• Supplementary Material •

DDPM-CD: Denoising Diffusion Probabilistic Models as Feature Extractors for Remote Sensing Change Detection

1. Related Work

1.1. Remote Sensing Change Detection

1.1.1 Classical change detection methods

Classical change detection methods in remote sensing can be primarily categorized into three groups: (1) algebraic, (2) transformation-based, and (3) classification-based techniques.

Algebraic methods, including image differencing (ImageDiff) [30], image regression (ImageRegr) [31], image ratioing (ImageRatio) [30], and change vector analysis (CVA) [35], rely on selecting thresholds to identify altered areas. These methods, except for CVA, are relatively simple to implement but cannot provide comprehensive matrices of change information. Their reliance on threshold selection remains a significant drawback.

Transformation-based methods, such as Principal Component Analysis (PCA)[14, 15], Karhunen-Loève Transform (KT)[30], Gramm–Schmidt (GS)[30], Multivariate Alteration Detection (MAD)[34], Re-weighted Multivariate Alteration Detection (IRMAD)[33], and Chi-square transformations[30], aim to reduce data redundancy between bands and emphasize different information in derived components. However, they often require threshold selection and encounter challenges in interpreting and labeling change information on transformed images.

Contrarily, classification-based methods like post-classification comparison [30], spectral–temporal combined change analysis [30], and expectation–maximization algorithm (EM) change detection [30], operate based on classified images. These methods heavily rely on the quality and quantity of training sample data to produce accurate classification results. They offer the advantage of providing change information matrices, mitigating external impacts from atmospheric and environmental differences between multi-temporal images. However, their modeling capacity and change detection quality are limited compared to modern deep learning-based approaches.

1.1.2 Deep learning-based change detection methods

The current research on remote sensing change detection has been significantly reshaped by deep learning owing to its powerful feature extraction ability [1]. Initially, it was primarily based on fully convolutional neural networks (CNNs) and did not utilize any form of pre-training; instead, it solely relied on supervised learning from labeled data in an end-to-end fashion. Examples of such approaches include Fully-Convolutional Early Fusion (FC-EF)[13], Fully-Convolutional Siamese Concatenation (FC-Siam-conc) [13], and Fully-Convolutional Siamese Difference (FC-Siam-diff) [13]. In the EF architecture, pre-change and post-change images are concatenated before passing them through the CNN, treating them as different color channels. In the Siamese network architecture, the encoding layers of the network are bifurcated into two streams of equal structure with shared weights, and each image is assigned to one of these streams. Subsequently, a feature difference (FC-Siam-diff) or feature concatenation (FC-Siam-conc) is applied before the final change classifier. In many cases, the Siamese difference/concatenation architecture has proven effective for change detection. Consequently, it became commonly utilized in later works for change detection purposes.

With the evolution of more potent CNN architectures such as VGG [36], ResNet [19], DenseNet [21], and the availability of their pre-trained models on large-scale natural image datasets like ImageNet, remote sensing methods employing transfer learning from natural images to remote sensing images have emerged. For instance, DSIFN (deeply supervised image fusion network) [44], DASNet (dual attentive Siamese network) [8], SemiCD (semi-supervised change detection) [3], and ADS-Net (attention-based deeply supervised network) [40] have utilized multi-scale features from VGG16 and ResNet50 pre-trained on ImageNet to train change detection networks.

The introduction of transformer networks [38], with the core component being multi-head self-attention (MHSA) [38] capable of capturing long-range context and

relationships between different positions, has seen adoption in remote sensing change detection. Inspired by the Vision Transformer (ViT) [16] approach, where the input image is divided into fixed-size patches forming tokens that are then processed by MHSA, BIT [6] was adapted for remote sensing change detection by operating on latent feature representations obtained from ImageNet pre-trained ResNet [19]. Furthermore, a recent work, ChangeFormer [4], proposed a fully transformer network devoid of 2D convolutions for change detection, achieving superior results compared to its counterparts. Later versions of transformer networks, such as the Swin Transformer [28], which substitutes the global MHSA with the shiftable window MHSA (WMHSA) to significantly reduce ViT's computational overhead, have also been adopted in remote sensing change detection, as seen in SwinSUNet [43].

However, transformers tend to be data-hungry and typically require a well-pre-trained model to achieve better performance. Most of the previously mentioned transformer networks proposed for change detection utilize pre-trained models on natural image datasets like ImageNet [24] and ADE20k [46], or are randomly initialized. This is sub-optimal because aerial images possess distinct characteristics creating a significant domain gap compared to natural images, including differences in view, color, texture, layout, objects, and more. To bridge this gap, these methods attempt to narrow it by further fine-tuning the pre-trained model on the remote sensing image dataset. Nevertheless, the systematic bias introduced by ImageNet pre-training has a noticeable impact on performance [41].

With the emergence of large-scale aerial scene classification datasets (such as MillionAID [29], fMoW [12], and BigEarthNet [37]), and access to publicly available large-scale unlabeled remote sensing datasets from various Earth observation programs, it is now possible to pre-train CNN and transformer backbones on remote sensing images. However, there have been few explorations in remote sensing pre-training, and it is still not as renowned as pre-training in the natural image domain. In Geographical Knowledge-driven Representation learning (GeoKR) [26], global land cover products are considered as labels and a mean-teacher framework is used to alleviate the influences of imaging time and resolution differences between RS images and geographical ones. The scarcity of large-scale remote sensing datasets is mainly in terms of category labels rather than images. Hence, it is promising to develop self-supervised pre-training methods, and some related methods have been developed.

For instance, SeCo [32] leverages seasonal changes to enforce consistency between positive samples, which are unique characteristics of aerial scenes. Meanwhile, in Geography-Aware Self-Supervised Learning [2], temporal information and geographical location are simultaneously fused into the MoCo-V2 [10, 18]. Moreover, exploration into remote sensing image colorization from multi-spectral images [39] and spatial properties of remote sensing images [23] has also been conducted.

Although these self-supervised methods do not rely on labeled data during pre-training, they still use paired multi-temporal images (like SeCo [32]), access to paired multi-band spectral images (as in remote sensing colorization [39]), or require spatially aligned remote sensing images with known geo-locations (as in geography-aware-ssl [2]). This limitation restricts their ability to easily harness information from millions of off-the-shelf remote sensing images.

