samples in ground truth as batch weight prior to alleviate the class imbalance problem. Then the edge extraction loss is defined:

$$
\begin{aligned}
\text{Loss}_{\text{Edge}} = \sum_{i,j=0}^{N} \Big[&\frac{1}{2}\frac{1}{n_{ne}}\big(1 - gt_{i,j}\big)d_{i,j}^2 \\
&+ \frac{1}{2}\frac{1}{n_e}gt_{i,j}\max\big(0, m - d_{i,j}\big)^2\Big],
\end{aligned}
\tag{3}
$$

where gt is the label map; 1 represents an edge pixel; d is the output of the edge detection network; and n_e, n_{ne} are the numbers of the edge pixels and nonedge pixels, respectively. To better reconstruct the contents in the image I_1 with the representation property of I_2, we have also to force the edge of I_1 to be close to that of I_2 as much as possible by simultaneously taking both gts of SAR and optical images as the label. To demonstrate the effectiveness of the edge extraction network, we illustrate the edges in Figure 3. Edges extracted by the Canny operator contain many noise, and the variance between those of optical and SAR images is large. After training the edge extraction network, the edges of the two types of images are more similar and can represent the main objects in the two types of images, respectively.
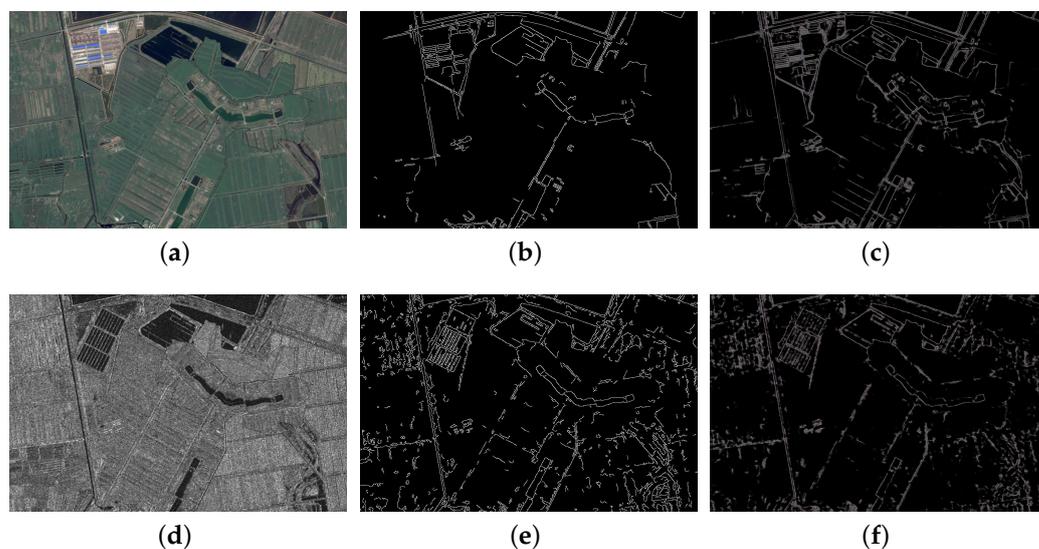

Figure 3. Illustration of the learned edge extraction network: (**a**) optical image, (**b**) edges of the optical image generated by the Canny operator, (**c**) edges of the optical image generated by the edge extraction network, (**d**) SAR image, (**e**) edges of the SAR image generated by the Canny operator, and (**f**) edges of the SAR image generated by the edge extraction network.

3.2. Reconstruction Network

After training the edge-extract network, we utilize an image-to-image translation network, pix2pix [38], for reconstructing optical images from edge images. For two heterogeneous images, we generally choose the optical image as the reference data for the reconstruction network since it contains more information and less noise. Essentially, we have only one sample for training the reconstruction network, which is the optical image and its edges obtained using the edge extraction network. Therefore, preprocessing is

necessary to meet our final need for change detection. Based on such an idea, the two images to be detected often have only part of the region that has changed. Taking the image I_2 as an example, first, we change some areas of it and obtain another image I'_2 with distortions. After that, with the edge information of I'_2 obtained by the edge extraction network as the input and the edge map generated via the Canny operator as the label, we can train a reconstruction network to reconstruct I'_2 from its edges. Assume that the image I_1 has a corresponding image $I'_{1->2}$ in the optical feature space with exactly the same content, which is the image we wish to obtain by transforming I_1 into the optical feature space. We can find that $I'_{1->2}$ and I'_2 correspond in characteristics, both with some regions changed compared with I_2. Since the edge of I_1 is also the edge of $I'_{1->2}$, by feeding the edge information of I_1 to the reconstruction network, it is feasible to reconstruct $I'_{1->2}$ in the optical feature space with the same content as I_1. That is equivalent to indirectly transforming I_1 into the optical feature space, which can then be directly compared with I_2.

In [39], a large number of unpaired optical and SAR images are used to pretrain a CycleGAN structured transformation network, where two generators can capture the transformation relationships in the feature space of optical and SAR images from the rich pretraining data. In this paper, the unsupervised approach is adopted completely, and no additional training data are required. The network does not directly perform feature transformation from SAR to optical, but rather performs feature transformation by extracting edge information from SAR images and reconstructing optical images from the edge information. In [23], Niu et al. also used a cGAN, but the translation network is trained with a pair of patches in two heterogeneous input images. However, with such a training method, the patches of changed regions may mislead the translation process. In changed regions, the contents are different in the heterogeneous images, while the objective of learning is to transform them as the same. Using such training data will certainly interfere with the final translation effect. In our proposed method, we use the strategy of artificially added changes to the input image when training the reconstruction network so that the feature content of the input image is changed. Then the edge information obtained through the edge extraction network is also changed, and the edge information and the changed image are used as a pair of input and label to train the reconstruction network. In this way, not only the training data are augmented, but also the reconstruction network is trained in a targeted manner, in which the changes of edge information are integrated into the learning of changes in the reconstructed image.

