Concatenated Deep-Learning Framework for Multitask Change Detection of Optical and SAR Images

Zhengshun Du, Xinghua Li, *Senior Member, IEEE*, Jianhao Miao, Yanyuan Huang, Huanfeng Shen, *Senior Member, IEEE*, and Liangpei Zhang, *Fellow, IEEE*

***Abstract*—Optical and synthetic aperture radar (SAR) images provide complementary information to each other. However, the heterogeneity of same-ground objects brings a large difficulty to change detection (CD). Correspondingly, transformation-based methods are developed with two independent tasks of image translation and CD. Most methods only utilize deep learning for image translation, and the simple cluster and threshold segmentation leads to poor CD results. Recently, a deep translation-based CD network (DTCDN) was proposed to apply deep learning for image translation and CD to improve the results. However, DTCDN requires the sequential training of the two independent subnetwork structures with a high computational cost. Toward this end, a concatenated deep-learning framework, multitask change detection network (MTCDN), of optical and SAR images is proposed by integrating the CD network into a complete generative adversarial network. This framework contains two generators and discriminators for optical and SAR image domains. Multitask refers to the combination of image identification by discriminators and CD based on an improved UNet++. The generators are responsible for image translation to unify the two images into the same feature domain. In the training and prediction stages, an end-to-end framework is realized to reduce cost. The experimental results on four optical and SAR datasets prove the effectiveness and robustness of the proposed framework over eight baselines.**

Manuscript received 28 October 2023; accepted 14 November 2023. Date of publication 17 November 2023; date of current version 4 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42171302, in part by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, under Grant KF-2021-06-003, in part by the Open Research Project of the Hubei Key Laboratory of Intelligent Geo-Information Processing under Grant KLIGIP-2021A02, in part by the Open Fund of Hubei Luojia Laboratory under Grant 220100055, in part by the Fundamental Research Funds for the Central Universities under Grant 2042022dx0002, Grant 2042023kfyq04, and Grant 2042023kf0197, and in part by the Knowledge Innovation Program of Wuhan–Shuguang Project under Grant 2023010201020242. *(Corresponding author: Xinghua Li.)*

Zhengshun Du, Jianhao Miao, and Yanyuan Huang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: maskedzs@126.com; jianhao_miao@whu.edu.cn; huang_yy@whu.edu.cn).

Xinghua Li is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China, also with the Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034, China, and also with the Hubei Luojia Laboratory, Wuhan 430079, China (e-mail: lixinghua5540@whu.edu.cn).

Huanfeng Shen is with the School of Resource and Environment Sciences, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

Liangpei Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: zlp62@whu.edu.cn).

The code is available at https://github.com/lixinghua5540/MTCDN.

Digital Object Identifier 10.1109/JSTARS.2023.3333959

***Index Terms*—Change detection (CD), deep learning, generative adversarial network (GAN), multitask network, optical and synthetic aperture radar (SAR) images.**

I. Introduction

CHANGE detection (CD), a classical and popular topic in remote sensing, utilizes multitemporal images of the same geographical region to identify the difference between ground objects [1]. It has been widely applied to different fields, such as urban planning [2], [3], resource survey [4], [5], [6], and natural disaster management [7], [8], [9]. Optical images contain rich spectral and spatial information and become an important data source of CD [10]. However, these images are limited by poor atmospheric conditions. Synthetic aperture radar (SAR) images can penetrate clouds and fogs to collect ground information but are influenced by speckle noises [11]. Therefore, combining the two types of images not only realizes the complementarity of information but is also beneficial for CD when homogeneous images are unavailable [12]. The challenges of heterogeneous CD can be concluded as follows.

1) Heterogeneous features between images: Given the different imaging mechanisms, the same-ground objects reflect the disparate features on each type of image. As a result, they are easily recognized as fake changed objects. However, this heterogeneity and the characteristics of unchanged areas lead to the difficulty of heterogeneous CD.
2) Incomplete objects and broken results: The difference image (DI) method inherited from the traditional CD of a single source is still widely used in heterogeneous images. DI is usually obtained by the difference or ratio operation between bitemporal images. The change map is generated from DI by a threshold and cluster method. The drawback is that it cannot be against noise, resulting in discontinuous ground objects and low accuracy.
3) Difficult generalization and explanation: Each type of image is treated individually on heterogeneous CD, which includes optical, SAR, and light detection and ranging images. Nevertheless, the quantity of sensors has been increasing due to the development of space-based technology, bringing a challenge to CD. Therefore, a general solution becomes critical for multisource CD.

Some heterogeneous CD studies have been carried out recently to solve the above problems. These methods can be

divided into traditional parameter-, classification-, invariant similarity measure-, and transformation-based methods.

1) Traditional parameter-based methods, such as Kullback–Leibler-based change measures [13], local joint distributions [14], and the Markov model for multimodal CD [15], apply multivariate or Gaussian distributions to extract relevant change. However, these methods are difficult to generalize.
2) Classification-based methods, such as multitemporal segmentation and mixed classification [16] and postclassification comparison [17], [18], obtain change information from previous classification or segmentation maps.
3) Invariant similarity measure-based methods use the similarity of pixels or objects in heterogeneous images to recognize change quantitatively, including similarity map estimation [19], improved nonlocal patch-based graph (INLPG) [20], [21], adaptive local structure consistency [22], patch similarity graph matrix [23], and pixel pair algorithm [24], [25].
4) Transformation-based methods translate heterogeneous or multimodal images to the same domain to facilitate image comparison. These methods are divided into traditional and deep-learning-based transformations.

Multivalue estimation methods with noise tolerance are used for traditional transformation, such as homogeneous pixel transformation (HPT) [26]. Deep-learning-based transformation for heterogeneous CD obtains exciting results, such as symmetric deep convolutional coupling network (SCCN) [27], conditional generative adversarial network (cGAN) [28], X-Net, ACE-Net [29], and deep translation-based CD network (DTCDN) [30]. The translated images are compared in the same feature space to obtain the accurate CD results. This type of method, with its ease of generalization and explanation, is a promising direction for future research.

According to DTCDN [30], the heterogeneous CD based on deep-learning transformation can be separated into two processes, i.e., image translation and CD. In some recently advanced methods, deep learning especially GAN is widely used for image translation. SCCN extracts feature maps from optical and SAR images through two weight-shared encoding networks to obtain a differential image. For cGAN, fake SAR images are generated by the translation network, and the original SAR images are adjusted by an approximation network. X-Net and ACE-Net take the affinity information as prior knowledge and are trained by the cyclic GAN. Although these methods can generate high-similarity images by the translation network, the translated image is directly exerted by the simple cluster and threshold segmentation, which cannot avoid noise. By contrast, some studies directly used the CD network without translation for heterogeneous images [31], [32]. A CD network can also learn some latent mappings between heterogeneous images. Nevertheless, it becomes difficult when the images have a big diversity in imaging patterns, such as optical and SAR images. The image translation has been demonstrated to improve the performance of CD. DTCDN is constructed by deep image translation and CD network and shows outstanding performance in heterogeneous images. However, the implementation of DTCDN is divided into two parts and its training process requires intermediate intervention, which is time-consuming. To address this problem, MTCDN, a novel fused framework for image translation and CD, is proposed. The main contributions of this article focus on the following aspects.