Unlike existing self-supervised methods in remote sensing, our research pioneers the use of DDPM [20], originally designed for image synthesis in generative AI, as a pre-training strategy for robust feature extraction from remote sensing images. This innovative pre-training approach only requires access to readily available remote sensing image datasets. Upon pre-training the DDPM, we utilize it to extract feature representations that can be leveraged to train a light-weight change detection model with annotated change images. The extraordinary capacity of DDPM to model complex training distributions more efficiently than other generative models (such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), etc.) enables the extraction of highly informative and compressed feature representations of a give image. Our experiments on multiple change detection datasets show that these representations obtained from pre-trained DDPM are pivotal in enhancing change detection performance significantly.

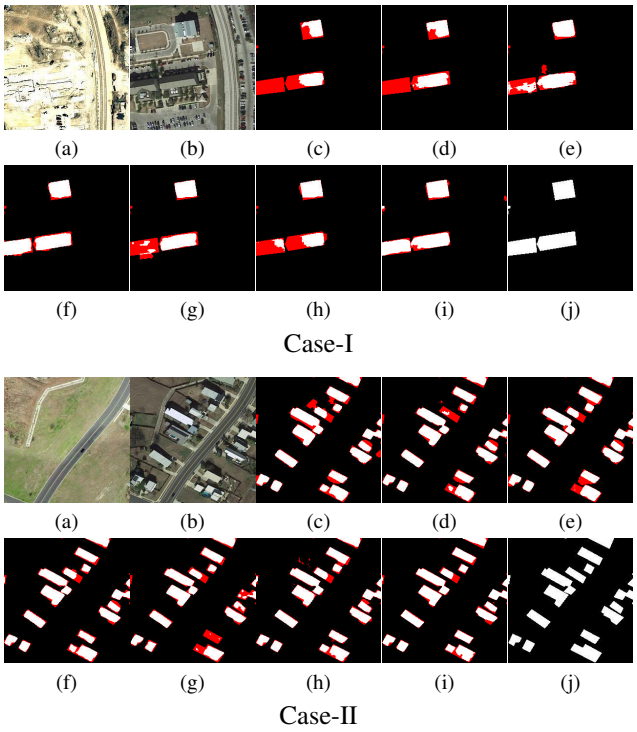

Figure 2. Comparison of different state-of-the-art change detection methods on **LEVIR-CD** dataset: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF, **(d)** FC-Siam-diff, **(e)** FC-Siam-conc, **(f)** DT-SCN, **(g)** BIT, **(h)** ChangeFormer, **(i)** DDPM-CD (ours), and **(j)** Ground-truth. *Note that true positives (change class) are indicated in white, true negatives (no-change class) are indicated in black, and false positives plus false negatives indicates in red.*

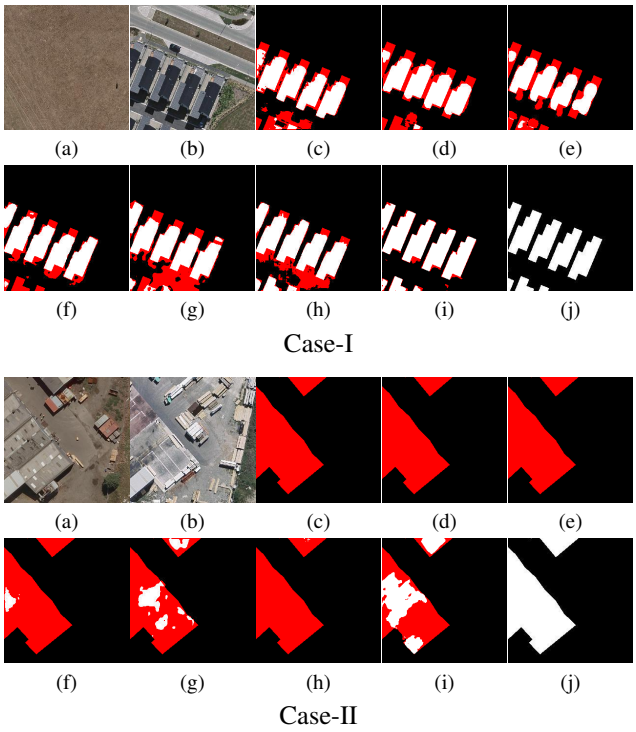

Figure 4. Comparison of different state-of-the-art change detection methods on WHU-CD dataset: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF, **(d)** FC-Siam-diff, **(e)** FC-Siam-conc, **(f)** DT-SCN, **(g)** BIT, **(h)** ChangeFormer, **(i)** DDPM-CD (ours), and **(j)** Ground-truth. *True positives (change class) are indicated in white, true negatives (no-change class) are indicated in black, and false positives plus false negatives indicates in red.*

2. Additional qualitative change detection results

Besides the quantitative results, we visually present predicted change maps to highlight the effectiveness of the proposed method compared to state-of-the-art methods. Figure 2, Figure 4, Figure 6, and Figure 8 display qualitative examples corresponding to the LEVIR-CD, WHU-CD, DSIFN-CD, and CDD datasets, respectively. In these visualizations, we represent the change class (positive class) in white, the no-change class (negative class) in black, and incorrectly predicted areas (false positives and false negatives) in red. Therefore, fewer red areas in a method indicate better performance in predicting both change and no-change classes.

For the LEVIR-CD dataset presented in Figure 2. The first example depicts three building changes, while in the second case, many buildings have appeared, resulting in numerous building changes. In the first case, we can observe that our DDPM-CD accurately captures all three building changes, while other methods like FC-EF, FC-Siam-diff, FC-Siam-conc, BIT, and Changeformer either miss the building in the left-middle or can only partially predict the changes. When considering the second case, although most previously proposed change detection methods can predict most of the building changes, the predictions of DDPM-CD are more accurate and have fewer red areas.

For the WHU-CD dataset shown in Figure 4, one with multiple building changes and the other with two building changes. In the first example, we can see that the change predictions from our DDPM-CD are more accurate and have sharper edges, while all the other methods struggle to predict the changes appearing at the bottom and struggling to differentiate building shadows with actual building parts. In the second case, which contains a very large building change on the left, challenging to recognize, all the other methods missed it, but our method was at least able to partially predict the change. Additionally, there is another change at the top, which was not predicted by any of the previous methods except BIT. However, our method has predicted most of the change area in that region and performed better than the prediction from BIT.