Specifically, we use the whole image instead of patches as input, and the two multitemporal images are the only training data we need, which makes data augmentation particularly important. In addition to the commonly used image rotation, we choose to add distortions by making global and local changes to the image. The global changes are achieved by distorting the whole image with different degrees of grid. It is intended for situations where the type of ground object is not changed, but only its shape and boundaries change. The implementation of local change is more complicated; first of all, the image is segmented into superpixels via the SLIC algorithm. The number of segmented blocks is chosen randomly in the set interval. The pixels in each superpixel after segmentation have a higher probability of belonging to the same ground object. Then one or several superpixels are randomly selected and taken out together with their neighboring superpixels as the region to be changed. These areas are distorted, rotated, scaled, and shifted with a certain probability to cover the original image, as shown in Figure 2. Some examples of distorted images, extracted edges, and reconstructed images are illustrated in Figure 4, where twisted images and artificial changes are shown. With the data augmentation, the reconstruction network can reconstruct the images from edges well.
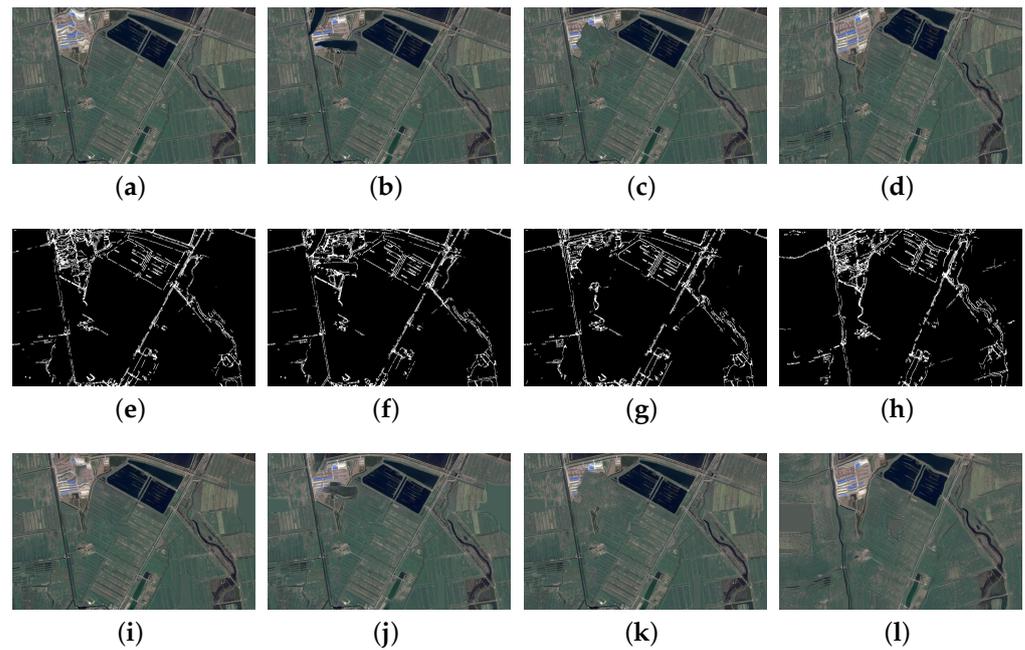
Figure 4. Distorted images, extracted edge maps, and reconstructed images from the edge maps: (**a–d**) distorted images, (**e–h**) extracted edges, and (**i–l**) reconstructed images.

3.3. Edge Denoising

Although only one image is actually needed for reconstruction, as shown in Figure 2, we still input another image I_1 to learn more consistent edges and features. After the distortion operation identical to that of I_2, its edge information E_1 is also extracted through the edge extraction network. After that, the common edge E' can be obtained by taking the intersection of E_2 and E_1, i.e., $E' = E_1 \bigcap E_2$. We use a simple iterative algorithm to complement E' with E_2 or E_1 as targets and derive the complemented images E'_2 and E'_1. For example, with E_2 as the target, E'_2 is initialized to E'. In each iteration, for each position in E_2 with value 1, if there exists any pixel with value 1 in its neighborhood in E'_2, we set the value of that position in E'_2 to 1:

$$E'_2(i,j) = \begin{cases} 1, & E_2(i,j) = 1 \,\&\, \sum_{(k,l) \in \Omega_{(i,j)}} E'_2(i,j) > 0 \\ 0, & otherwise, \end{cases} \tag{4}$$

where $\Omega_{(i,j)}$ denotes the neighborhood of the pixel position of (i,j). It can be easily implemented by using a convolution kernel with the size of the neighborhood. If it is set to 3×3, the points that get restored in each iteration must be adjacent to the edges that already exist in E'_2. If it is larger than 3×3, some points that are not connected to the edges in E'_2 may also be recovered. Most of the edges in E' are common overlapping edges in the unchanged region, and the rest belong to the changed region, where the edges in E_2 and E_1 can only overlap a small fraction. By using this iterative algorithm, the incomplete edge of the changed region can be restored to its complete state in E_2. In the meantime, in the operation of taking the intersection filtering out most of the isolated noise, and unless the noise point is connected to an existing edge in E'_2, they will not be recovered. In the same way, we can also obtain E'_1, which is the final input of the reconstruction network during change detection. The complementary edges with different sizes of kernels are shown in Figure 5.