1) A novel MTCDN is proposed to realize end-to-end CD for optical and SAR images. This method fuses image generation, image identification, and CD in a GAN to make the multitask structure compact, thereby promoting efficiency and keeping accuracy.
2) A cyclic structure is constructed for the optical and SAR image translation task. By this translation, the main heterogeneous features are eliminated, and the changed features become evident for the accurate heterogeneous CD.
3) An improved Siamese UNet++, which can realize the function of classifying real or fake images and obtaining CD results, is used as a discriminator. A feature-enhanced model (FEM) is introduced to strengthen the connection between low- and high-level features.

The rest of this article is organized as follows. In Section II, MTCDN is described in detail. The experiments and results are presented in Section III followed by the comparisons and analyses in Section IV. Finally, Section V concludes this article.

II. Methodology

The details of the proposed MTCDN and how it solves the heterogeneous CD problem are introduced in this section. As generative model emerges in the field of computer vision, the applications of remote sensing have been greatly improved, including normalization [33], CD [30], and super-resolution [34], [35]. Our proposed MTCDN also takes the advantage of generative model CycleGAN [36]. As shown in Fig. 1, the framework of MTCDN displays a translation mode of optical-to-SAR images and CD. This framework contains three tasks, i.e., image generation, image identification, and CD. Specifically, an optical image x is imported to two generations G_x and G_y in domains X and Y, respectively, to produce the reconstruction optical image x_r and the translated SAR image y_t. Then, y_t and original SAR image y in the same feature space are fed into the discriminator of domain Y (D_y) for image identification and CD. Besides, G_x translates y_t back to the optical domain and generates the cycle optical x_c. After that one loop is completed, improved results are achieved by repeating the process. Another translation mode of SAR-to-optical has a similar process. Next, the multitask discriminator D and generator G of MTCDN will be described in detail.

A. Multitask Discriminator of MTCDN

CycleGAN has been widely applied to image translation task because of the cycle-consistent loss. However, the CycleGAN consists of complex different parts, which include an encoder, generator, and discriminator. Nice-GAN is used to solve this problem by blending the encoder in the discriminator to combine classification and encoding [37]. Recently, multitask methods based on GAN are gradually applied to remote sensing. An advanced multitask GAN was proposed for hyperspectral image

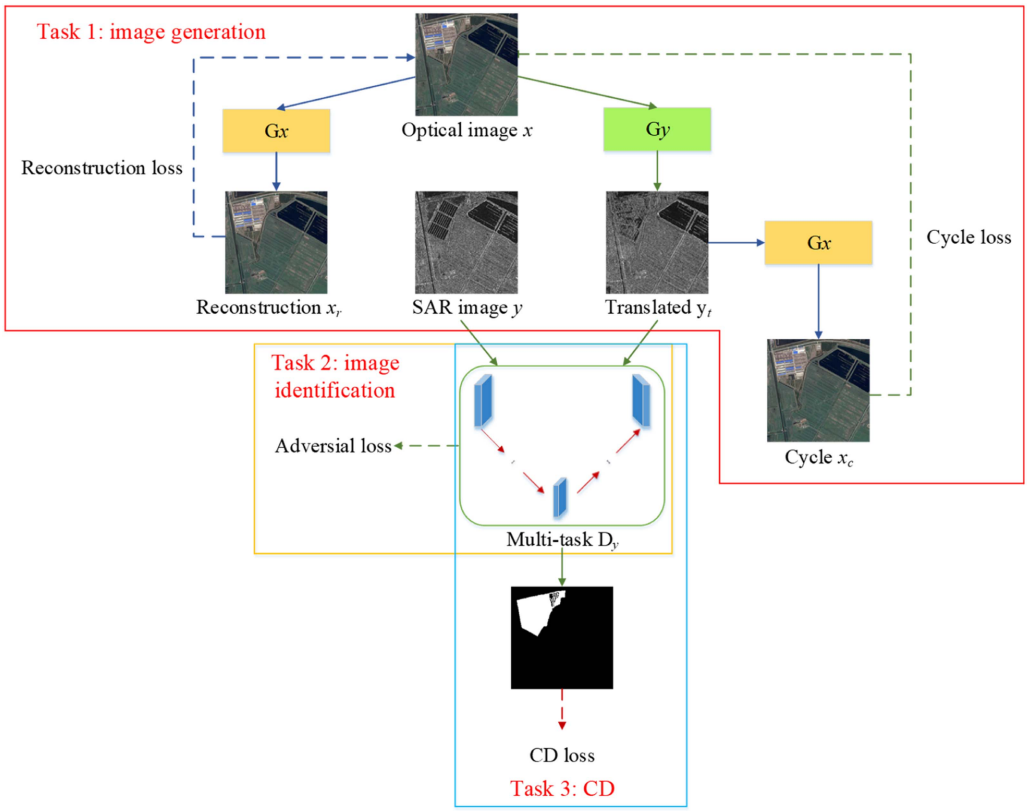


Fig. 1. General framework of the proposed MTCDN method, including image generation, image identification, and CD.

classification [38]. The combination of GAN, pose estimator, and auxiliary classifier is used for SAR image generation [39]. These methods have extended the limitation of single-task GAN. Accordingly, a multitask discriminator for CD and image identification is constructed. This discriminator is improved from SNUNet [40], and the network structure is shown in Fig. 2. A Siamese network is adopted because the discriminator needs to classify two-group images in MTCDN. SNUNet contains an encoder to extract rich semantic information and a decoder to concatenate and integrate different-level features. Besides, FEM and ensemble channel attention module (ECAM) are introduced to strengthen features. These modules are presented in the following sections.

1) Image Identification: The image identification aims to distinguish whether the image imported to the discriminator is real or fake. As depicted in Fig. 2, $A^{i,j}$ denotes the features in the ith depth and jth nested layers with skip connection. The two types of images, i.e., an original T_1 image of $A_1^{0,0}$ and a translated T_2 image $A_2^{0,0}$ in the same domain, are fed into the discriminator. $A_1^{0,0}$ and $A_2^{0,0}$ are operated by a series of convolutions and downsamplings to extract deep features. The filter sizes are set to {32, 64, 128, 256, 512}. The image identifications are connected to the last depth features of the encoder. $A_1^{4,0}$ and $A_2^{4,0}$ are imported into a convolution layer with a kernel size of 4×4 and padding to generate features with one channel. The sigmoid activation function is used to obtain the outputs. Then, L_{Da} is calculated by the labels of real or fake and identification outputs.