Differing from building change detection, let's now consider the visual quality of predictions on general change de-

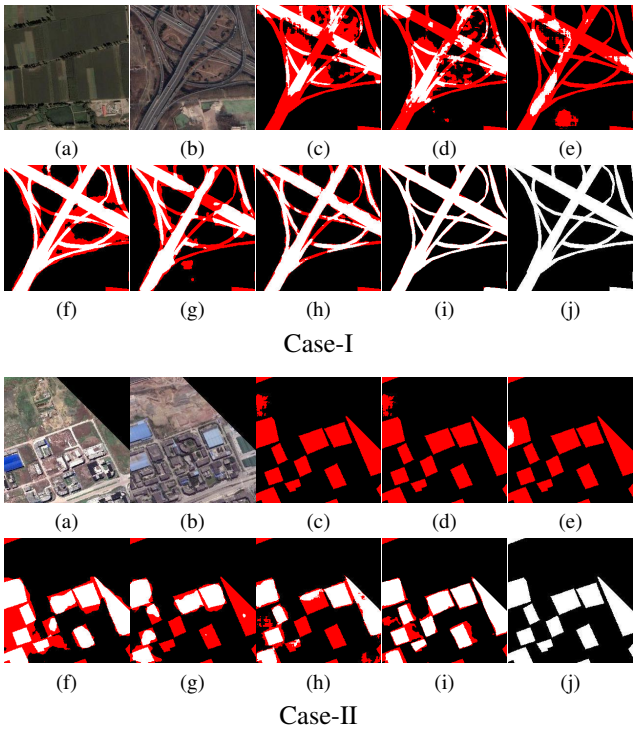

Figure 6. Comparison of different state-of-the-art change detection methods on **DSIFN-CD** dataset: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF, **(d)** FC-Siam-diff, **(e)** FC-Siam-conc, **(f)** DT-SCN, **(g)** BIT, **(h)** ChangeFormer, **(i)** DDPM-CD (ours), and **(j)** Ground-truth. *Note that true positives (change class) are indicated in white, true negatives (no-change class) are indicated in black, and false positives plus false negatives indicates in red.*

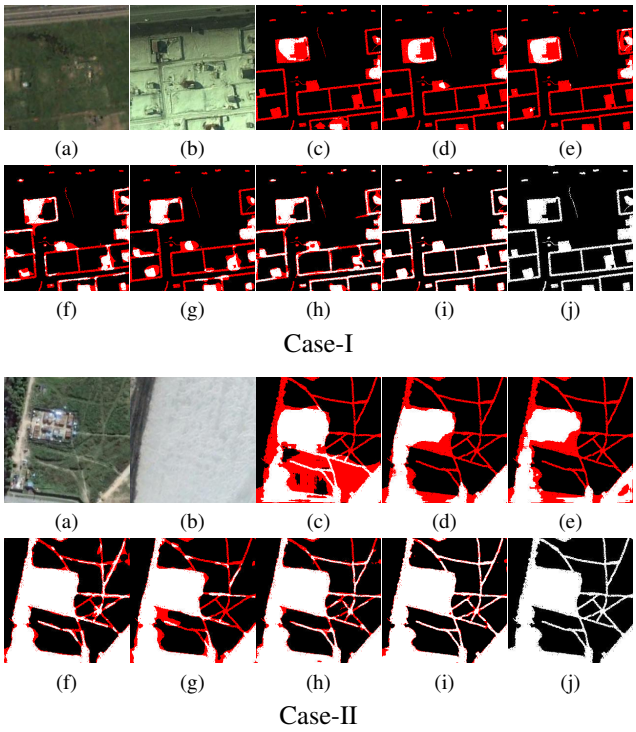

Figure 8. Comparison of different state-of-the-art change detection methods on **CDD** dataset: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF, **(d)** FC-Siam-diff, **(e)** FC-Siam-conc, **(f)** DT-SCN, **(g)** BIT, **(h)** ChangeFormer, **(i)** DDPM-CD (ours), and **(j)** Ground-truth. *Note that true positives (change class) are indicated in white, true negatives (no-change class) are indicated in black, and false positives plus false negatives indicates in red.*

tection datasets like DSIFN-CD and CDD. We showcase prediction results for two examples from the DSIFN-CD dataset in Figure 6. The first case includes changes due to highway construction, while the other contains changes related to new buildings. Given the nature of highways with numerous narrow and curved parts, all other methods miss most of these changes because it's challenging to predict due to the similarities in colors between highways and forests. However, our method can easily differentiate between highway and forest regions, resulting in highly accurate change predictions. In the second example, several challenging-to-recognize building changes appear, and our method accurately detects these regions better than all other methods, particularly in the changes visible on the left.

We also present two examples from the CDD dataset in Figure 8. The first example exhibits changes in buildings and roadways. However, the post-change image was captured during the snow season, making those changes challenging to recognize and predict. As observed, all other methods struggle to capture these changes, but our method accurately predicts them. In the second example, the narrow roadways and buildings visible in the pre-change image disappear in the second image. While the building changes are clearly visible, the narrow roadways, obscured by the forest, are challenging to predict. While state-of-the-art methods predict building change areas, they face difficulties with the narrow and obscured roadways. However, our method accurately predicts these narrow roadways, resulting in a high-quality change map.

All of these qualitative comparisons underscore the effectiveness of our proposed DDPM-CD method compared to the existing state-of-the-art methods. Moreover, it demonstrates the extraordinary ability of DDPM to deliver robust and discriminative features that are useful in downstream applications like change detection.

3. Ablation Studies

3.1. Ablation study on multi-timestep features

This ablation study investigates the impact of utilizing different multi-timestep features ($t \in [0, T]$) from the diffusion model on change detection performance. We fine-tune the change detection classifiers using features obtained at various timesteps t from the diffusion model to identify the timestep range that provides optimal semantics for change detection. Table 1 illustrates how change detection performance on the validation set varies when utilizing features sampled from different timesteps: $t = 5, 50, 100, (50 \text{ and } 100), (50, 100 \text{ and } 400)$, and $(50, 100, \text{ and } 650)$ as inputs for training the hierarchical change classifier.

