Unsupervised Change Detection Based on Image Reconstruction Loss

Hyeoncheol Noh *
Hanbat National University
hyeoncheol.noh@edu.hanbat.ac.kr

Jingi Ju *
Hanbat National University
jingi.ju@edu.hanbat.ac.kr

Minseok Seo *
SI Analytics Inc
minseok.seo@si-analytics.ai

Jongchan Park
Lunit Inc
jcpark@lunit.io

Dong-Geol Choi †
Hanbat National University
dgchoi@hanbat.ac.kr

Abstract

To train a change detector, bi-temporal images taken at different times in the same area are used. However, collecting labeled bi-temporal images is expensive and time consuming. To solve this problem, various unsupervised change detection methods have been proposed, but they still require unlabeled bi-temporal images. In this paper, we propose an unsupervised change detection method based on image reconstruction loss, which uses only a single-temporal unlabeled image. The image reconstruction model was trained to reconstruct the original source image by receiving the source image and photometrically transformed source image as a pair. During inference, the model receives bi-temporal images as input and aims to reconstruct one of the inputs. The changed region between bi-temporal images shows high reconstruction loss. Our change detector demonstrated significant performance on various change detection benchmark datasets even though only a single-temporal source image was used. The code and trained models are available in https://github.com/cjf8899/CDRL

1. Introduction

In earth vision, change detection is a task to detect semantic changes in two high spatial resolution (HSR) images of the same area at different times (i.e., bi-temporal images). Change detection is a very important task in the field of earth vision and is used for urban expansion, urban planning, environmental monitoring, and disaster assessment [15, 32].

However, manual comparison and change detection between two HSR images is very labor-intensive and expensive. To solve this problem, deep learning-based change detection methods [5, 11] have recently been proposed, and their results are promising. Due to the data-driven nature of deep learning methods, a large-scale training dataset of bi-temporal images and corresponding change labels is essential for supervised approaches [5, 11]. The challenge lies in the expense of dataset creation; collecting the correctly registered bi-temporal HSR images is expensive, and annotating the changes between them is more costly than for general semantic segmentation [28] or object detection datasets [31]. Another challenge is dataset imbalance. A change detection dataset requires two images taken at different times in the same area, and in real-world scenarios, changes are rare, so it is more difficult to collect a change detection dataset in which changes exist (i.e., class balanced).

To solve this data collection problem, various unsupervised change detection (UCD) methods [10, 14, 21, 25] have been proposed. UCD approaches effectively solve the prob-

(a) Unchange (b) Change (c) Pseudo-unchange

Figure 1. Qualitative comparison of (a) unchanged, (b) changed, and (c) pseudo-unchanged bi-temporal pair images. The pseudo-unchanged pair was created by photometric transform.

*equal contribution
†corresponding author

lem of expensive annotations in change detection, but they still require correctly registered bi-temporal HSR images or their performance is low compared to supervised learning methods.

In the existing UCD setting, because it does not explicitly train change and unchange, the prediction result is noisy both when change occurs and when it does not. To solve this problem, most UCD methods use post-processing. However, they do not disclose the post-processing method or they overfit for specific situations e.g., pixels with small area are removed).

Inspired by the study of unsupervised anomaly detection [13, 20], we rethink the UCD setup. In both change detection and anomaly detection, change/anomaly situations are rare in the real world. Unsupervised anomaly detection methods [13, 20] train the image reconstruction model with only normal data, and the model is fitted to a normal distribution. During inference, normal inputs will be well reconstructed, as they fall within the normal distribution; meanwhile, anomaly inputs will have high reconstruction error, as they fall outside the normal distribution. *Can UCD use reconstruction errors like unsupervised anomaly detection?* Because unchanged pairs can be generated synthetically, we can train an image reconstruction model that trains normal distributions. For example, if the change detector is trained on the unchanged area by pairing X^{t1} with itself, it can be trained on the unchanged area without X^{t2} images or changed labels.

In this paper, we propose unsupervised change detection based on image reconstruction loss (CDRL) using only unlabeled single-temporal single-source images. The proposed method explicitly solves the challenges of data collection in change detection, as it does not require expensive bi-temporal HSR images, expensive annotations, or balanced datasets with sufficient changes. CDRL is trained to reconstruct the original source image by receiving the source image and photometrically transformed source image as a pair. The purpose of photometric transformations is to create pseudo-unchanged pairs that mimic unchanged pairs, as show in Fig. 1-(a) and Fig. 1-(c). In the unchanged pairs, there are no structural changes by definition, only style changes or photometric changes. The pseudo-changed pairs can be used to train CDRL instead of changed pairs. Similar to unsupervised anomaly detection, CDRL receives only (pseudo-)unchanged pair images during training and is trained to reconstruct the original source image, so if untrained cases (changed pair Fig. 1-(b)) are input during inference, the reconstruction loss is high for that area.