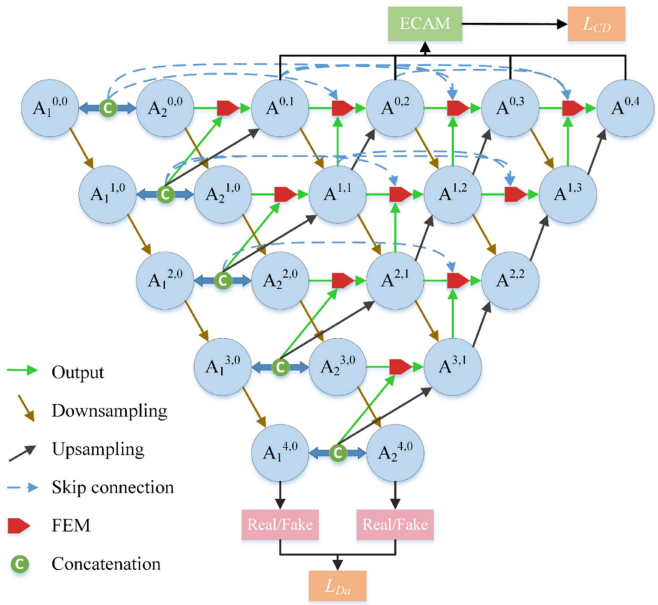


Fig. 2. Overview of the structure of the multitask discriminator.

2) Feature-Enhanced Model: The nested dense blocks at the same level in UNet++ are designed to generate rich semantic information. However, the different-level information is only connected by upsampling, which limits the further fusion of high- and low-level features. Li et al. [41] utilized an attention

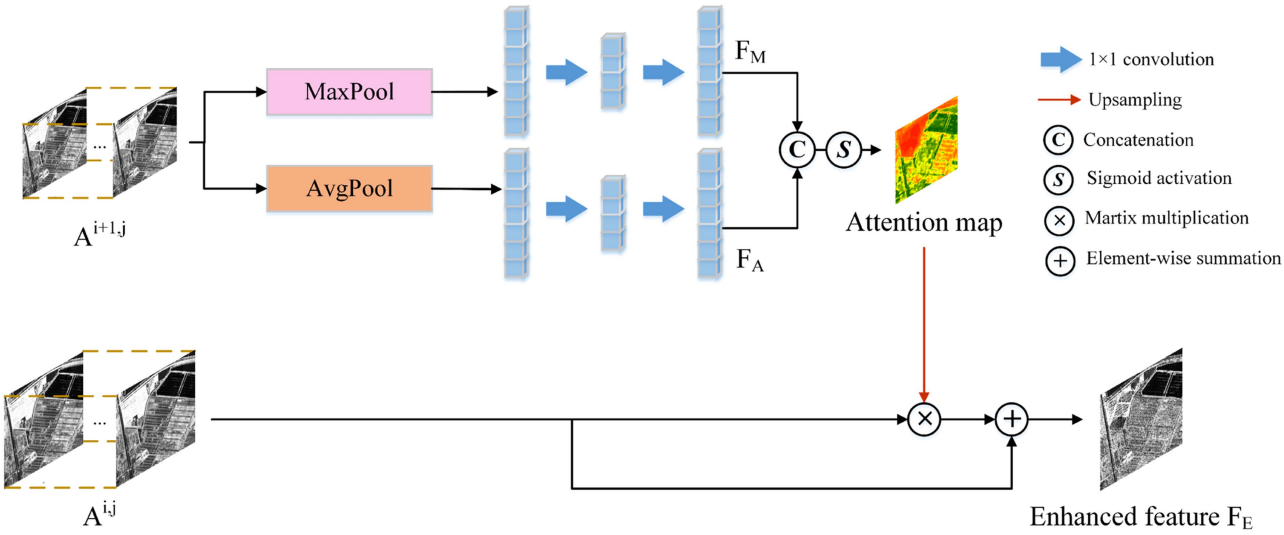


Fig. 3. Architecture of FEM.

gate to improve the propagation of semantic information while combining the upsampling and encoder features at the first stage. Therefore, an FEM between adjacent depth features is proposed for UNet++ to localize and focus on changed areas and inhibit irrelevant areas. The architecture of FEM is shown in Fig. 3. $A^{i,j}$ in the ith depth and $A^{i+1,j}$ in the next depth are taken as input. $A^{i+1,j}$ is operated by max and average pooling operations followed by two successive 1×1 convolution, respectively, to obtain two local features F_M and F_A. Then, the attention map is generated by concatenating them in channel dimension. After the attention map upsamples to double sizes, an elementwise multiplication combined with an addition operation is utilized to acquire enhanced feature F_E. The formulations of F_M, F_A, and F_E are listed in (1)–(3), respectively. The features reflect the relevant information of change areas by FEM

$$F_M = \text{Conv}\left(\text{MaxPool}\left(A^{i+1,j}\right)\right) \tag{1}$$

$$F_A = \text{Conv}\left(\text{AvgPool}\left(A^{i+1,j}\right)\right) \tag{2}$$

$$F_E = A^{i,j} + \left(A^{i,j} \otimes \text{Up}\left(\sigma\left(\text{Concat}\left(F_M, F_A\right)\right)\right)\right) \tag{3}$$

where MaxPool and AvgPool are the max pooling and average pooling, respectively, Conv represents two successive 1×1 convolutional operation, σ represents the sigmoid function, $\otimes$ means the operation of matrix multiplication, Concat is the operation of concatenation, and Up is the upsampling operation.

3) Ensemble Channel Attention Module: ECAM [40] is introduced for fusing the output of top layers because they contain different-depth contextual information of U-Net. Compared with multiple-side outputs fusion [42], which only uses a simple concatenation, ECAM assigns corresponding weight to each block, focusing on efficient parts to promote CD results. It can be described by the following equations:

$$\text{CAM} = \sigma(\text{MLP}\left(\text{MaxPool}\left(X\right)\right) + \sigma(\text{MLP}\left(\text{AvgPool}\left(X\right)\right) \tag{4}$$

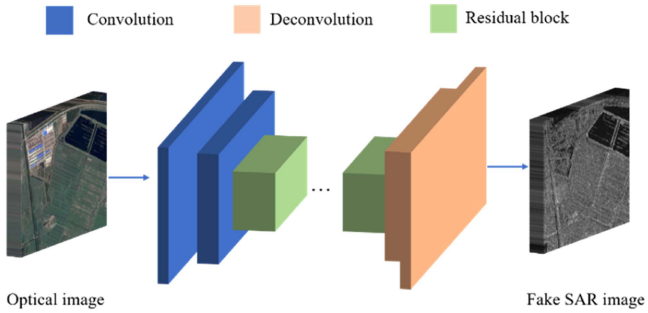


Fig. 4. Structure of generator in MTCDN.

$$A^{0,5} = \left[A^{0,1}, A^{0,2}, A^{0,3}, A^{0,4}\right] \tag{5}$$

$$\text{ECAM} = (A^{0,5} + \boldsymbol{repeat}_{(4)}(\text{CAM}(A^{0,1} + A^{0,2} + A^{0,3} + A^{0,4}))) \otimes A^{0,5} \tag{6}$$

where σ means the sigmoid activation function, and $\boldsymbol{repeat}_{(k)}$ denotes the operation of repeating the attention map k times and concatenating in the channel dimension.

B. Generator of MTCDN

The generator learns the mapping between the original and target domains to produce images with the same resolution. The generator is based on an encoder–decoder with residual blocks in MTCDN (see Fig. 4). The translation mode of an optical-to-SAR image is drawn as an example. An optical image is inputted into two blocks of convolution and downsampling to extract global and local information in the encoder stage. Then, a series of residual blocks are utilized to obtain deep features. These blocks can extract high-level features and ensure accuracy with the network structure deepening, thereby improving efficiency and robustness. To recover the original resolution, high-level features are fed to the decoder with two deconvolution blocks.