Our observations indicate that the most favorable change detection performance across all datasets is achieved when utilizing feature representations sampled within the range of $t \in [100, 400]$. Moreover, combining feature representations from multiple time samples, such as $t = 50, 100$, and 400, further enhances change detection performance. Consequently, we designate feature representations sampled at t =50, 100, and 400 as the default configuration for multi-timestep features, which is employed to report results on the test sets of all datasets presented in Table **??**.

3.2. Comparison of Computational Complexity

Table 2 compares the computational complexity of the proposed DDPM-CD with the existing methods. We benchmark our method for pre-change and post-change images of spatial resolution 256×256 and use an NVIDIA Quadro RTX 8000 GPU.

Our DDPM has a total of 390.95 million trainable parameters. The hierarchical change classifier has 39.08 million parameters if single-timestep features are used, 43.96 million parameters if two timesteps are used, and 46.41 million if three time-steps are used. Since we fine-tune only the hierarchical change classifier and keep the DDPM frozen, the total trainable parameters during the fine-tuning come from the hierarchical change detector. DDPMs usually require a higher number of parameters to enable their modeling capability, and more recent DDPMs have even higher parameter counts.

When considering GLOPs and inference time, the DDPM consumes 716.40 GLOPs per image pair and takes about 28.75 ms per image pair for one step forward pass. Since we utilize DDPM for feature extraction during fine-tuning and inference, it requires 1-3 forward passes to extract multi-step features, whereas if we use it in the synthesis, which usually involves 1000s of time-steps, it requires $\times 1000$ times. For our best model, which utilizes features corresponding to three time steps, it requires 3×716.40 = 2149.2 GLOPs and takes 28.75×3 = 86.25 ms. The hierarchical change detector, which processes those features and outputs a change map, requires 32.84 GLOPs and takes 2.56 ms when utilize features of three timesteps. Therefore, for the best model, it requires a total of 2149.2 + 32.84 = 2182.04 GLOPs and takes 86.25 + 2.56 = 88.81 ms.

In comparison to other state-of-the-art methods, our method exhibits higher counts of trainable parameters, GLOPs, and inference time. This observation is understandable because the DDPM necessitates a large network to enable its modeling power, allowing it to accurately capture the training distribution, unlike other architectures. We believe that despite the higher number of parameters and GLOPs, the final performance of our method outweighs these metrics when compared to other state-of-the-art methods. Exploring ways to reduce its model size while retaining its modeling capabilities and decreasing inference time would be both intriguing and timely. Presently, the current trend in diffusion models leans toward larger sizes, a direction driven by the demanding nature of handling extremely complex input data distributions, the need for high-quality image synthesis, and the increasing complexity of multi-modal data in the natural image domain.

4. Results on LEVID-CD+ Dataset

Table 1. **The ablation study on the timestep t used to extract multi-timestep feature representations**. We show that combining feature representations belonging to multiple timesteps improves the change detection performance on the `val`-set of LEVIR-CD, WHU-CD, DSIFN-CD, and CDD.

Time step t	LEVIR-CD [7]			WHU-CD [22]			DSIFN-CD [45]			CDD [25]		
	F1	**IoU**	**OA**	**F1**	**IoU**	**OA**	**F1**	**IoU**	**OA**	**F1**	**IoU**	**OA**
5	89.71	81.35	99.15	91.57	84.46	99.19	93.87	88.39	96.09	91.24	83.89	91.24
50	90.66	82.90	99.23	92.74	86.47	99.31	94.17	88.99	96.29	93.78	88.28	98.60
100	90.50	82.65	99.21	92.78	86.54	99.31	94.95	90.39	96.77	94.32	89.25	98.72
150	90.08	81.95	99.18	92.34	85.77	99.27	94.59	89.74	96.54	94.34	89.29	98.75
50, 100	**91.02**	**83.52**	**99.26**	**93.09**	**87.07**	**99.34**	**94.51**	**89.61**	**96.51**	**94.91**	**90.31**	**98.85**
50, 100, 400	**91.26**	**83.92**	**99.28**	**93.50**	**87.80**	**99.38**	**95.38**	**91.18**	**94.05**	**95.64**	**91.64**	**99.00**
50, 100, 650	**91.10**	**83.67**	**99.26**	**93.02**	**86.95**	**99.33**	**95.07**	**90.62**	**96.87**	**95.24**	**90.90**	**98.92**

Table 2. Comparison of computational complexity of different methods. We consider pre-change and post-change images of size 256×256.

Method	Trainable Params. (M)	GLOPs	Inference Time (ms)
SimSiam [9]	12.49	4.76	1.04
MoCo-v2 [11]	11.24	4.76	1.92
DenseCL [42]	11.69	4.76	2.66
CMC [5]	22.48	4.66	1.55
SeCo [32]	12.16	9.52	3.62
DDPM	390.95	716.40	28.75
CD w/ $n = 1$	39.08	25.99	1.85
CD w/ n=2	43.96	30.56	2.46
CD w/ n=3	46.41	32.84	2.56
DDPM-CD (n=1)	39.08	$1 \times 716.49 + 25.99 = 742.48$	$1 \times 28.75 + 1.85 = 30.6$
DDPM-CD (n=2)	43.96	$2 \times 716.49 + 30.56 = 1458.97$	$2 \times 28.75 + 2.46 = 59.35$
DDPM-CD (n=3)	46.41	$3 \times 716.49 + 32.84 = 2175.46$	$3 \times 28.75 + 2.56 = 88.10$

Table 3. Accuracy assessment for different binary CD models on the LEVIR-CD+ adapted from changemamba[17].