It is also observed that the speckle noise of radar images leads to missing and intermittent edges, while the edges of optical images are generally more coherent. Therefore, we add some additional noise to E'_2. We take pepper noise as the basis, generate random

pepper noise at different scales, and overlap them to mask E_2' to obtain the E_2'', which is the actual input for the reconstruction network during training. Figure 6 shows the edge of SAR, E_2, E_2', and E_2''. By adding the noise, the generated edges from the optical image are more close to that from the SAR image.
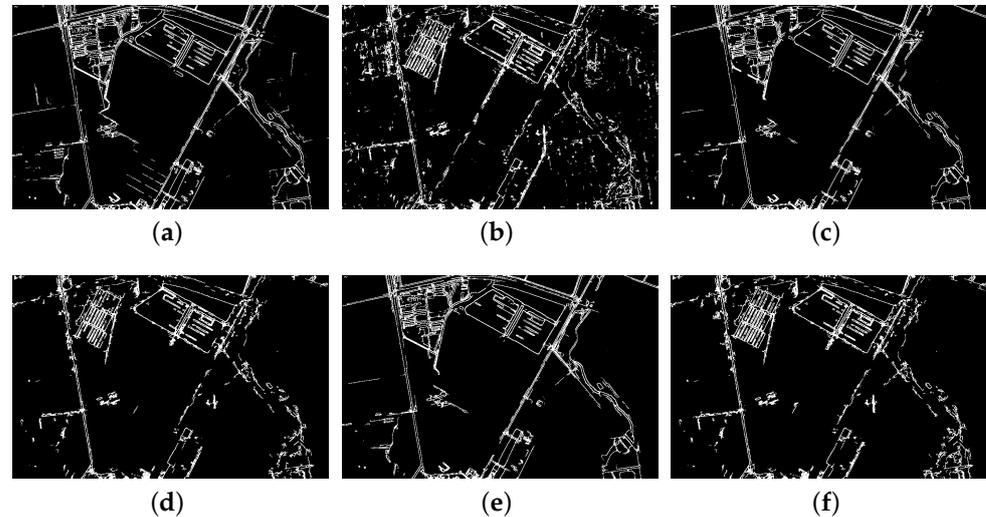

Figure 5. Complementary edges: (**a**) optical image, (**b**) SAR image, (**c**) optical image with kernel size 3, (**d**) SAR image with kernel size 3, (**e**) optical image with kernel size 5, and (**f**) SAR image with kernel size 5.
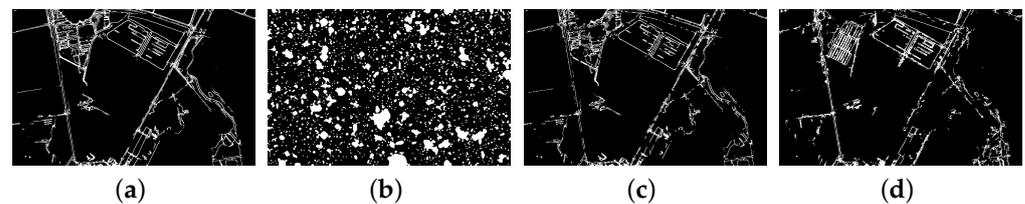
Figure 6. Noise in edge generation: (**a**) optical image, (**b**) multiscale pepper noise, (**c**) optical image with noise, and (**d**) SAR image.

4. Experimental Study

We use five datasets to evaluate the proposed EO-GAN, as shown in Figures 7–11. The first dataset consists of one optical image and one SAR image with a size of 291 × 343 pixels, as shown in Figure 7a,b, respectively. The second dataset is composed of one RGB optical image and one SAR image with a size of 548 × 340 pixels, as shown in Figure 8a,b, respectively. These two datasets both cover a section of the Yellow River. The SAR images were captured by Radarsat-2 in June 2008. The optical image of the first dataset captured in September 2010 was obtained from Google Earth, and the optical image of the second dataset acquired in May 2020 was obtained from satellite images of HERE Maps. These two datasets show the changes of the Yellow River bank caused by the scouring of the river channels. The actual changed regions are shown in Figures 7c and 8c.

The third dataset consists of two RGB optical images with the same size of 680 × 540 pixels. It includes some plants of several construction and machinery companies. The two images were acquired in May 2021 and September 2017, as shown in Figure 9a,b. The changed area corresponds to several new buildings, as shown in Figure 9c. The fourth dataset also consists of two RGB optical images like the third dataset. The two images have the same size of 736 × 1140, covering a piece of a suburb area of Guangzhou City in China. The first image was acquired in July 2017, and the second image was acquired in November 2013, as shown in Figure 10a,b. The reference image is shown in Figure 10c. This dataset is chosen from Google Earth, which was constructed by Peng et al. [38]. Although the two datasets

are homogeneous, their two images are affected by different lights, climates, and seasons, as well as other factors, and have different feature representations.

The fifth dataset also consists of one RGB optical image and one SAR image, as shown in Figure 11a,b, respectively. It was taken at the Shuguang Village of Dongying City in China, including farmlands and some factory buildings. The optical image was acquired in September 2012, and the SAR image was acquired in June 2008. Both images have a size of 921 × 593 pixels. The reference image shown in Figure 11c indicates the change of buildings over the years.