The modes of optical-to-SAR and SAR-to-optical images are conducted simultaneously in the training stage.

C. Loss Function

The loss functions of MTCDN can be divided into two main patterns, i.e., L_G and L_D, which represent the losses of generator and discriminator, respectively. As shown in Fig. 1, one translation mode predominantly contains three different losses, i.e., adversarial, reconstruction, and cycle losses. The final loss is the sum of two translation modes, which are defined by L_{Ga} (7), L_{recon} (8), and L_{cyc} (9), respectively

$$L_{Ga} = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[1 - D_y(G_y(x))^2\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[1 - D_x(G_x(y))^2\right] \tag{7}$$

$$L_{\text{recon}} = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\|G_x(x) - x\|_1\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\|G_y(y) - y\|_1\right] \tag{8}$$

$$L_{\text{cyc}} = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\|G_x(G_y(x)) - x\|_1\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\|G_y(G_x(y)) - y\|_1\right] \tag{9}$$

$$L_G = L_{Ga} + \lambda_1 L_{\text{recon}} + \lambda_2 L_{\text{cyc}} \tag{10}$$

where $G_x(x)$ is the generated image in domain X through G_x, and $D_x(x)$ attempts to distinguish image x in domain X. $x \sim p_{\text{data}}(x)$ and $y \sim p_{\text{data}}(y)$ are the image distribution probabilities, and $\|\ \|_1$ is the l_1 norm. λ_1 and λ_2 are the weights of L_{recon} and L_{cyc}, respectively, in total loss L_G.

The multitask discriminator combines image identification and CD into one network. Therefore, the loss of discriminator L_D contains the adversarial loss L_{Da} and CD loss L_{CD}. To our knowledge, CD image sources usually consist of unchanged pixels and few changed pixels, leading to class imbalance. The focal loss [43] is introduced to focus on the class of change to address the issue. Three losses are expressed as follows:

$$L_{Da} = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[D_y(G_y(x))^2\right] + \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[1 - D_x(x)^2\right] \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[D_x(G_x(y))^2\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[1 - D_y(y)^2\right] \tag{11}$$

$$L_{\text{CD}} = -\alpha_t(1 - p_t)^{\gamma}\log(p_t) \tag{12}$$

$$L_D = L_{Da} + \beta L_{\text{CD}} \tag{13}$$

where α_t is the weighting factor, γ is the tunable focusing factor, p_t is the output probability of ECAM in location t, and β is the weight of L_{CD}. To obtain the desired results, L_{Da} and L_{CD} should be as small as possible.

During the training process of MTCDN, the adversarial loss in initial epochs does not decrease to the ideal stage, and the quality of generated images may not support a high-accuracy CD. Therefore, a stepwise training strategy is used in which β is set on different stages to control the performance of CD. In a word, a low β is set in the initial iteration, and β increases step-by-step. Therefore, the discriminator obtains improved generated images to guide the CD task.

(a) (b) (c)

Fig. 5. Shuguang village dataset. (a) Pre-event image. (b) Postevent image. (c) Ground truth. White areas are changed in (c).

III. Experiments and Results

In this section, optical and SAR datasets are illustrated. The experiments were performed on four datasets separately. Eight baselines were selected for comparison to prove the effectiveness and robustness of MTCDN. They include five heterogeneous CD methods based on the deep-learning transformation, i.e., SCCN, cGAN, ACE-Net, X-Net, and DTCDN. Besides, an unsupervised invariant similarity-based method INLPG [20] and a supervised traditional transformation-based approach HPT [26] were selected. At the same time, the results of four datasets are analyzed in detail. The experiment without translation was carried out to verify the effect of image translation. The CD results are presented in terms of both accuracy assessment and visualization.

A. Experimental Datasets

Four pairs of bitemporal optical and SAR images for experiments are presented, which came from Shuguang village [15], [44], Gloucester [15], [44], [45], and California [46]. Each dataset has a pair of optical and SAR images collected before and after an event. And the detailed information is shown in Table I.

The first dataset contains a building construction and river change in Shuguang village, China. The SAR image was captured by the Radarsat-2 in June 2008, as shown in Fig. 5(a). An optical image [see Fig. 5(b)] was captured by Google Earth in September 2012. The image size is only 419×342. It is notable that due to the small size of the images, the cross-validation method was used in the experiments for the two main changed parts after clipping.

As shown in Fig. 6, a flood, as a change event, occurred in Gloucester, U.K. QuickBird 2 captured the optical image [see Fig. 6(a)] in July 2006, and TerraSAR-X provided the postevent SAR image [see Fig. 6(b)] in July 2007. The ground truth was generated manually.

As shown in Fig. 7, the third dataset also came from Gloucester, and the change event is a flood. The pre-event optical image [see Fig. 7(a)] was collected by SPOT in September 1999, and the postevent SAR image was captured by ERS-1 in November 2000. This image dataset was applied to the 2009 and 2010 IEEE GRSS Data Fusion Contest [45].

The fourth dataset, which came from California (USA) (https://sites.google.com/view/luppino/data), also describes a flood event. Different from the above datasets, the fourth dataset contains a SAR image with three bands [see Fig. 8(b)] acquired by

TABLE I
DESCRIPTION OF FOUR OPTICAL AND SAR DATASETS

Dataset	Date	Image size	Changed event	Spatial resolution	Satellite
Shuguang	June 2008/September 2012	419*342	Building Construction	8 m	QuickBird 2/ TerraSAR-X
Gloucester I	July 2006/July 2007	2325*4135	Flooding	0.65 m	SPOT/ ERS-1
Gloucester II	September 1999/November 2000	1250*2600	Flooding	≈25 m	Landsat 8/ Sentinel 1-A
California	January 2017/February 2017	2000*3500	Flooding	≈15 m	Radarsat-2/ Google Earth

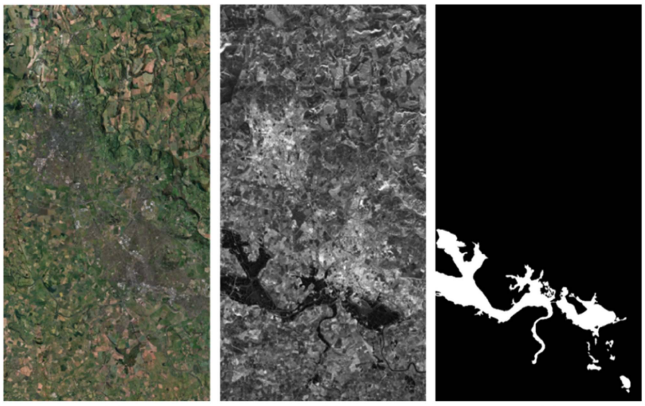


Fig. 6. Gloucester I dataset. (a) Pre-event image. (b) Postevent image. (c) Ground truth. White areas are changed in (c).