Type	Method	OA	F1	IoU
$\mathcal{C}$	FC-EF	97.54	70.42	54.34
	FC-Siam-Diff	98.26	77.57	63.36
	FC-Siam-Conc	98.24	78.44	64.53
	SiamCRNN-18	98.56	82.71	70.52
	SiamCRNN-34	98.61	83.08	71.05
	SiamCRNN-50	98.68	83.46	71.61
	SiamCRNN-101	98.67	83.20	71.23
	DSIFN	98.70	84.07	72.52
	SNUNet	97.83	74.70	59.62
	HANet	98.22	77.56	63.34
	CGNet	98.63	83.68	71.94
	SEIFNet	98.66	83.32	71.41
$\mathcal{T}$	ChangeFormerV1	98.38	79.51	65.98
	ChangeFormerV2	98.36	80.20	66.94
	ChangeFormerV3	98.44	80.65	67.58
	ChangeFormerV4	98.01	75.87	61.12
	ChangeFormerV5	98.23	78.23	64.24
	ChangeFormerV6	97.60	72.71	57.12
	BIT-18	98.58	82.28	69.90
	BIT-34	98.68	83.34	71.44
	BIT-50	98.67	83.40	71.53
	BIT-101	98.60	82.53	70.26
	TransUNetCD	98.66	83.63	71.86
	SwinSUNet	98.92	85.60	74.82
	CTDFormer	98.40	80.30	67.09
$\mathcal{DDPM}$	DDPM-CD	98.44	84.85	76.43
$\mathcal{M}$	MambaBCD-Tiny	99.03	88.04	78.63

5. Additional qualitative results

5.1. LEVIR-CD dataset

Figure 9, 10, 11, and 12 show additional qualitative results on LEVIR-CD dataset.

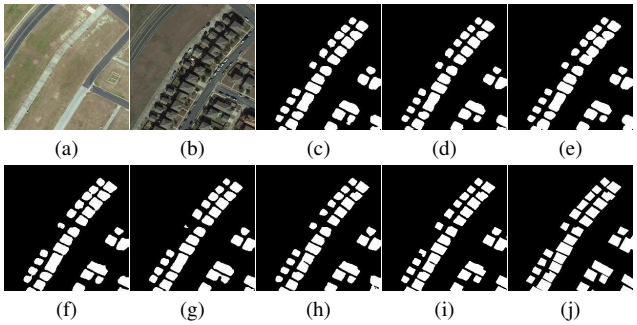

Figure 9. Comparison of different state-of-the-art CD methods on **LEVIR-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

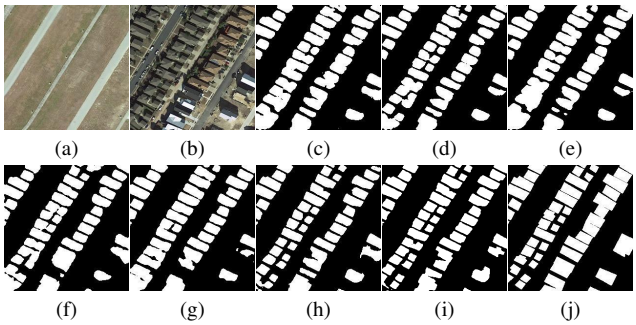

Figure 10. Comparison of different state-of-the-art CD methods on **LEVIR-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

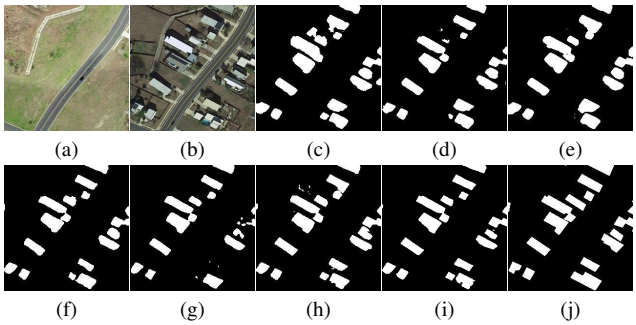

Figure 11. Comparision of different state-of-the-art CD methods on **LEVIR-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

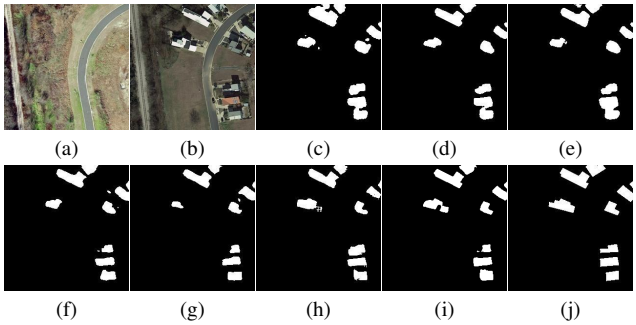

Figure 12. Comparision of different state-of-the-art CD methods on **LEVIR-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

5.2. WHU-CD dataset

Figure 13, 14, 15, 16 and 17 show additional qualitative results on WHU-CD dataset.

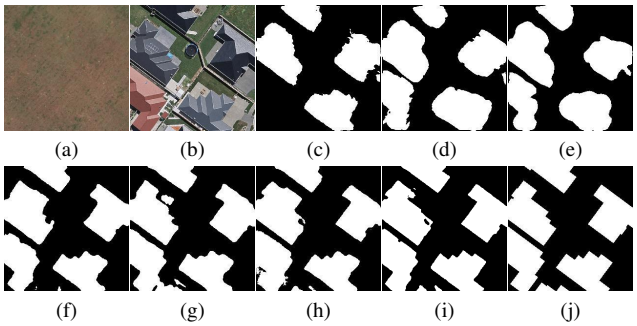

Figure 13. Comparision of different state-of-the-art CD methods on **WHU-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

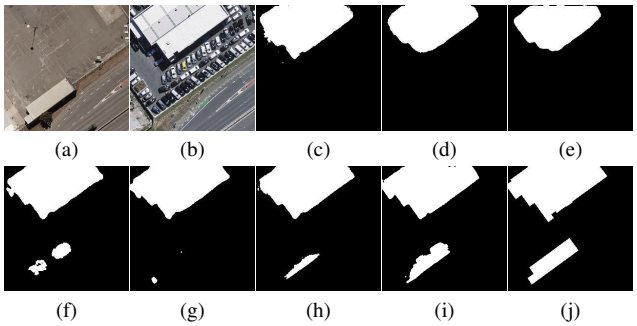

Figure 14. Comparision of different state-of-the-art CD methods on **WHU-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

5.3. DSIFN-CD dataset

Figure 18, 19, 20 and 21 show additional qualitative results on LEVIR-CD dataset.

5.4. CDD dataset

Figure 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31 show additional qualitative results on LEVIR-CD dataset.