We use several criteria to evaluate our method, including areas under the ROC curve (AUC), false positive (FP), false negative (FN), overall error (OE), classification accuracy (CA), and kappa coefficient (KC). We choose the SCCN [22], cGAN [23], and HTP [24] as comparison methods.
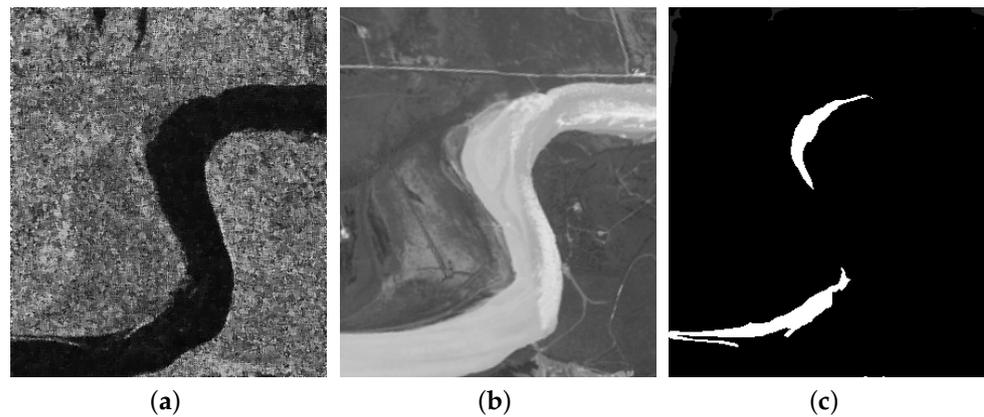

(**a**)　　　　　　　(**b**)　　　　　　　(**c**)

Figure 7. YR_1 dataset that shows the change of part of the Yellow River in China: (**a**) optical image, (**b**) SAR image, and (**c**) reference image.
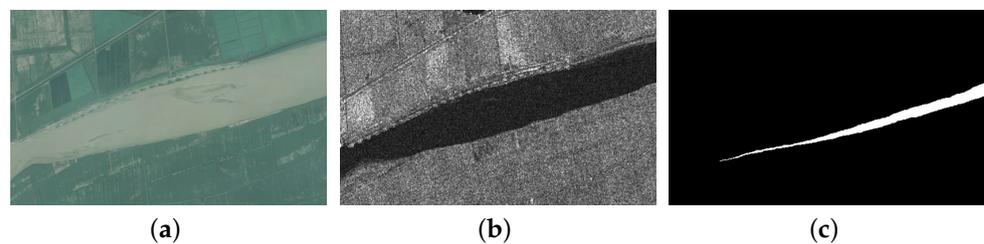

(**a**)　　　　　　　(**b**)　　　　　　　(**c**)

Figure 8. YR_2 dataset that shows the change of part of the Yellow River in China: (**a**) optical image, (**b**) SAR image, and (**c**) reference image.
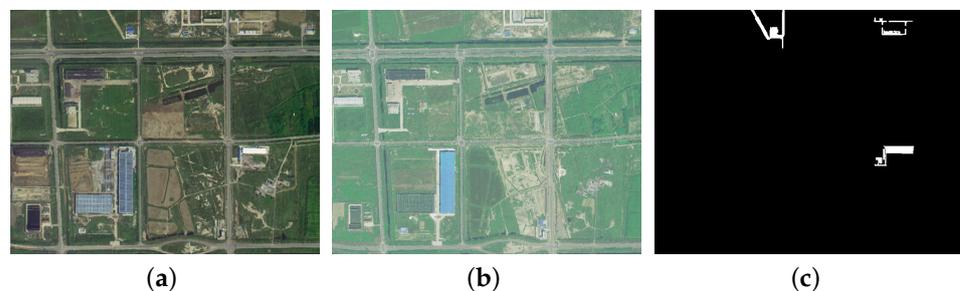

(**a**)　　　　　　　(**b**)　　　　　　　(**c**)

Figure 9. Dongying dataset that covers an area in the Tangtou Village of Dongying City in China: (**a**) image acquired in May 2021, (**b**) image acquired in September 2017, and (**c**) reference image.
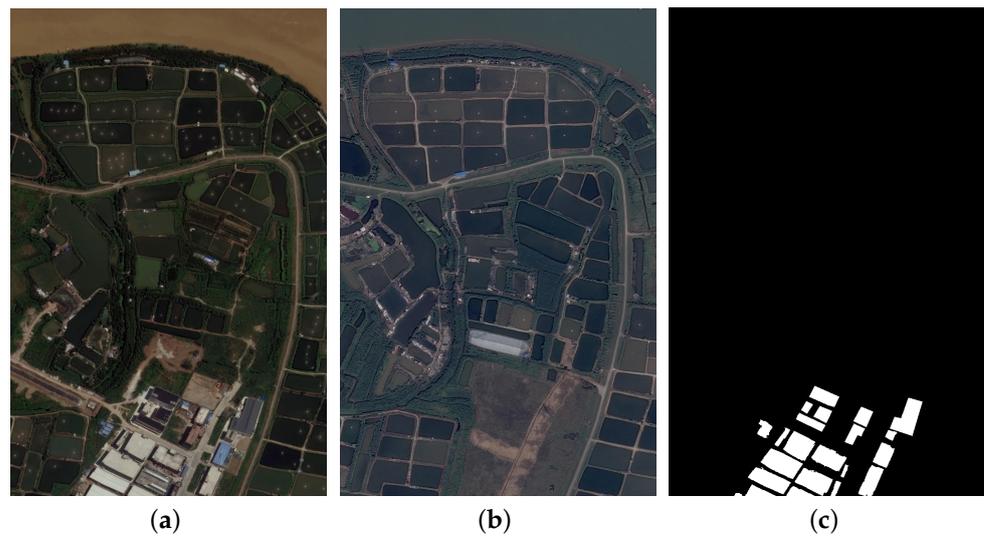
Figure 10. Guangzhou dataset that covers a piece of a suburb area of Guangzhou City in China: (**a**) image acquired in July 2017, (**b**) image acquired in November 2013, and (**c**) reference image.
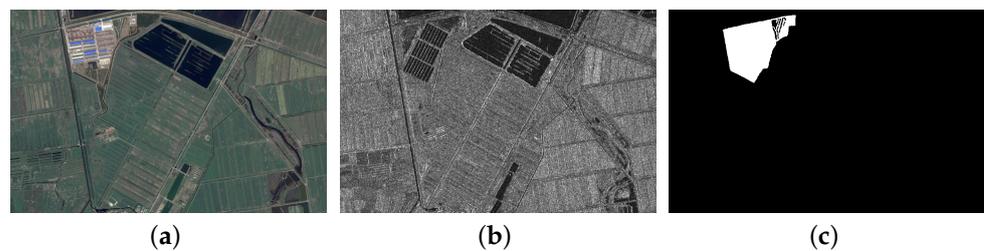

Figure 11. Shuguang dataset that was taken at the Shuguang Village in Dongying City of China: (**a**) optical image, (**b**) radar image, and (**c**) reference image.