However, unlike existing unsupervised anomaly detection approaches, change detection usually receives two images, so there are two major problems. First, the image reconstruction models should be able to reconstruct the original source image by receiving two paired images. Second, the reconstruction models should focus more on the structure information of the photometrically transformed source image. To solve this problem, we propose an image reconstruction model using encoder–decoder-based generative adversarial networks. CDRL consists of a shared encoder to extract features from each image and a decoder to fuse the features from the two images for image reconstruction. To focus on the structure information of the photometrically transformed source image, spatial attention was performed only on the photometrically transformed source image.

To validate the efficacy of our proposed CDRL, we evaluated it on LEVIR-CD [6] and WHU-CD [22]. Although CDRL does not use bi-temporal pairs or pre-trained weights, it outperforms the existing UCD method using bi-temporal pairs and the UCD method using pre-trained weights by a large margin.

Our major contributions can be summarized as follows:

- We propose CDRL, a method to train a change detector on a single-temporal single-source image in UCD. To the best of our knowledge, this is the first time a single-temporal single-source image has been used in UCD.
- We propose an encoder–decoder-based generative adversarial network that receives paired images as input.
- We evaluate CDRL on various change detection datasets, and it outperforms existing UCD methods by a large margin.

2. Related Work

The problem we are trying to solve involves the following issues: (1) obtaining a matched bi-temporal image including a changed area is more difficult than obtaining a general single-temporal image, and (2) pairwise annotation is very expensive and time consuming. Therefore, this section focuses on the problems encountered by existing change detection approaches and briefly introduces the field of anomaly detection, which inspired our proposed approach.

2.1. Supervised Change Detection (SCD)

Supervised change detection (SCD) is largely divided into a method that uses only single-temporal information and an approach that performs modeling, e.g., temporal information modeling [33]. A change detector that uses only single-temporal information, called post-classification comparison (PCC), trains a semantic segmentation model during training [33, 34]. After that, the semantic segmentation model predicts the change area through the xor operation of the results obtained by predicting images from two different times during inference. PCC has the great advantage of not requiring coregistrated pair images, but this method

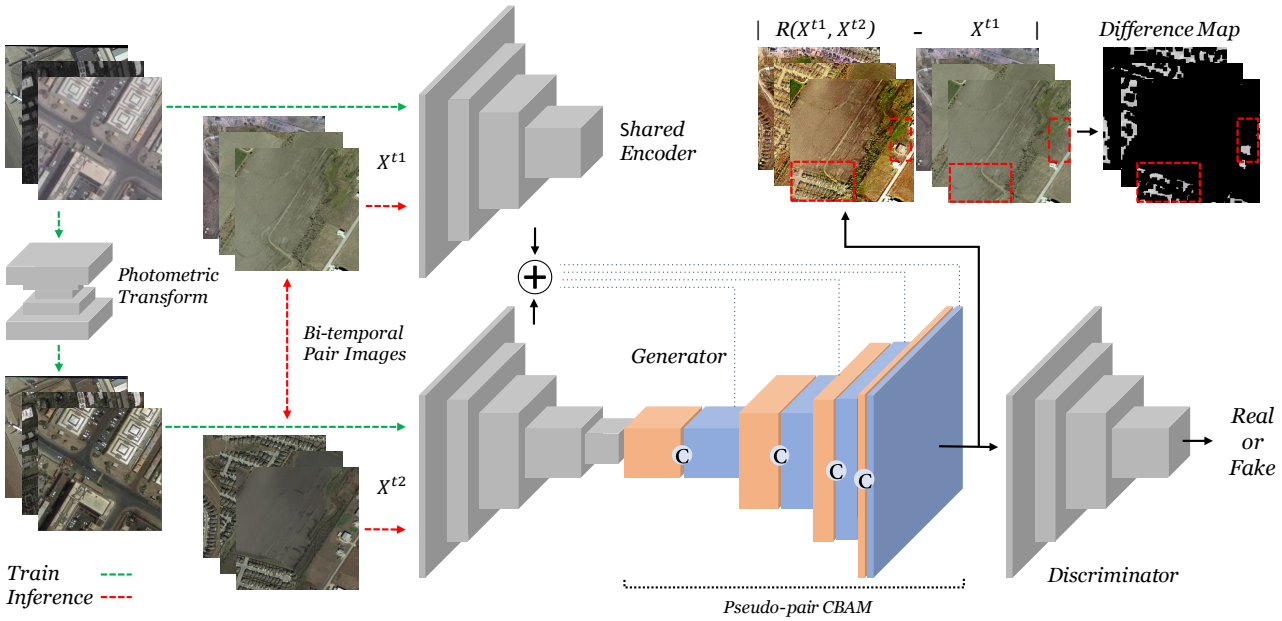

Figure 2. Overview of the overall CDRL framework. CDRL is trained to reconstruct X^{t1} by receiving a pseudo-unchanged pair during training, and when a changed bi-temporal pair that is not learned during training is input during inference, the reconstruction loss is large in the region with large structure change.

only simply treats the change detection task as a semantic segmentation task and ignores the temporal information modeling, thus significantly decreasing the performance.