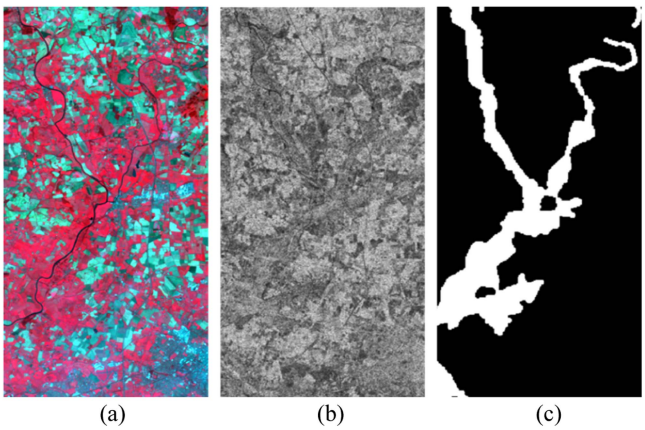


Fig. 7. Gloucester II dataset. (a) Pre-event image. (b) Postevent image. (c) Ground truth. White areas are changed in (c).

Sentinel 1-A. The SAR image was made up of the polarizations of VV and VH and their ratio band. The optical image [see Fig. 8(a)] was collected by the Landsat 8. The ground truth was obtained by applying a classic CD method between two single-polarization SAR images acquired at similar times as the Landsat 8 and Sentinel 1-A images.

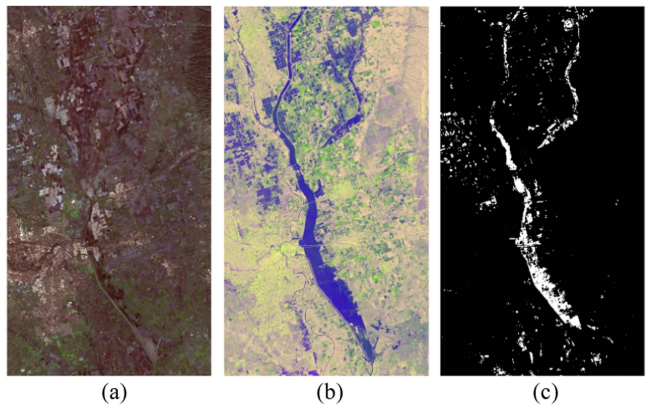


Fig. 8. California dataset. (a) Pre-event image. (b) Postevent image. (c) Ground truth. White areas are changed in (c).

B. Implementation Details and Metrics

1) Implementation Details: During our experiments, the model was trained with the PyTorch framework on an NVIDIA RTX 2080Ti (11G) GPU and Intel Core i7-9700 (32G) CPU. The number of residual blocks in the generator was set to 9. The learning rate of the generator and discriminator was 0.0001 with 500 epochs. For L_G, $\lambda_1 = 5$ and $\lambda_2 = 10$ were chosen. The weight and focus tunable factors α and γ were 0.25 and 2, respectively. In the stepwise training strategy, β was set to 0.1 in the first 20 iterations and increased to 0.5 and 2.5 after 20 and 40 iterations, respectively.

2) Evaluation Metrics: For the assessment of results, the criteria from the two-class confusion matrix, which includes precision (Precision), recall (Recall), $F1$ score ($F1$), and overall accuracy (OA), were applied

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) \tag{14}$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) \tag{15}$$

$$F1 = 2{*}\text{Precision} \times \text{Recall}/(\text{Precision}+\text{Recall}) \tag{16}$$

$$\text{OA} = (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \tag{17}$$

where the true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) are the ratings from the two-class confusion matrix.

TABLE II
QUANTITATIVE EVALUATIONS ON SHUGUANG DATASET

Methods	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
INLPG [20]	39.90	90.01	55.29	90.40
HPT [26]	51.81	89.99	65.76	96.25
SCCN [27]	30.21	43.42	35.62	89.97
cGAN [28]	54.03	87.23	66.73	95.52
ACE-Net [29]	79.17	75.87	75.34	97.13
X-Net [29]	68.19	75.82	71.80	96.37
Translation$^-$	80.15	86.16	84.11	97.34
DTCDN [30]	92.92	90.25	91.56	**99.75**
MTCDN^S	88.54	**91.43**	89.86	98.03
MTCDN^O	**93.64**	88.75	**92.69**	99.32

The best result of each metric is highlighted in bold.

TABLE III
QUANTITATIVE EVALUATIONS ON GLOUCESTER I DATASET

Methods	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
INLPG [20]	28.51	61.27	38.91	80.70
HPT [26]	31.75	65.18	42.68	82.46
SCCN [27]	59.74	61.68	60.70	91.99
cGAN [28]	33.84	68.68	45.34	83.42
ACE-Net [29]	65.65	70.73	68.10	93.35
X-Net [29]	66.26	78.75	71.97	93.85
Translation$^-$	87.89	80.45	84.01	96.93
DTCDN [30]	**89.96**	**89.93**	**89.95**	**97.98**
MTCDN^S	87.69	85.10	86.38	97.31
MTCDN^O	88.86	87.59	88.22	97.65

The best result of each metric is highlighted in bold.

TABLE IV
QUANTITATIVE EVALUATIONS ON GLOUCESTER II DATASET

Methods	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
INLPG [20]	42.48	42.47	42.48	75.36
HPT [26]	50.46	49.65	50.05	88.52
SCCN [27]	34.12	35.36	34.73	72.35
cGAN [28]	36.53	52.13	42.96	77.01
ACE-Net [29]	57.12	39.84	46.94	85.69
X-Net [29]	53.26	37.23	43.83	82.14
Translation$^-$	86.67	86.40	86.53	95.88
DTCDN [30]	**90.78**	86.66	88.67	96.33
MTCDN^S	76.27	84.30	80.08	93.06
MTCDN^O	89.49	**88.25**	**88.87**	**96.34**

The best result of each metric is highlighted in bold.

TABLE V
QUANTITATIVE EVALUATIONS ON CALIFORNIA DATASET

Methods	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
INLPG [20]	30.92	80.27	44.65	90.58
HPT [26]	53.95	51.32	52.60	95.62
SCCN [27]	31.11	54.01	39.47	84.55
cGAN [28]	32.43	59.20	41.91	92.82
ACE-Net [29]	30.59	76.49	43.71	91.38
X-Net [29]	35.34	59.11	44.24	90.83
Translation$^-$	57.59	67.03	61.95	96.32
DTCDN [30]	**66.73**	78.24	72.03	97.61
MTCDN^S	65.47	**80.61**	**72.26**	**97.83**
MTCDN^O	55.20	69.52	61.54	95.83

The best result of each metric is highlighted in bold.

C. Comparison to SOTA Methods

The quantitative evaluations of the Shuguang dataset are presented in Table II. In particular, translation$^-$ means without translation and only uses the CD network of MTCDN, i.e., original optical and SAR images are directly imported to the CD network. MTCDN^S and MTCDN^O indicate the optical-to-SAR and SAR-to-optical modes, respectively. MTCDN^O obtained the highest precision and $F1$-score, which was 1% higher than DTCDN approximately. By contrast, DTCDN, which included increased unchanged pixels, obtained the highest OA. MTCDN^O was better than the other mode, the reason might be the optical image could gather useful information. X-Net and ACE-Net introduced extra prior information to ensure good performance. SCCN performed the worst. In precision and $F1$, MTCDN^O and DTCDN outperformed translation$^-$ by more than 7%. Therefore, the image translation works well in heterogeneous CD and increases the accuracy.