4.1. Experiments on Yellow River Datasets

The difference images generated by compared and proposed methods on the two Yellow River datasets are shown in Figures 12 and 13. The two datasets are with heterogeneous images. The cGAN and HTP have difficulty recognizing the changed regions due to the influence of changed regions during training. The SCCN uses pseudo labels to mark the changed region. The proposed method uses the edges as the link between the two types of images, and edge denoising operators are proposed to improve the consistency. The proposed method can better restrain the unchanged regions. The final change detection results are also shown in Figures 12 and 13. Most of the unchanged regions are avoided by the proposed method, and the changed regions can be accurately detected. The quantitative evaluations on the two datasets are listed in Tables 1 and 2. Based on the criteria, the proposed method achieves the best result among the compared methods, which demonstrates the effectiveness of the proposed method on heterogeneous images.

Table 1. Evaluation metrics for the different methods experimented on the YR_1 dataset.

Methods	AUC	FP	FN	OE	CA	KC
SCCN	0.9688	1060	1235	2295	0.9770	0.6154
cGAN	0.9267	1652	1284	2936	0.9706	0.5466
HTP	0.9526	2356	838	3194	0.9680	0.5771
Proposed	0.9714	1048	937	1985	0.9801	0.6816

(**a**)   (**b**)   (**c**)   (**d**)

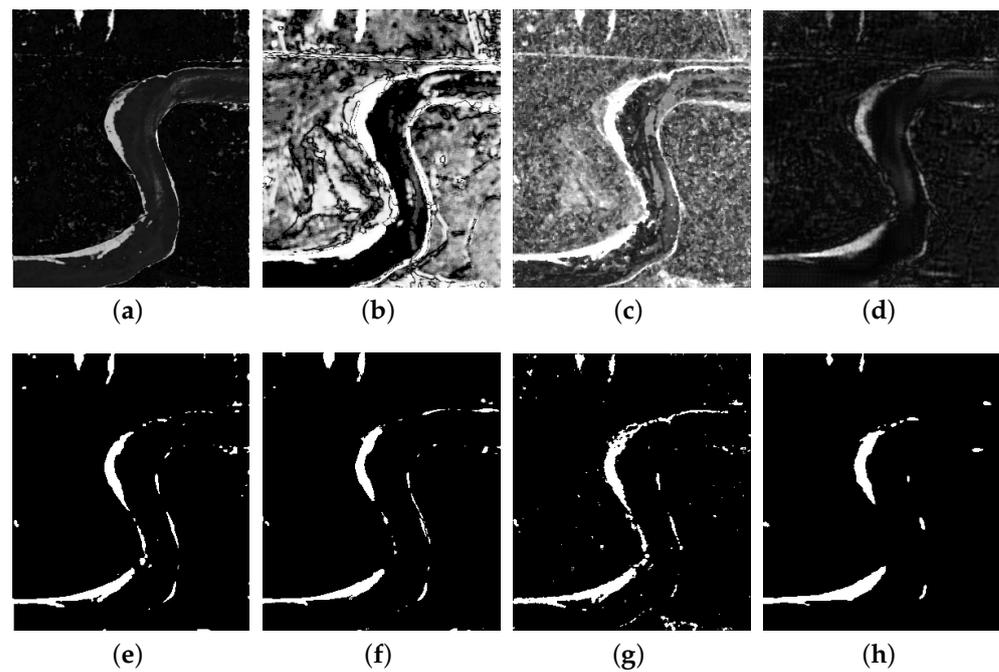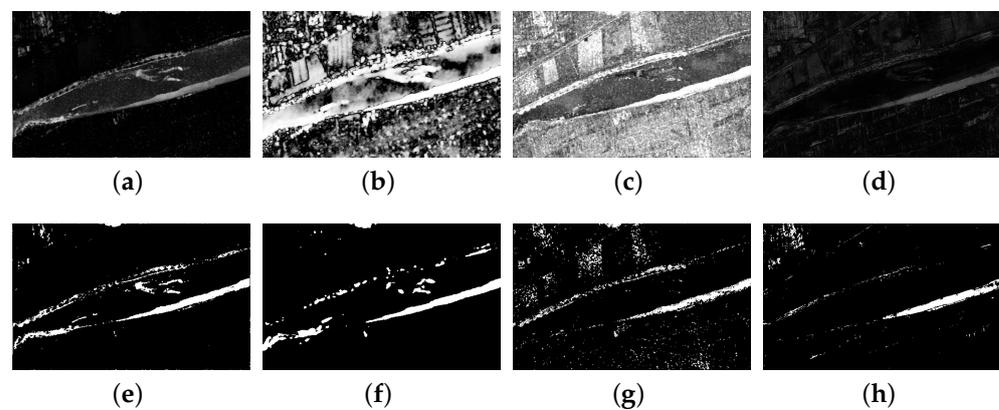(**e**)   (**f**)   (**g**)   (**h**)

Figure 12. Difference images and change detection results of the compared methods on the YR_1 dataset: (**a**) difference image of SCCN, (**b**) difference image of cGAN, (**c**) difference image of HTP, (**d**) difference image of the proposed method, (**e**) result of SCCN, (**f**) result of cGAN, (**g**) result of HTP, and (**h**) result of the proposed method.