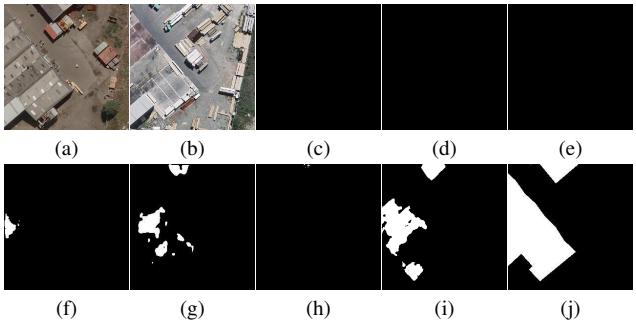

Figure 15. Comparision of different state-of-the-art CD methods on **WHU-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

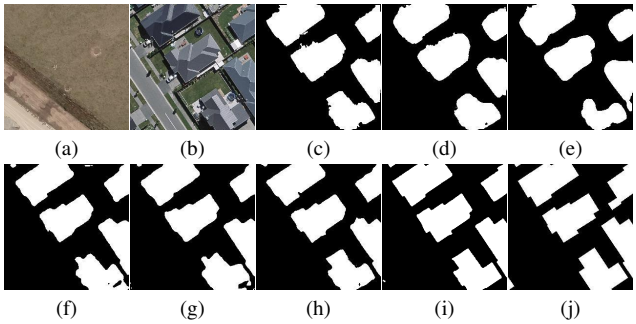

Figure 16. Comparision of different state-of-the-art CD methods on **WHU-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

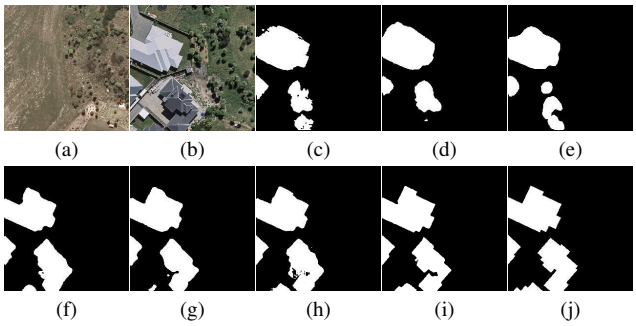

Figure 17. Comparision of different state-of-the-art CD methods on **WHU-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

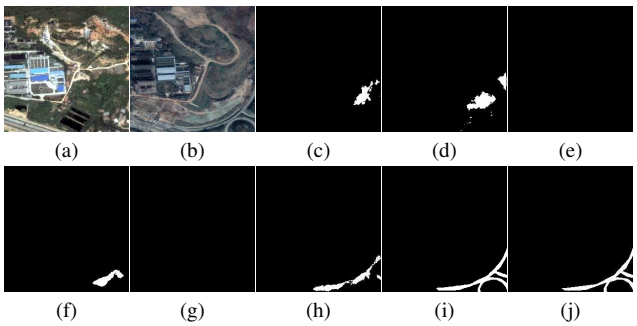

Figure 18. Comparison of different state-of-the-art CD methods on **DSIFN-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

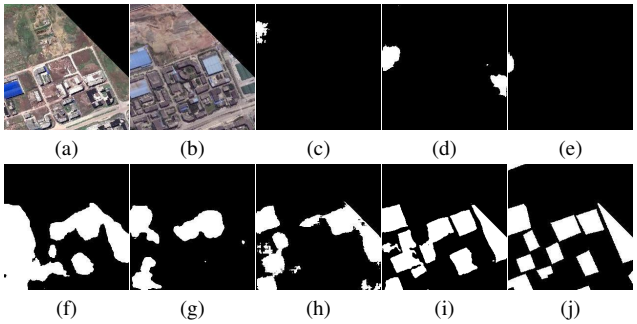

Figure 19. Comparison of different state-of-the-art CD methods on **DSIFN-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

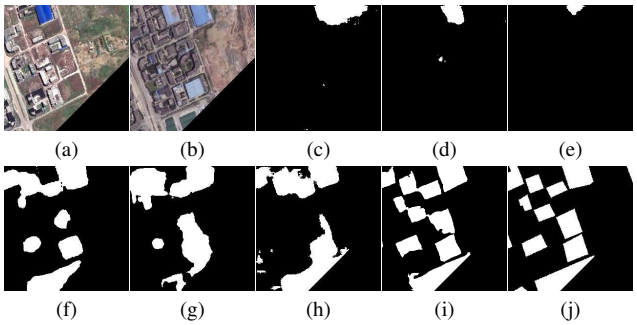

Figure 20. Comparison of different state-of-the-art CD methods on **DSIFN-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

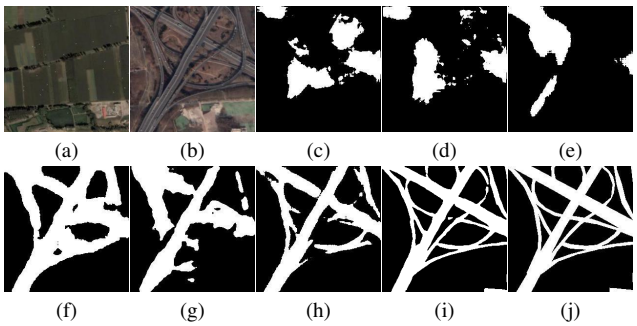

Figure 21. Comparison of different state-of-the-art CD methods on **DSIFN-CD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