To solve this problem, change detection methods [5, 11] have been proposed for temporal information modeling between pair images taken at different times in the same area. All of these methods have achieved high performance, but the generalization ability of these models is not guaranteed because of the small size of the change detection benchmark datasets [1, 4, 7–9, 12, 18, 19, 27]. The reason the change detection benchmark datasets are small is that collecting bi-temporal pair images is much more difficult than collecting single-temporal images, and pairwise annotation is very expensive and time-consuming.

Because our proposed CDRL performs UCD using only unlabeled single-temporal single-source images, it can alleviate the problem of collecting bi-temporal pair images and the cost and time consumption of labeling.

2.2. Unsupervised Change Detection (UCD)

UCD is usually divided into a method [3,16,26,30] based on the concept of change vector analysis (CVA) [23] and a method [25] based on a generative adversarial network (GAN) using an unlabeled bi-temporal pair image. However, because they use pre-trained weights without direct training on the dataset, the performance is low, or large-scale unlabeled bi-temporal pair images are required to train the GAN model.

Our proposed CDRL can be explicitly trained on an unchanged area and can be trained without bi-temporal pair images.

2.3. Unsupervised Anomaly Detection

We were inspired by a reconstruction-based anomaly detection method [2, 24]. Reconstruction-based methods typically utilize generative models like autoencoders or GANs to encode and reconstruct the normal data. These methods hold the insights that anomalies cannot be reconstructed because they do not exist in the training samples. These unsupervised anomaly detection methods achieved an area under receiver operating characteristic curve (AUC) performance of over 95 o-n various benchmark datasets [2], even without explicitly training the anomaly data.

We also applied the fact that only unchanged pair (normal) images are trained during training like this reconstruction-based anomaly detection, and when a changed pair (anomaly) is input during inference, the reconstruction loss is high.

3. Method

This section describes the components of the CDRL in detail. First, the training pipeline will be briefly described in Sec. 3.1, and then a method of performing photometric transform based on a single-temporal single-source image is described in Sec. 3.2. Sec. 3.3 describes the reconstructor that receives pair images and is trained as an objective to

reconstruct the original source image. Finally, Sec. 3.4 describes the entire objective function of the CDRL, including the GAN model.

3.1. Overall Pipeline

CDRL performs photometric transform to create a pair image as a single-temporal single-source image. Photometric augmentation of simple rules, such as brightness control and channel shuffling, does not sufficiently express the style change of the corresponding bi-temporal pair image in the real-world unchanged area. Therefore, to express the style change of the corresponding bi-temporal pair image of an unchanged area in the real-world, we perform photometric transform by style transfer using CycleGAN [35].

Next, the generated pair image is input to the U-Net-based original source image reconstructor during training, as shown in Fig. 2. For our purpose, to train the original source image reconstructor with high reconstruction loss for the region where the change has occurred, we must focus on the channel information in the original source image and the spatial information in the photometric transformed image. To achieve this, we applied spatial attention to the photometric transform image and channel attention to the original source image using the convolutional block attention module (CBAM) [29].

Despite these efforts, the original source image reconstructor has a problem of overfitting the original source image too easily during training. Therefore, to prevent overfitting, we made a discriminator and conducted adversarial training with the image reconstructor.

3.2. Photometric Transformation

The purpose of photometric transform for training CDRL is to create a natural style change while maintaining the structure, like in an actual unchanged bi-temporal pair image, as a single-temporal single-source image. To achieve this, we adopted CycleGAN [35], which receives unpaired images and changes the style while maintaining the structure. In the existing CycleGan, when there are two domains $\{x_1, x_2, ..., x_n\} \in X$ and $\{y_1, y_2, ..., y_n\} \in Y$, it receives two samples x_i and y_j and is trained to optimize the parameters of two mapping functions $G : X \rightarrow Y$, $F : Y \rightarrow X$. However, since we need to perform unpaired style transfer in one domain, we train a function that maps two randomly selected samples $x_i^{t1} \in X^{t1}$ and $x_i^{t2} \in X^{t2}$ in one domain X. Therefore, when there is discriminator D_{t2} for mapping function $G : X^{t1} \rightarrow X^{t2}$ and discriminator D_{t1} for $F : X^{t2} \rightarrow X^{t1}$, our objective function is as follows:

$$L(G, F, D_{t1}, D_{t2}) = L_{GAN}(G, D_{t2}, X^{t1}, X^{t2}) + L_{GAN}(F, D_{t1}, X^{t2}, X^{t1}) + \lambda L_{cyc}(G, F), \quad (1)$$

where λ controls the relative importance of the two objectives.

3.3. Pair Image-based Source Image Reconstructor

The pair image-based Source Image reconstructor $R(.)$ is trained to reconstruct X^{t1} by receiving the pseudo unchanged pair image X^{t1}, X^{t2} previously created in Sec. 3.2 as an input. To achieve this, the pair image-based source image reconstructor consists of a shared encoder and a decoder that concatenates and fuses each feature map of the pair image output from the encoder. R is trained to optimize the objective function as follows:

$$L_{mae}(R) = MAE(R(X^{t1}, X^{t2}), X^{t1}), \quad (2)$$

where MAE is the mean absolute error between the reconstructed image and the source image.