The Gloucester I dataset describes the conditions before and after the flood. The results of MTCDN with two modes and eight other baselines are shown in Table III. Different from the Shuguang dataset, DTCDN achieved the best $F1$-score and OA. In terms of OA, MTCDN^O was lower than DTCDN by about 0.6%. The reason is obviously that two separate tasks had superiority because the image translation and CD cannot be influenced by each other on this dataset. Among these methods, INLPG was the worst. SCCN, X-Net, and ACE-Net showed superiority over the HPT-based traditional transformation method.

The third dataset, which was obtained from Gloucester about a flood disaster, is the same as the second dataset. The results of quantitative metrics are shown in Table IV. MTCDN^O showed its superiority over most methods, and its accuracy was higher than that of DTCDN. Moreover, MTCDN^O, which could discover changed pixels that were difficult to detect, had a slight improvement in recall. Compared with the Gloucester I dataset, HPT surpassed some deep-learning-based transformation methods, including cGAN, X-Net, and ACE-Net.

The fourth California dataset contains a SAR image with multipolarized three bands of VV, VH, and their ratio. The results of MTCDN and other SOTA methods are shown in Table V. The supervised HPT achieved better results compared with other unsupervised methods. MTCDN^S obtained competitive results on all evaluations, especially on the recall, which was over 80%, while DTCDN achieved the best precision. The $F1$ scores of DTCDN and MTCDN^S were higher than 70%. Different from

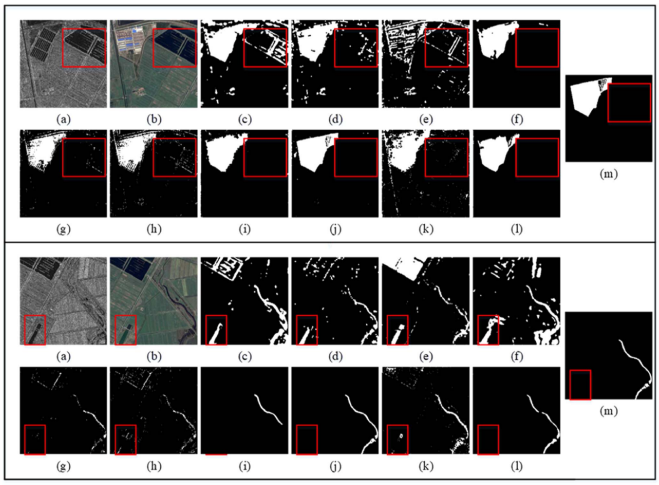


Fig. 9. Visualization of binary maps on the Shuguang dataset. (a) Pre-event image. (b) Postevent image. (c) INLPG [20]. (d) HPT [26]. (e) SCCN [27]. (f) cGAN [28]. (g) ACE-Net [29]. (h) X-Net [29]. (i) Translation^{-}. (j) DTCDN [30]. (k) MTCDN^{S}. (l) MTCDN^{O}. (m) Ground truth.

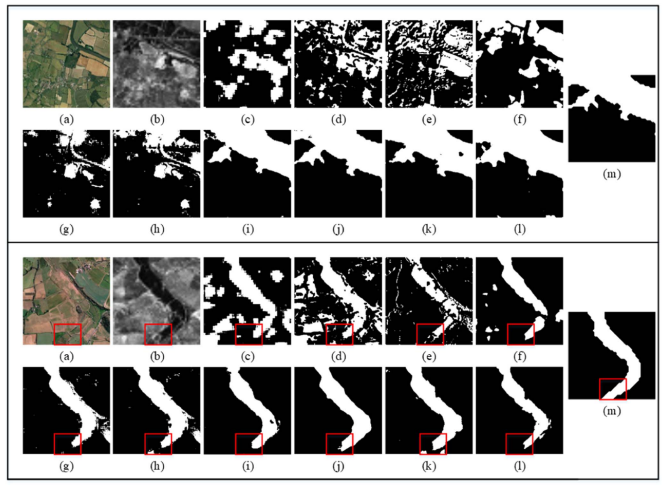


Fig. 10. Visualization of binary maps on the Gloucester I dataset. (a) Pre-event image. (b) Postevent image. (c) INLPG [20]. (d) HPT [26]. (e) SCCN [27]. (f) cGAN [28]. (g) ACE-Net [29]. (h) X-Net [29]. (i) Translation^{-}. (j) DTCDN [30]. (k) MTCDN^{S}. (l) MTCDN^{O}. (m) Ground truth.

the above three datasets, this dataset obtained a better assessment in the optical-to-SAR mode than the SAR-to-optical mode. Moreover, the output of the SAR-to-optical mode was worse than that without translation.

D. Visualization Results

The binary CD maps of the above approaches were also compared with MTCDN results fairly. The examples of CD visual results on the Shuguang dataset are displayed in Fig. 9. This dataset reflected a building and river change. Most methods could recognize the boundary of a building. Among these methods, unsupervised methods were limited by the DI drawback, which easily produced noise in unchanged areas. The supervised deep-learning-based methods DTCDN and MTCDN could suppress noise to improve the effect. Besides, the visual map MTCDN^{O} was better than another mode in MTCDN^{S}. MTCDN^{O} was not good at detecting the complete shape of buildings, and the result of DTCDN still had noise around the buildings. This finding also reflected that the precision of MTCDN^{O} was higher than that of DTCDN, while the recall was lower. For the change of river, MTCDN^{O} also showed its remarkable performance compared with most methods.

The examples of the Gloucester I dataset are shown in Fig. 10. The flood area usually covered a connected neighbor. However, INLPG, HPT, and SCCN had broken stripes and contained noises in many areas. Most results of MTCDN, DTCDN, and translation^{-} could detect the whole flood object and had improved visual effects. Similarly, their results of Gloucester II in Fig. 11 were integrated as a whole, which was consistent with the actual characteristics of the flood. Compared with DTCDN, the result of MTCDN^{O} mode was closer to the ground truth.

The California dataset is a middle-resolution remote sensing imagery. Some dispersed buildings are presented in Fig. 12. Visually, the SCCN mistakenly detected some buildings in unchanged areas. MTCDN^{S} could recognize many buildings but took some unchanged pixels as changed buildings. Compared with MTCDN, DTCDN did not cause too much misclassification but ignored some small buildings.