(**a**)   (**b**)   (**c**)   (**d**)

(**e**)   (**f**)   (**g**)   (**h**)

Figure 13. Difference images and change detection results of the compared methods on the YR_2 dataset: (**a**) difference image of SCCN, (**b**) difference image of cGAN, (**c**) difference image of HTP, (**d**) difference image of the proposed method, (**e**) result of SCCN, (**f**) result of cGAN, (**g**) result of HTP, and (**h**) result of the proposed method.

Table 2. Evaluation metrics for the different methods experimented on the YR_2 dataset.

Methods	AUC	FP	FN	OE	CA	KC
SCCN	0.9404	6408	1537	7945	0.9574	0.4494
cGAN	0.9577	5223	807	6030	0.9676	0.5693
HTP	0.9263	6930	1969	8899	0.9522	0.3869
Proposed	0.9837	2287	1345	3632	0.9805	0.6610

4.2. Experiments on Dongying and Guangzhou Datasets

In theory, change detection methods for heterogeneous images are compatible with homogeneous images. The multitemporal images in the two datasets are both optical images. However, there are many unimportant changes, such as change of seasons, which can be avoided by the methods for heterogeneous images. The difference images and change detection results on the two datasets are shown in Figures 14 and 15. Similarly, the proposed method is able to restrain the irrelevant changes and highlight the most critical changes. Even though the cGAN and HTP can generate the same changed regions, the background objects are also highlighted. The quantitative evaluations on the two datasets are listed in Tables 3 and 4. Based on the criteria, the proposed method achieves the best result among the compared methods, which demonstrates the effectiveness of the proposed method in homogeneous images.
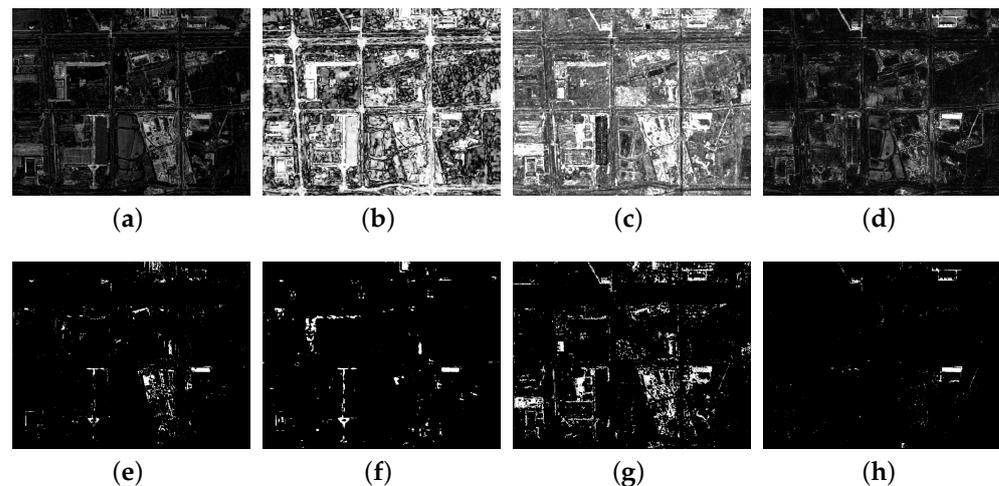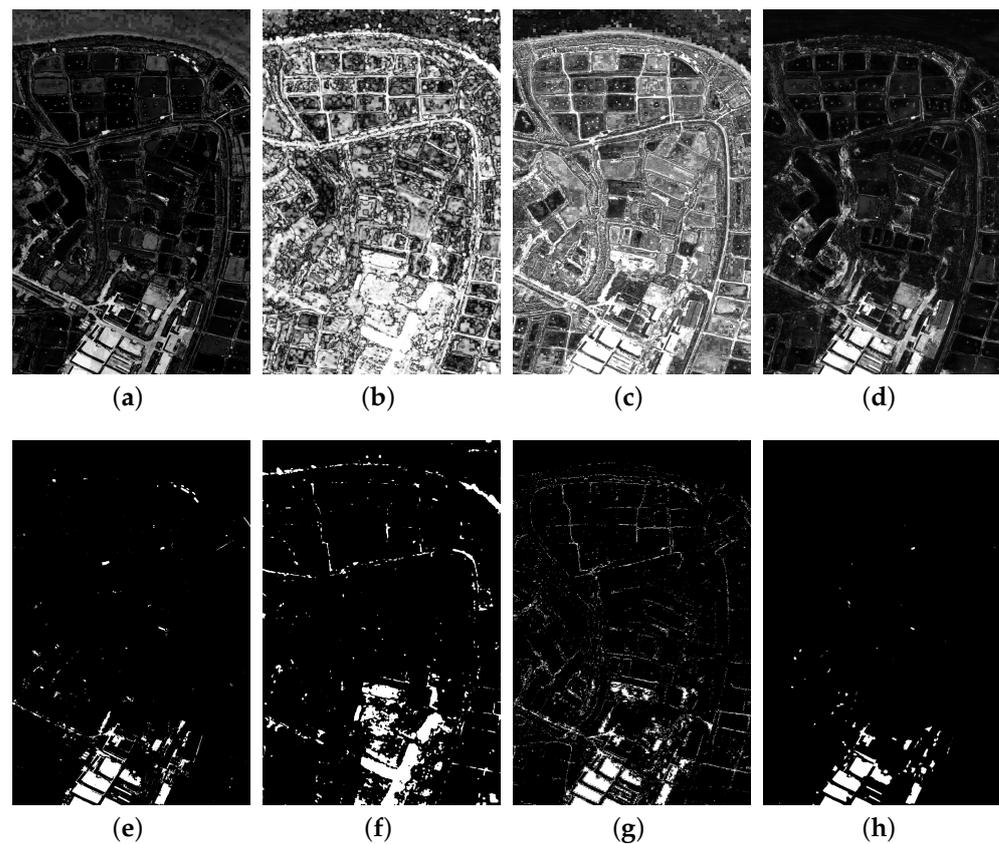

Figure 14. Difference images and change detection results of the compared methods on the Dongying dataset: (**a**) difference image of SCCN, (**b**) difference image of cGAN, (**c**) difference image of HTP, (**d**) difference image of the proposed method, (**e**) result of SCCN, (**f**) result of cGAN, (**g**) result of HTP, and (**h**) result of the proposed method.