A source image reconstructor trained only on pseudo unchanged pair images during training should have a high reconstruction loss when a changed pair image is received during inference. However, if the source image reconstructor relying only on the structure information of the source image regardless of the photometrically transformed image, the reconstruction loss will be low, even when a changed pair image is input. To alleviate this problem, we modified the CBAM structure to perform spatial attention on the photometric transformed image and channel attention on the original source image. Through this process, the source image reconstructor was trained by focusing on the structure information of the photometric transformed image and on the style information in the source image. Figure 3 shows our attention structure modified from the CBAM structure. As shown in the figure, channel attention is performed on the X^{t1} image and spatial attention is performed on the X^{t2} image, which is then added and concatenated for training.

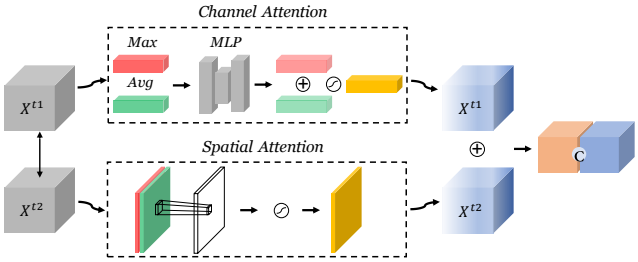

Figure 3. Our proposed pseudo-pair CBAM structure. Channels are applied to the X^{t1} image, and spatial attention is applied to the X^{t2} image

3.4. GAN for detailed structure reconstruction

As in a previous study [17], if only the MLE loss is used, the reconstruction image does not reconstruct the

structure well and is blurry. If the reconstruction result is blurry, the performance of the CDRL is degraded because it is insensitive to structure changes. Therefore, we used a GAN as in [17] to solve this problem. Therefore, given the discriminator D_r, the objective function of the discriminator is

$$L_{gan}(R, D_r, X^{t1}, X^{t2}) = log(D_r(X^{t1})) + log(1 - D_r(R(X^{t1}, X^{t2}))), \quad (3)$$

where R aims to reconstruct images $R(X^{t1}, X^{t2})$ that look similar to images from X^{t1} , while D_r aims to distinguish between translated samples $R(X^{t1}, X^{t2})$ and original source image X_i^{t1}. R aims to minimize this objective against an adversary D_r that tries to maximize it, i.e., $min_R max_{D_r} L(R, D_r, X^{t1}, X^{t2})$.

The final objective function of the source image reconstructor that combines the GAN loss and the MAE loss is

$$L_{total} = L_{gan} + \lambda L_{mae} \quad (4)$$

where λ controls the relative importance of the two objectives. We use λ as 100 in all experiments.

4. Experiments

In this section, CDRL is evaluated on two HSR remote sensing change detection datasets. Section Sec. 4.1 describes the experimental setting in detail, and section Sec. 4.2 describes the loss analysis results in detail. Also, in Sec. 4.3, pixel-level change detection, which is the same as existing change detection experimental settings, is evaluated, and in Sec. 4.4, we describe the patch level change detection results suitable for our proposed change detector application situation. Finally, in Sec. 4.5, qualitative results, and in Sec. 4.6, describes the results of an ablation study.

4.1. Experimental Setting

Datasets Two HSR remote sensing change detection datasets were used to train and evaluate the performance of object change detection.

- **LEVIR-CD** [7]**.** The LEVIR-CD dataset contains 637 bi-temporal pairs of HSR images and 31,333 change labels on building instances. Each image has a spatial size of 1,024 × 1,024 pixels with a spatial resolution of 0.5 m. The change labels provide information about the construction of new buildings and the disappearance of existing buildings. This dataset provides an official split of 445 training, 128 validation, and 64 test pairs. The evaluation results are computed in the test pair set.

- **WHU building change detection** [18]**.** The WHU dataset has one pair of aerial images of size 15,354 × 32,507 pixels obtained in 2012 and 2016 at the same area. It provides 12,796 and 16,077 building instances labels, respectively, and changed labels across the pair. We will use the change labels later to evaluate change detectors. Training, validation, and test sets are given specific areas containing 4,736, 1,036, and 2,416 tiles respectively.

Implementation details CycleGAN was used to generate pseudo unchanged pairs for training of CDRL. The two datasets X^{t1} and X^{t2} of CycleGAN are randomly divided into datasets X. All implementation details strictly follow the official CycleGAN code[1]. For the data augmentation, RandomRotate90, HorizontalFlip, VerticalFlip, Transpose, RandomBrightnessContrast, and Sharpen of albumentations [2] were used, and the probability p of almost all applications is 0.3.

We trained the source image reconstructor using the Adam optimizer with beta values of (0.5, 0.999). The learning rate was set to 0.0002 and batch size was 1 for model training. Evaluation was conducted on both the LEVIR-CD and WHU validation (or test) sets. Our all models were implemented on PyTorch and trained using a single NVIDIA Quadro RTX 8000 GPU.