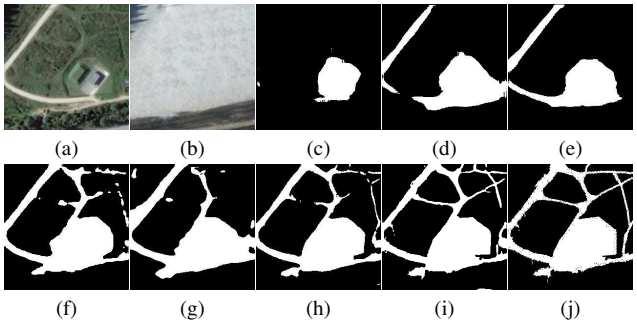

Figure 22. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

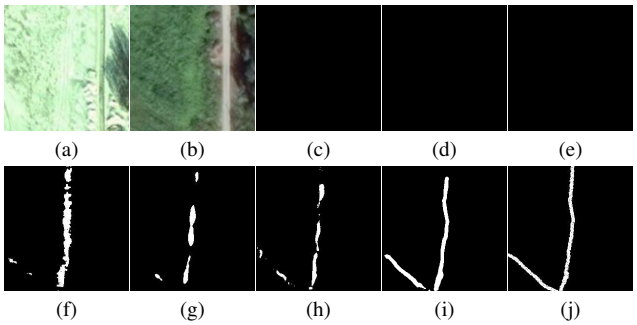

Figure 23. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

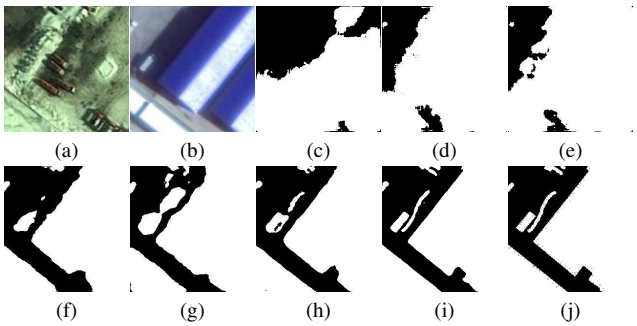

Figure 24. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

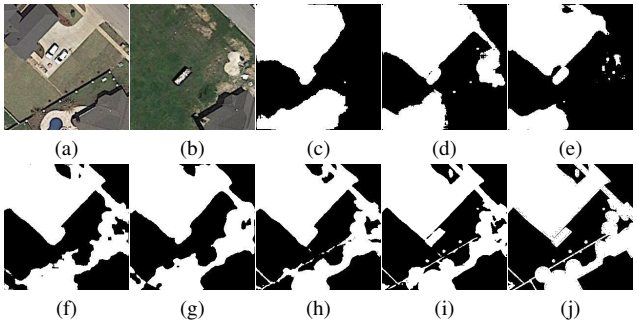

Figure 25. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

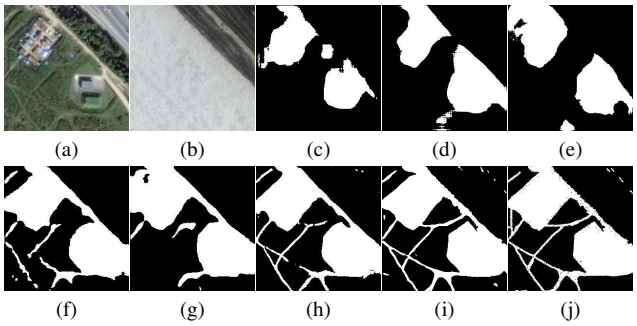

Figure 26. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

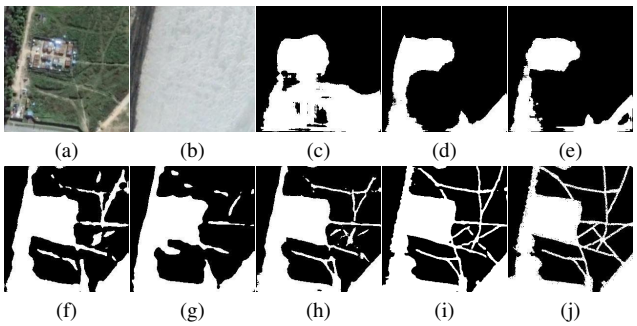

Figure 27. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

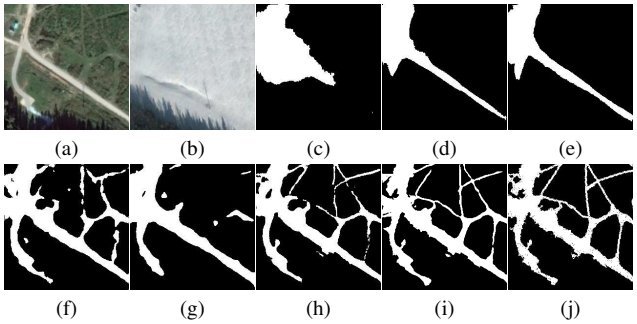

Figure 28. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

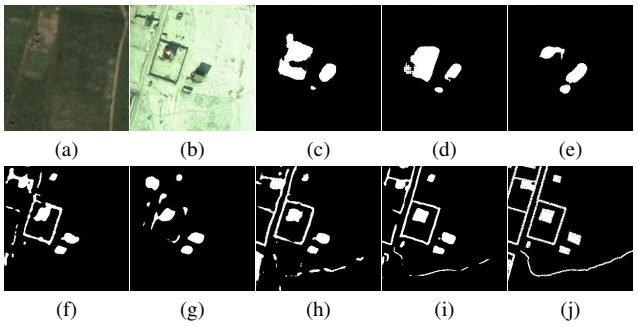

Figure 29. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

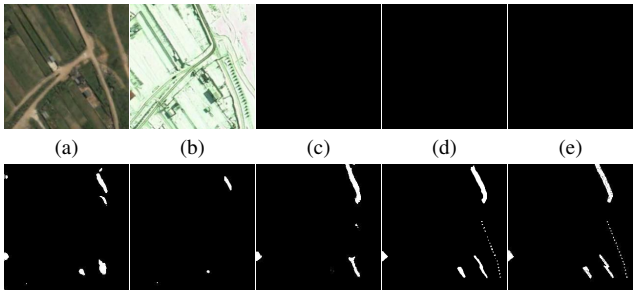

Figure 30. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

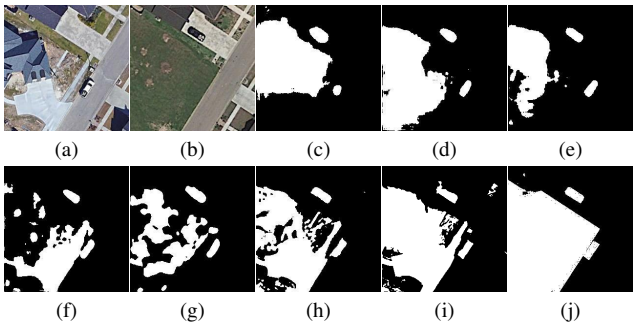

Figure 31. Comparison of different state-of-the-art CD methods on **CDD dataset**: **(a)** Pre-change image, **(b)** Post-change image, **(c)** FC-EF [13], **(d)** FC-Siam-Di [13], **(e)** FC-Siam-Conc [13], **(f)** DT-SCN [27], **(g)** BIT [6], **(h)** ChangeFormer [4], **(i)** ddpm-CD (*ours*), and **(j)** Ground-truth.