Table 3. Evaluation metrics for the different methods on the Dongying dataset.

Methods	AUC	FP	FN	OE	CA	KC
SCCN	0.8254	5895	2710	8605	0.9923	0.1776
cGAN	0.8149	5807	2901	8708	0.9763	0.1455
HTP	0.8966	20,901	1610	22,511	0.9387	0.1423
Proposed	0.9494	748	2080	2828	0.9923	0.5316

Table 4. Evaluation metrics for the different methods on the Guangzhou dataset.

Methods	AUC	FP	FN	OE	CA	KC
SCCN	0.8337	9574	18,190	27,764	0.9669	0.5233
cGAN	0.7160	45,673	26,883	72,556	0.9135	0.1299
HTP	0.7961	26,710	23,905	18,981	0.9455	0.3761
Proposed	0.9187	4580	19,325	18,981	0.9715	0.5456

Figure 15. Difference images and change detection results of the compared methods on the Guangzhou dataset: (**a**) difference image of SCCN, (**b**) difference image of cGAN, (**c**) difference image of HTP, (**d**) difference image of the proposed method, (**e**) result of SCCN, (**f**) result of cGAN, (**g**) result of HTP, and (**h**) result of the proposed method.

4.3. Experiments on Shuguang Dataset

The Shuguang dataset contains a SAR and an optical image, and there are many types of ground objects, such as lake, farmland, building, and river. The difference images and the change detection results are shown in Figure 16. All the compared methods can generate the changed region, but the proposed method better restrains the impact of the background. Moreover, with edge denoising, there is much less noise in the results of the proposed method. The quantitative evaluation on the Shuguang dataset is listed in Table 5. Based on the criteria, the proposed method achieves the best result among the compared methods, which demonstrates the effectiveness of the proposed method for complex scenarios.

Table 5. Evaluation metrics for the different methods on the Yellow River dataset.

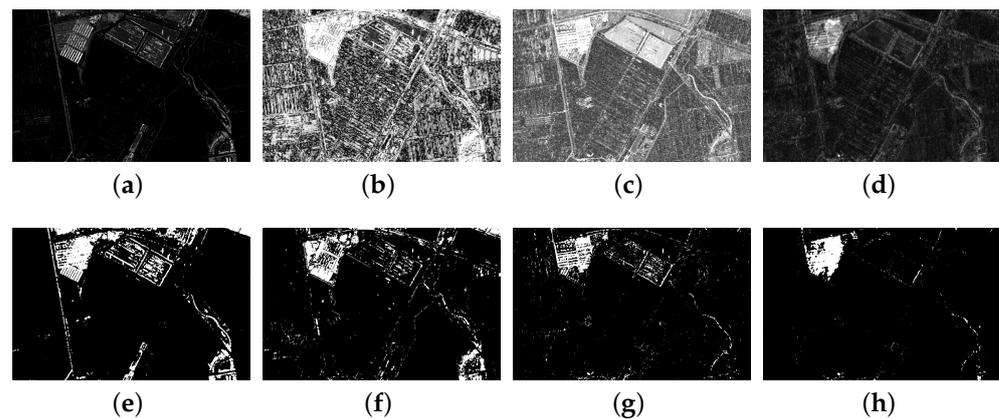| Methods	AUC	FP	FN	OE	CA	KC
SCCN	0.9703	1250	11778	13028	0.9761	0.6050
cGAN	0.9762	1933	9994	11927	0.9782	0.6616
HTP	0.9301	9648	10251	19899	0.9636	0.5273
Proposed	0.9784	2775	7293	10068	0.9816	0.7385

Figure 16. Difference images and change detection results of the compared methods on the Yellow River dataset: (**a**) difference image of SCCN, (**b**) difference image of cGAN, (**c**) difference image of HTP, (**d**) difference image of the proposed method, (**e**) result of SCCN, (**f**) result of cGAN, (**g**) result of HTP, and (**h**) result of the proposed method.

5. Conclusions

In this paper, we propose an edge-oriented GAN (EO-GAN) for change detection based on heterogeneous images by translating one image into one of another style. In particular, unlike the usual homogeneous transformation method, we use an indirect approach, with the edge information that is approximately common in heterogeneous images as the medium of transformation. Through the two processes of edge extraction and reconstruction from the edge based on a cGAN, the function of reconstructing the corresponding optical image from the edge of a radar image is realized. A super-pixel-based method is designed for distortion in order to prompt the network to build connections between edge changes and actual content changes. The experimental results on both homogeneous images and heterogeneous images demonstrate the effectiveness of our proposed method. In future work, we will focus on more complex scenarios such as multiview high-resolution images and design registration translators based on a GAN.