Evaluation Metrics

- **Pixel Level Change Detection.** We use the common metrics in pixel-by-pixel binary classification tasks and object change detection tasks: intersection over union (IoU), recall, and precision score. This is because our goal is also to classify whether it has changed or not at the pixel level.

- **Patch Level Change Detection.** We used classification AUC for patch level change detection. Note that if all the output mask values of the change detector are 0, it is set to unchanged; if at least one is 1, it is set to changed.

4.2. Loss Analysis Results

We planned a loss analysis experiment to check whether the source image reconstructor has a high reconstruction loss in the part where the structure is changed a lot.

We divided the dataset into unchanged, small change (the changed part is less than 30% of the total image), and large change.

When the dataset was split based on these criteria, the LEVIR-CD dataset was split into 8 unchanged pairs, 35

[1]https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix
[2]https://github.com/albumentations-team/albumentations

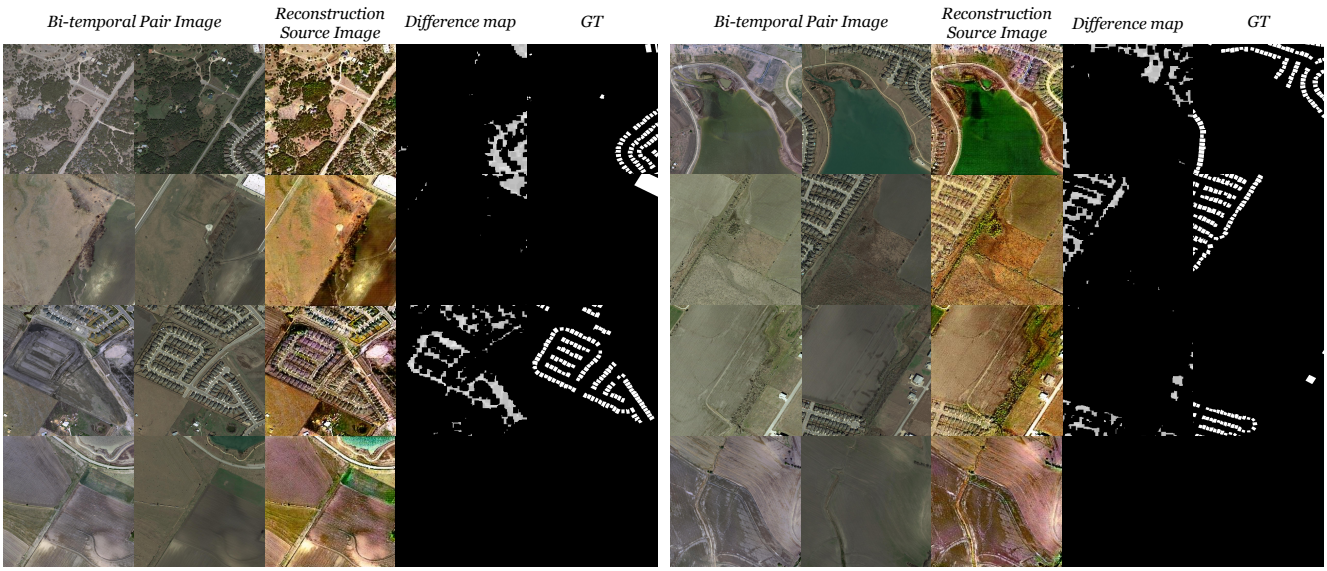

Figure 4. Qualitative analysis of CDRL. The top 3 lines are the qualitative results of CDRL in the region where the change has occurred, and the bottom line is the qualitative result of CDRL in the unchanged pair. It can be seen that CDRL localizes the area where the change occurred.

Method	Dataset	Reconstruction Loss		
		Un	Small	Large
CDRL (Aug)	LEVIR-CD	24.79	36.81	44.91
CDRL (Pseudo Unchange Pair)	LEVIR-CD	14.45	34.74	43.49
CDRL (Aug+Pseudo Unchange Pair)	LEVIR-CD	10.15	31.41	38.53
CDRL (Aug)	WHU	35.95	41.11	50.89
CDRL (Pseudo Unchange Pair)	WHU	22.89	39.37	49.46
CDRL (Aug+Pseudo Unchange Pair)	WHU	17.20	38.03	47.65

Table 1. Loss analysis results of CDRL in LEVIR-CD dataset and WHU dataset.

small change pairs, and 21 large change pairs. The WHU dataset was split into 377 unchange pairs, 145 small change pairs, and 138 large change pairs.

Table 1 lists the results of loss analysis of CDRL on the LEVIR-CD and WHU test datasets. As shown in the table, the loss for the unchanged pairs was the lowest on both datasets, and the loss of the large change pair was the highest. These experimental results indicate that the source image reconstructor is not good at reconstructing the source image when a pair with a large change in structure is input during the test because only pseudo unchanged pairs were input during training, as we intended. Also, the fact that the loss was low for the unchanged pairs indicates that our pseudo unchanged pair was generated at a level similar to that of the actual unchanged pair.