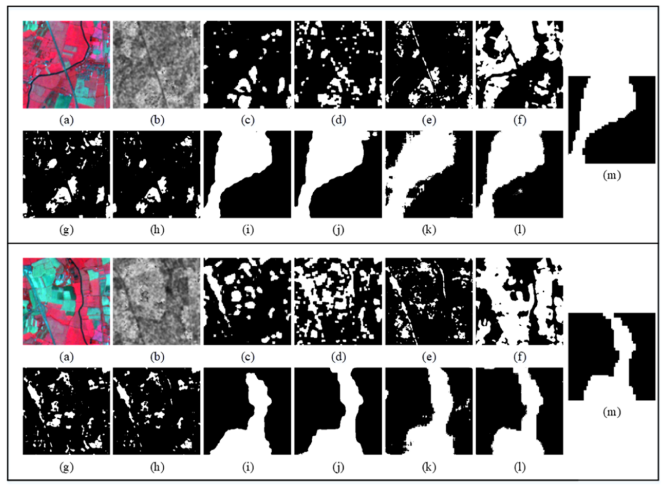


Fig. 11. Visualization of binary maps on the Gloucester II dataset. (a) Pre-event image. (b) Postevent image. (c) INLPG [20]. (d) HPT [26]. (e) SCCN [27]. (f) cGAN [28]. (g) ACE-Net [29]. (h) X-Net [29]. (i) Translation^{-}. (j) DTCDN [30]. (k) MTCDN^{S}. (l) MTCDN^{O}. (m) Ground truth.

IV. Comparisons and Analyses

In this section, the extended comparison experiments of MTCDN and DTCDN were conducted. To explore the CD effect of heterogeneous images, different image pairs were used in CD experiments. Moreover, the ablation experiments were carried out to analyze how the advanced block in MTCDN affected the heterogeneous CD.

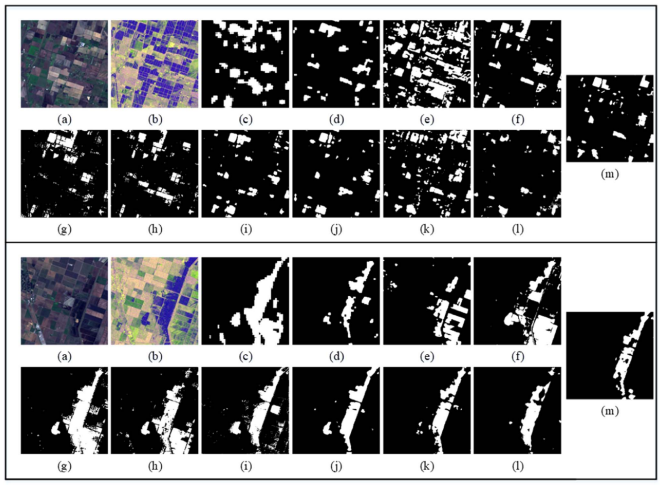


Fig. 12. Visualization of binary maps on the California dataset. (a) Pre-event image. (b) Postevent image. (c) INLPG [20]. (d) HPT [26]. (e) SCCN [27]. (f) cGAN [28]. (g) ACE-Net [29]. (h) X-Net [29]. (i) Translation^{-}. (j) DTCDN [30]. (k) MTCDN^{S}. (l) MTCDN^{O}. (m) Ground truth.

TABLE VI
COMPARISON OF MTCDN AND DTCDN ON EFFICIENCY

Model	Number of parameters	Inference Time(s)
MTCDN	48.06M	0.54
DTCDN	119.14M	1.13

A. Comparison With MTCDN and DTCDN

DTCDN [28] can be divided into two steps, i.e., deep translation and CD. DTCDN splits them into completely independent steps. Conversely, the proposed MTCDN combines them to reduce unnecessary processing. The number of the two models' parameters and the inference time of one image are shown in Table VI.

As for the number of parameters, MTCDN only has about 48.06 M, which is far less than that of DTCDN (119.14 M), showing a decrease of over 50%. The inference time was reduced for fast prediction. As mentioned previously, MTCDN and DTCDN achieved high accuracy for four datasets. However, compared with MTCDN, DTCDN requires a manual data augmentation after image translation to get a robust result. Therefore, MTCDN is more suitable for small sample datasets and keeps a high accuracy.

B. Comparison of Heterogeneous and Homogeneous Images

To further analyze the effect of heterogeneous images on CD, two other corresponding optical and SAR images from Gloucester II were collected. As shown in Fig. 13, the Gloucester II consisted of optical and SAR images from both time steps. Therefore, they could be divided into three image pairs, i.e., homogeneous optical [see Fig. 13(a) and (b)], homogeneous SAR [see Fig. 13(c) and (d)], and heterogeneous [see Fig. 13(a) and (d)] pairs. Homogenous pairs were directly imported into the discriminator for CD because the translation was unnecessary.

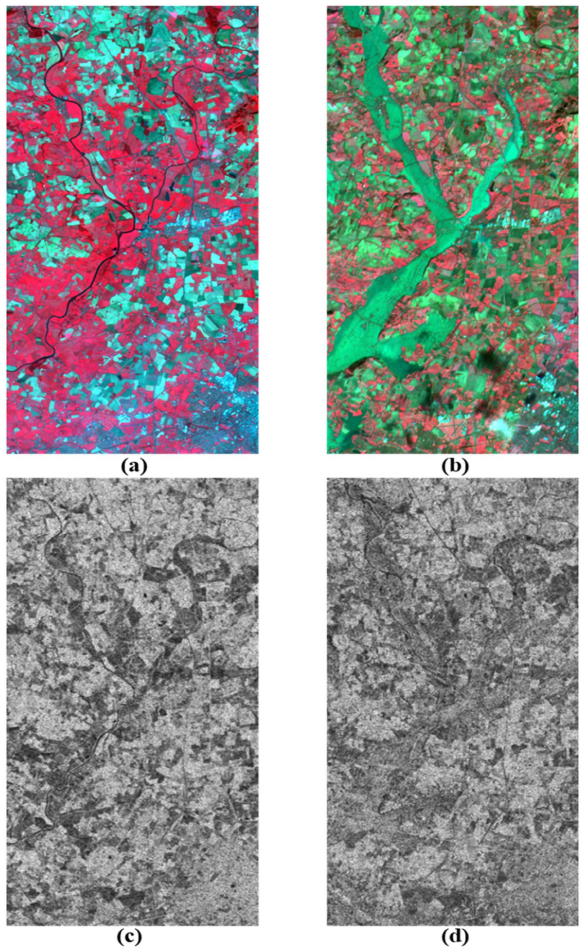


Fig. 13. Optical and SAR images for pre- and postevents. (a) T_1 optical. (b) T_2 optical. (c) T_1 SAR. (d) T_2 SAR images.

TABLE VII
RESULTS OF THREE IMAGE PAIRS FOR GLOUCESTER II DATASET

Data	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
Optical only	88.22	**92.53**	**90.32**	96.59
SAR only	87.67	87.41	87.54	95.88
Optical and SAR	**88.86**	87.59	88.22	**97.65**

The best result of each metric is highlighted in bold.

Different from DTCDN, all image pairs were not processed by data augment.

The experimental CD results of three different pairs are shown in Table VII. Optical-only image pair had an advantage in recall and $F1$-score because clear optical images provided rich information to improve CD. Although it did not get the highest score, the optical and SAR image pairs surpassed two other image pairs in precision and OA. This finding indicated that the results of MTCDN classified to changed pixels were consistent with the ground truth. SAR images contained a lot of speckle noise, which might affect the recall of SAR only and optical and SAR image pairs.