References

1. Varghese, A.; Gubbi, J.; Ramaswamy, A.; Balamuralidhar, P. ChangeNet: A deep learning architecture for visual change detection. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 129–145.
2. Bruzzone, L.; Bovolo, F. A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. *Proc. IEEE* **2012**, *101*, 609–630. [CrossRef]
3. Tang, Y.; Feng, S.; Zhao, C.; Fan, Y.; Shi, Q.; Li, W.; Tao, R. An Object Fine-Grained Change Detection Method Based on Frequency Decoupling Interaction for High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–13. [CrossRef]

4. Zhang, W.; Zhang, Y.; Gao, S.; Lu, X.; Tang, Y.; Liu, S. Spectrum-Induced Transformer-Based Feature Learning for Multiple Change Detection in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–12. [CrossRef]

5. Zhao, X.; Li, S.; Geng, T.; Wang, X. GTransCD: Graph Transformer-Guided Multitemporal Information United Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–13. [CrossRef]

6. Alatalo, J.; Sipola, T.; Rantonen, M. Improved Difference Images for Change Detection Classifiers in SAR Imagery Using Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [CrossRef]

7. Chen, Z.; Song, Y.; Ma, Y.; Li, G.; Wang, R.; Hu, H. Interaction in Transformer for Change Detection in VHR Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–12. [CrossRef]

8. Chen, H.; Zhang, H.; Chen, K.; Zhou, C.; Chen, S.; Zou, Z.; Shi, Z. Continuous Cross-Resolution Remote Sensing Image Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–20. [CrossRef]

9. Dong, W.; Yang, Y.; Qu, J.; Xiao, S.; Li, Y. Local Information-Enhanced Graph-Transformer for Hyperspectral Image Change Detection With Limited Training Samples. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [CrossRef]

10. Dong, W.; Zhao, J.; Qu, J.; Xiao, S.; Li, N.; Hou, S.; Li, Y. Abundance Matrix Correlation Analysis Network Based on Hierarchical Multihead Self-Cross-Hybrid Attention for Hyperspectral Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–13. [CrossRef]

11. Huang, X.; Cao, Y.; Li, J. An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images. *Remote Sens. Environ.* **2020**, *244*, 111802. [CrossRef]

12. Rußwurm, M.; Korner, M. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.

13. Seydi, S.T.; Hasanlou, M. A new land-cover match-based change detection for hyperspectral imagery. *Eur. J. Remote Sens.* **2017**, *50*, 517–533. [CrossRef]

14. Farahani, M.; Mohammadzadeh, A. Domain adaptation for unsupervised change detection of multisensor multitemporal remote-sensing images. *Int. J. Remote Sens.* **2020**, *41*, 3902–3923. [CrossRef]

15. Ma, W.; Yang, H.; Wu, Y.; Xiong, Y.; Hu, T.; Jiao, L.; Hou, B. Change detection based on multi-grained cascade forest and multi-scale fusion for SAR images. *Remote Sens.* **2019**, *11*, 142. [CrossRef]

16. Qu, X.; Gao, F.; Dong, J.; Du, Q.; Li, H.C. Change detection in synthetic aperture radar images using a dual-domain network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]

17. Zhao, W.; Mou, L.; Chen, J.; Bo, Y.; Emery, W.J. Incorporating metric learning and adversarial network for seasonal invariant change detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2720–2731. [CrossRef]

18. Wan, L.; Xiang, Y.; You, H. An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9941–9959. [CrossRef]

19. Dalla Mura, M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson, J.A. Challenges and opportunities of multimodality and data fusion in remote sensing. *Proc. IEEE* **2015**, *103*, 1585–1601. [CrossRef]

20. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; et al. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–39. [CrossRef]

21. Gong, M.; Niu, X.; Zhan, T.; Zhang, M. A coupling translation network for change detection in heterogeneous images. *Int. J. Remote Sens.* **2019**, *40*, 3647–3672. [CrossRef]

22. Liu, J.; Gong, M.; Qin, K.; Zhang, P. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *29*, 545–559. [CrossRef]

23. Niu, X.; Gong, M.; Zhan, T.; Yang, Y. A conditional adversarial network for change detection in heterogeneous images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 45–49. [CrossRef]

24. Liu, Z.; Li, G.; Mercier, G.; He, Y.; Pan, Q. Change detection in heterogenous remote sensing images via homogeneous pixel transformation. *IEEE Trans. Image Process.* **2017**, *27*, 1822–1834. [CrossRef]

25. Li, H.; Gong, M.; Zhang, M.; Wu, Y. Spatially self-paced convolutional networks for change detection in heterogeneous images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4966–4979. [CrossRef]

26. Jiang, X.; Li, G.; Liu, Y.; Zhang, X.P.; He, Y. Change detection in heterogeneous optical and SAR remote sensing images via deep homogeneous feature fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1551–1566. [CrossRef]

27. Kittler, J. On the accuracy of the Sobel edge detector. *Image Vis. Comput.* **1983**, *1*, 37–42. [CrossRef]

28. Martin, D.R.; Fowlkes, C.C.; Malik, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 530–549. [CrossRef]

29. He, J.; Zhang, S.; Yang, M.; Shan, Y.; Huang, T. Bi-directional cascade network for perceptual edge detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3828–3837.

30. Xie, S.; Tu, Z. Holistically-nested edge detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1395–1403.

31. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]

32. Wang, M.; Liu, X.; Gao, Y.; Ma, X.; Soomro, N.Q. Superpixel segmentation: A benchmark. *Signal Process. Image Commun.* **2017**, *56*, 28–39. [CrossRef]

33. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]

34. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.

35. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.

36. Lee, J.S. Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *PAMI-2*, 165–168. [CrossRef] [PubMed]

37. 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]

38. Peng, D.; Bruzzone, L.; Zhang, Y.; Guan, H.; Ding, H.; Huang, X. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5891–5906. [CrossRef]

39. Chen, Z.; Liu, J.; Liu, F.; Zhang, W.; Xiao, L.; Shi, J. Learning Transformations between Heterogeneous SAR and Optical Images for Change Detection. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 3243–3246.