However, if the source image reconstructor worked perfectly as we intended, the reconstruction loss would be close to 0 when an unchanged pair is input. As shown in Table 1, the unchanged pair showed the lowest loss, but the value was not small. The reason for this is that the LEVIR-CD and WHU datasets are only labeled with changes in buildings, and in fact, the unchanged pair includes many structural changes such as land becoming lakes, roads that did not exist, and cars. A more detailed analysis result is described with an example in Sec. 4.5.

4.3. Pixel Level Change Detection Results

We compared and analyzed the performance of CDRL with other UCD and SCD methods. To compare our CDRL with existing UCD methods, we reproduced all [10, 21, 21] methods and tested them in LEVIR-CD and WHU. As shown in Table 2, the existing UCD methods have high recall values and low precision because prediction results are very noisy and vulnerable to small structural changes. These experimental results indicate that our CDRL is robust to both small structural changes and style changes.

However, when compared with BIT, which is a state-of-the-art supervised change detection method, the performance of CDRL was low due to a large gap. The reason for this is twofold. First, unlike supervised change detection, CDRL does not learn the pixel-level change area explicitly, so it can localize only the approximate location. Therefore, compared to BIT, our CDRL has similar recall value but clearly lower precision. Second, supervised change detection can be explicitly trained on information about changed objects of interest, so it can explicitly learn that a car is created or a lake is changed or unchanged, but our CDRL predicts that they are all changed.

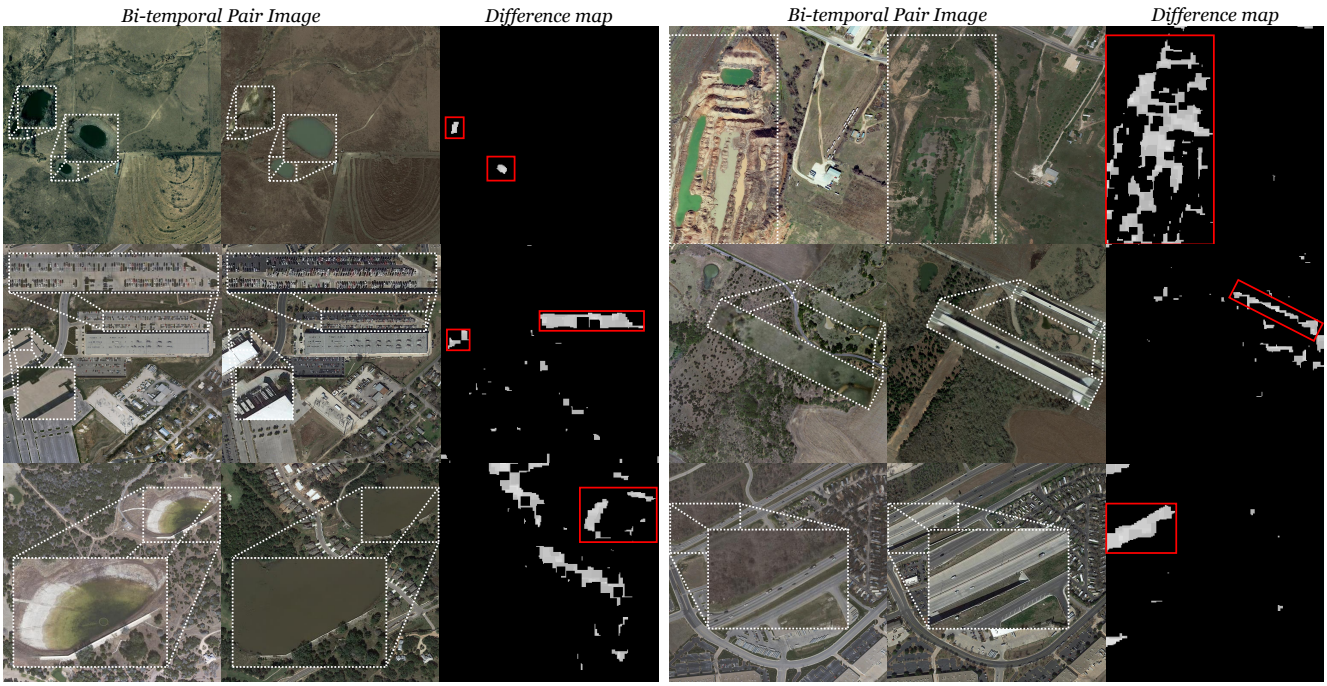

Figure 5. Qualitative analysis result of unchanged bi-temporal images for which CDRL showed high reconstruction loss. CDRL predicted that when roads, lakes, and cars were created, it was all change.