References

[1] Anju Asokan and J Anitha. Change detection techniques for remote sensing applications: a survey. *Earth Science Informatics*, 12(2):143–160, 2019. 1

[2] Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. *ICCV*, 2021. 2

[3] Wele Gedara Chaminda Bandara and Vishal M Patel. Revisiting consistency regularization for semi-supervised change detection in remote sensing images. *arXiv preprint arXiv:2204.08454*, 2022. 1

[4] Wele Gedara Chaminda Bandara and Vishal M Patel. A transformer-based siamese network for change detection. *arXiv preprint arXiv:2201.01293*, 2022. 2, 8, 9, 10, 11, 12

[5] Keumgang Cha, Junghoon Seo, and Yeji Choi. Contrastive multiview coding with electro-optics for sar semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021. 6

[6] Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 2021. 2, 8, 9, 10, 11, 12

[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. 6

[8] Jie Chen, Ziyang Yuan, Jian Peng, Li Chen, Haozhe Huang, Jiawei Zhu, Yu Liu, and Haifeng Li. Dasnet: Dual attentive fully convolutional siamese networks for change detection in high-resolution satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1194–1206, 2021. 1

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 6

[10] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 2

[11] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 6

[12] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6172–6180, 2018. 2

[13] Rodrigo Caye Daudt, Bertr Le Saux, and Alexandre Boulch. Fully convolutional siamese networks for change detection. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 4063–4067. IEEE, 2018. 1, 8, 9, 10, 11, 12

[14] JS Deng, K Wang, YH Deng, and GJ Qi. Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16):4823–4838, 2008. 1

[15] M Dharani and G Sreenivasulu. Land use and land cover change detection by using principal component analysis and morphological operations in remote sensing applications. *International Journal of Computers and Applications*, 43(5):462–471, 2021. 1

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[17] Song Chen Hao Chen, Wenyuan Li and Zhenwei Shi. Semantic-aware dense representation learning for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–18, 2022. 7

[18] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *arXiv preprint arXiv:1911.05722*, 2019. 2

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2

[20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 1

[22] 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. 6

[23] Jian Kang, Ruben Fernandez-Beltran, Puhong Duan, Sicong Liu, and Antonio J. Plaza. Deep unsupervised embedding for remotely sensed images based on spatially augmented momentum contrast. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):2598–2610, 2021. 2

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional

neural networks. *Advances in neural information processing systems*, 25, 2012. 2

[25] 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. 6

[26] Wenyuan Li, Keyan Chen, Hao Chen, and Zhenwei Shi. Geographical knowledge-driven representation learning for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2021. 2

[27] Yi Liu, Chao Pang, Zongqian Zhan, Xiaomeng Zhang, and Xue Yang. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geoscience and Remote Sensing Letters*, 18(5):811–815, 2020. 8, 9, 10, 11, 12

[28] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[29] Yang Long, Gui-Song Xia, Shengyang Li, Wen Yang, Michael Ying Yang, Xiao Xiang Zhu, Liangpei Zhang, and Deren Li. On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-aid. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4205–4230, 2021. 2

[30] Dengsheng Lu, Paul Mausel, Eduardo Brondizio, and Emilio Moran. Change detection techniques. *International journal of remote sensing*, 25(12):2365–2401, 2004. 1

[31] Luigi Tommaso Luppino, Filippo Maria Bianchi, Gabriele Moser, and Stian Normann Anfinsen. Unsupervised image regression for heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9960–9975, 2019. 1

[32] Oscar Mañas, Alexandre Lacoste, Xavier Giro-i Nieto, David Vazquez, and Pau Rodriguez. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9414–9423, 2021. 2, 6

[33] Allan Aasbjerg Nielsen. The regularized iteratively reweighted mad method for change detection in multi- and hyperspectral data. *IEEE Transactions on Image processing*, 16(2):463–478, 2007. 1

[34] Allan A Nielsen, Knut Conradsen, and James J Simpson. Multivariate alteration detection (mad) and maf postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sensing of Environment*, 64(1):1–19, 1998. 1

[35] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone. Unsupervised deep change vector analysis for multiple-change detection in vhr images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6):3677–3693, 2019. 1

[36] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[37] Gencer Sumbul, Marcela Charfuelan, Begüm Demir, and Volker Markl. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 5901–5904. IEEE, 2019. 2

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[39] Stefano Vincenzi, Angelo Porrello, Pietro Buzzega, Marco Cipriano, Pietro Fronte, Roberto Cuccu, Carla Ippoliti, Annamaria Conte, and Simone Calderara. The color out of space: learning self-supervised representations for earth observation imagery. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3034–3041. IEEE, 2021. 2

[40] Decheng Wang, Xiangning Chen, Mingyong Jiang, Shuhan Du, Bijie Xu, and Junda Wang. Ads-net:an attention-based deeply supervised network for remote sensing image change detection. *International Journal of Applied Earth Observation and Geoinformation*, 101:102348, 2021. 1

[41] Di Wang, Jing Zhang, Bo Du, Gui-Song Xia, and Dacheng Tao. An empirical study of remote sensing pretraining. *IEEE Transactions on Geoscience and Remote Sensing*, 2022. 2

[42] Xinlong Wang, Rufeng Zhang, Chunhua Shen, Tao Kong, and Lei Li. Dense contrastive learning for self-supervised visual pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3024–3033, 2021. 6

[43] Cui Zhang, Liejun Wang, Shuli Cheng, and Yongming Li. Swinsunet: Pure transformer network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. 2

[44] Chenxiao Zhang, Peng Yue, Deodato Tapete, Liangcun Jiang, Boyi Shangguan, Li Huang, and Guangchao Liu. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:183–200, 2020. 1

[45] Chenxiao Zhang, Peng Yue, Deodato Tapete, Liangcun Jiang, Boyi Shangguan, Li Huang, and Guangchao Liu. A deeply supervised image fusion network for change detection in high resolution bitemporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:183–200, 2020. 6

[46] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017. 2