Method	Dataset	Supervision	Precision	Recall	IoU
NMCD [21]	LEVIR-CD	Unsup	0.13	0.71	0.07
UCNN [10]	LEVIR-CD	Unsup	0.16	0.79	0.09
UCDGAN [21]	LEVIR-CD	Unsup	0.20	0.66	0.15
CDRL	LEVIR-CD	Unsup	**0.63**	**0.92**	**0.59**
BIT [5]	LEVIR-CD	Sup	0.89	0.89	0.80
NMCD [21]	WHU	Unsup	0.07	0.96	0.03
UCNN [10]	WHU	Unsup	0.07	0.95	0.03
UCDGAN [21]	WHU	Unsup	0.09	0.93	0.08
CDRL	WHU	Unsup	**0.52**	**0.93**	**0.50**
BIT [5]	WHU	Sup	0.86	0.81	0.72

Table 2. Quantitative comparison results of CDRL and UCD, SCD methods. Note that, since there are no post-processing implementation details of the existing UCD methods, it was not applied.

4.4. Patch Level Change Detection Results

Many real-world applications that use change detectors do not rely solely on change detectors. In these situations, the role of change detectors is to reduce human labor intensity by providing information on the patches or areas where the change occurred among hundreds of patches. Considering this application situation, we aimed to solve change detection with patch level classification.

Table 3 shows the patch level change detection results for our CDRL on the LEVIR-CD and WHU datasets. As shown in the table, a high AUC was achieved on both datasets despite using only single-temporal single-source images Note that CVA-based methods predict that there is a change in all patches because the output result is noisy.

Dataset	AUC
LEVIR-CD	83.52
WHU	87.18

Table 3. Patch level binary classification results in the LEVIR-CD dataset and the WHU dataset.

4.5. Qualitative Results

CDRL was qualitatively analyzed on the LEVIR-CD dataset. As shown in Fig. 4, the CDRL detects the changed part well because the reconstruction loss is high in the part where the structure change is large.

Also, for the unchanged pair, even if the style change is large, because there is no structure change, it can be seen that the reconstruction loss is low and no change is predicted. Therefore, it is thought that CDRL will be useful in applications where it is not important to localize the exact location but to know the approximate location or whether or not changes have occurred in units of patches.

Fig. 5 shows the qualitative analysis results for samples with poor CDRL performance. As shown in the figure, because CDRL cannot designate a specific change object of

interest, it predicts that a change has occurred when a car is created, the ground becomes a lake, or a lake becomes ground. Therefore, the low IoU in all our experiments is dominant for the above reasons.

4.6. Ablation Study

To compare and analyze the effects of the attention module and adversarial training constituting the CDRL, we conducted an ablation study on the LEVIR-CD dataset.

Attention Modules Table 4 lists the patch level classification results of CDRL in the LEVIR-CD dataset according to the existence of channel attention and spatial attention. As shown in the table, performance was higher with CBAM than without CBAM. Also, in our proposed CDRL, the pseudo-pair CBAM, which provides channels attention to the X^{t1} image and spatial attention to the X^{t2} image, has the highest performance.

CDRL	AUC
w/o Attention	77.38
w/ CBAM	80.90
w/ Pseudo-pair CBAM	83.52

Table 4. Patch level binary cleavage results of CDRL with or without attention module.

Adversarial Training We designed an experiment to check whether adversarial training solves the blurring problem of reconstruction images like [17] in CDRL.

w/o D_r w/ D_r

Figure 6. Qualitative comparison of reconstructed images with and without discriminator.

Fig. 6 is a sample of the reconstructed image with and without discriminator. As shown in the figure, the reconstructed image is blurred when there is no discriminator. Meanwhile, if there is a discriminator, it can be seen that the boundary is reconstructed more clearly. These results indicate that adversarial training is also effective in CDRL

CDRL	AUC
w/o discriminator	69.09
w/ discriminator	83.52

Table 5. Patch level binary cleavage results of CDRL with or without discriminator.

Table 5 lists the results of measuring the AUC of CDRL with and without discriminator on the LEVIR-CD dataset. As shown in the table, higher AUC was achieved with the discriminator. These experimental results show that the sensitivity of the CDRL to the structure of the pair image helps the performance, as intended.

5. Discussion and Future work

In this study, we performed change detection with source image reconstruction loss using only unlabeled single-temporal single-source images. However, the semantic change we are interested in can exist in a variety of ways, such as buildings, seasons, cars, and trees. The prerequisite for CDRL is to make the reconstruction loss appear high in the part where the structure change occurs regardless of the style change. However, the semantic change that we are interested in can be diverse, such as natural scenery, artificial objects, weather, and environmental changes. Therefore, in future work, based on the fact that CDRL has a significant performance improvement in UCD, we plan to study semi-supervised change detection to efficiently detect changes of interest (change of specific objects).

6. Conclusion

In this paper, to solve the problem that it is difficult to construct a bi-temporal pair dataset containing semantic changes, we proposed a CDRL, which performs unsupervised change detection using only a single-temporal single-source image. To solve the unsupervised change detection problem as a reconstruction-based unsupervised anomaly detection problem, CDRL defined normal data as unchanged pairs and anomaly data as changed pairs. After that, a change detector (reconstructor) that receives pair images was proposed. We verified the CDRL on the WHU and LEVIR-CD datasets and achieved significant results despite unsupervised change detection using single-temporal single-source images. We hope that CDRL will be widely used in real-world scenarios where it is difficult to obtain labeled bi-temporal pair images.

References

[1] Csaba Benedek and Tamás Szirányi. Change detection in optical aerial images by a multilayer conditional mixed markov model. *IEEE Transactions on Geoscience and Remote Sensing*, 47(10):3416–3430, 2009. 3

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 3

[3] Thomas Blaschke and Geoffrey J Hay. Object-oriented image analysis and scale-space: theory and methods for modeling and evaluating multiscale landscape structure. *International Archives of Photogrammetry and Remote Sensing*, 34(4):22–29, 2001. 3

[4] Nicolas Bourdis, Denis Marraud, and Hichem Sahbi. Constrained optical flow for aerial image change detection. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 4176–4179. IEEE, 2011. 3

[5] Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 2021. 1, 3, 7

[6] Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10), 2020. 2

[7] Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10):1662, 2020. 3, 5

[8] Rodrigo Caye Daudt, Bertr Le Saux, Alexandre Boulch, and Yann Gousseau. Urban change detection for multispectral earth observation using convolutional neural networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 2115–2118. IEEE, 2018. 3

[9] Rodrigo Caye Daudt, Bertrand Le Saux, Alexandre Boulch, and Yann Gousseau. Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding*, 187:102783, 2019. 3

[10] Kevin Louis de Jong and Anna Sergeevna Bosman. Unsupervised change detection in satellite images using convolutional neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. 1, 6, 7

[11] Sheng Fang, Kaiyu Li, Jinyuan Shao, and Zhe Li. Snunet-cd: A densely connected siamese network for change detection of vhr images. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021. 1, 3

[12] Aito Fujita, Ken Sakurada, Tomoyuki Imaizumi, Riho Ito, Shuhei Hikosaka, and Ryosuke Nakamura. Damage detection from aerial images via convolutional neural networks. In *2017 Fifteenth IAPR international conference on machine vision applications (MVA)*, pages 5–8. IEEE, 2017. 3

[13] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1705–1714, 2019. 2

[14] Pengfei He, Xiangwei Zhao, Yuli Shi, and Liping Cai. Unsupervised change detection from remotely sensed images based on multi-scale visual saliency coarse-to-fine fusion. *Remote Sensing*, 13(4):630, 2021. 1

[15] Masroor Hussain, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of photogrammetry and remote sensing*, 80:91–106, 2013. 1

[16] Jungho Im, JR Jensen, and JA Tullis. Object-based change detection using correlation image analysis and image segmentation. *International journal of remote sensing*, 29(2):399–423, 2008. 3

[17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 4, 5, 8

[18] Shunping Ji, Shiqing Wei, and Meng Lu. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):574–586, 2018. 3, 5

[19] MA Lebedev, Yu V Vizilter, OV Vygolov, VA Knyaz, and A Yu Rubis. Change detection in remote sensing images using conditional adversarial networks. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2), 2018. 3

[20] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021. 2

[21] Xuelong Li, Zhenghang Yuan, and Qi Wang. Unsupervised deep noise modeling for hyperspectral image change detection. *Remote Sensing*, 11(3):258, 2019. 1, 6, 7

[22] Jin Liu and Shunping Ji. A novel recurrent encoder-decoder structure for large-scale multi-view stereo reconstruction from an open aerial dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6050–6059, 2020. 2

[23] William A Malila. Change vector analysis: an approach for detecting forest changes with landsat. In *LARS symposia*, page 385, 1980. 3

[24] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2898–2906, 2019. 3

[25] Caijun Ren, Xiangyu Wang, Jian Gao, Xiren Zhou, and Huanhuan Chen. Unsupervised change detection in satellite images with generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12):10047–10061, 2020. 1, 3

[26] Frank Thonfeld, Hannes Feilhauer, Matthias Braun, and Gunter Menz. Robust change vector analysis (rcva) for

multi-sensor very high resolution optical satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 50:131–140, 2016. 3

[27] Shiqi Tian, Ailong Ma, Zhuo Zheng, and Yanfei Zhong. Hi-ucd: A large-scale dataset for urban semantic change detection in remote sensing imagery. *arXiv preprint arXiv:2011.03247*, 2020. 3

[28] Syed Waqas Zamir, Aditya Arora, Akshita Gupta, Salman Khan, Guolei Sun, Fahad Shahbaz Khan, Fan Zhu, Ling Shao, Gui-Song Xia, and Xiang Bai. isaid: A large-scale dataset for instance segmentation in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 28–37, 2019. 1

[29] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 4

[30] Chen Wu, Bo Du, and Liangpei Zhang. Slow feature analysis for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5):2858–2874, 2013. 3

[31] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018. 1

[32] Xueliang Zhang, Pengfeng Xiao, Xuezhi Feng, and Min Yuan. Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area. *Remote Sensing of Environment*, 201:243–255, 2017. 1

[33] Zhuo Zheng, Ailong Ma, Liangpei Zhang, and Yanfei Zhong. Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15193–15202, 2021. 2

[34] Zhuo Zheng, Yanfei Zhong, Junjue Wang, and Ailong Ma. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4096–4105, 2020. 2

[35] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 4