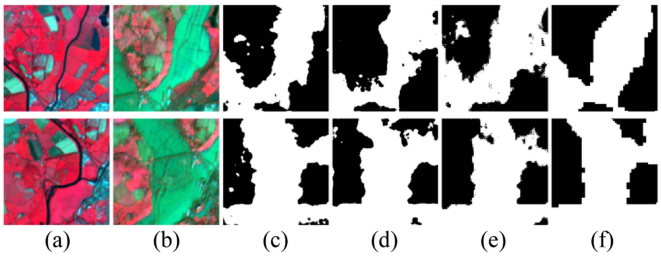


Fig. 14. Evaluations of three image pairs on the Gloucester II dataset. (a) T_1 optical data. (b) T_2 optical data. (c) Optical pairs. (d) SAR pair. (e) MTCDN. (f) Ground truth.

TABLE VIII
ABLATION STUDY OF FEM

Name	Models	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
Shuguang	FEM$^-$	87.92	**90.12**	89.01	97.15
	FEM$^+$	**93.64**	88.75	**92.69**	**99.32**
Gloucester I	FEM$^-$	85.42	**88.62**	86.99	97.34
	FEM$^+$	**88.86**	87.59	**88.22**	**97.65**
Gloucester II	FEM$^-$	88.15	84.88	86.48	95.61
	FEM$^+$	**89.49**	**88.25**	**88.87**	**96.34**
California	FEM$^-$	**66.14**	78.50	71.79	96.78
	FEM$^+$	65.47	**80.61**	**72.26**	**97.83**

The best result of each metric is highlighted in bold.

The binary CD maps of three image pairs are shown in Fig. 14. The ground truth is displayed in Fig. 14(f) in which the white pixels represent the flood area. Optical image results generated more salt noise than the other image pairs. The result of MTCDN contained some holes in the whole flood objects, which should be improved in further research.

C. Ablation Study

The UNet++ relies on upsampling and downsampling to vary the size of receptive field and catch the different levels of information. FEM, as an extra information-enhanced method between high- and low-level features, was proposed. For further analysis, an ablation study was carried out. Table VIII presents the assessment results of FEM effectiveness. All four datasets had an improvement on most evaluations. CD results proved that FEM could enhance the connection of low- and high-level features, thereby promoting the network's ability on feature fusion. Although removed FEM had achieved little improvement in recall on the Shuguang and Gloucester I datasets, this method considered many unchanged pixels as changed pixels, reflecting a decrease in precision. The California dataset with FEM had lower results than without FEM in precision because of the complexity of its SAR image.

Some binary CD maps of four datasets are shown in Fig. 15. In the areas of change, model with FEM module is more superior because it enhances fusion at different levels. The four datasets with FEM obtained complete detection and improved detail.

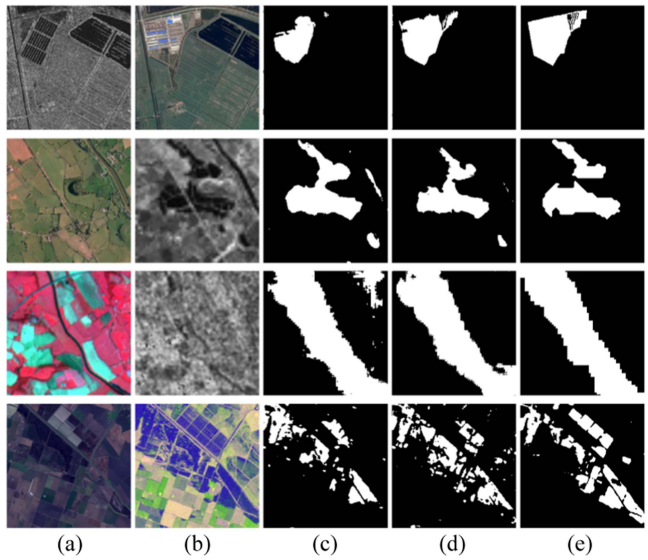


Fig. 15. CD examples of FEM ablation study on four datasets. (a) Pre-event image. (b) Postevent image. (c) Without FEM. (d) MTCDN. (e) Ground truth.

TABLE IX
QUANTITATIVE EVALUATIONS ON DIFFERENT λ_1

λ_1	mode	Precision (%)↑	Recall (%)↑	F1 (%)↑	OA (%)↑
0.5	MTCDN^S	80.69	89.00	84.64	96.74
	MTCDN^O	81.51	73.75	77.44	95.66
1	MTCDN^S	72.62	89.74	80.28	95.55
	MTCDN^O	69.06	73.04	71.00	93.98
2.5	MTCDN^S	85.79	86.92	86.35	97.23
	MTCDN^O	**88.86**	80.84	84.66	97.04
5	MTCDN^S	87.69	85.10	86.38	97.31
	MTCDN^O	**88.86**	87.59	**88.22**	**97.65**
7.5	MTCDN^S	65.69	76.94	70.87	93.62
	MTCDN^O	83.24	75.51	79.19	95.99
10	MTCDN^S	73.84	**92.38**	82.08	95.93
	MTCDN^O	87.01	65.44	74.70	95.53

The best result of each metric is highlighted in bold.

FEM brought some computations to our network. Thus, how to balance this problem should be studied in the future.

D. Parameters Analysis

Parameters λ_1 and λ_2 in (10) and α and γ in (12) were set to 5, 10, 0.25, and 2, respectively, to get an ideal result. Given that the effectiveness of focal loss [43] was studied in the previous research and $\alpha = 0.25$, $\gamma = 2$ was proved to be the best, we followed this conclusion. Besides, all domain transfer methods with a cyclic structure, including CycleGAN [36], NiceGAN [37], and DTCDN [30], set the weight of cycle loss to 10, a slight difference over these methods is the weight of reconstruction loss. Therefore, a comparative study of λ_1 was carried out.

Table IX presents the assessments on different λ_1 with Gloucester I dataset. As can be seen in Table IX, the optimal

and suboptimal metrics are set to bold font and blue color, respectively. Although the performance of MTCDN is fluctuating with the increase of λ_1, it is obvious that an excessively big or small parameter is inappropriate. Besides, the best performance of MTCDN lies near the range of [2.5, 5] and $\lambda_1 = 5$ performs better.

V. Conclusion

In this article, a multitask CD framework of optical and SAR images called MTCDN is proposed. It integrates the image and CD in an improved GAN to benefit from the effective information of the translated image, which is achieved by reusing the discriminator in GAN. Compared with eight SOTA methods, MTCDN shows outstanding ability in extracting change areas on four optical and SAR datasets. It can reduce misclassification and is robust against noise, resulting in improved precision and $F1$-score. Unlike DTCDN, MTCDN avoids complicated operations and a large number of parameters to realize the real end-to-end heterogeneous CD. Moreover, FEM is imposed in UNet++ to strengthen the low-level features with high-level features, thereby improving the accuracy significantly. Although MTCDN has obtained ideal results in the CD of optical and SAR datasets quantitatively and visually, it is still a simple framework with room for improvement. Evidently, it can be generalized to different types of sensors and is not limited to optical and SAR images. Therefore, MTCDN will be further explored so as to apply to CD of multisource remote sensing data.

Acknowledgment

We would like to thank Prof. M. Migonotte for providing the Gloucester and Shuguang dataset, L. T. Luppino for sharing the California dataset, and the professional reviewers for their helpful remarks.

References

[1] S. Ashbindu, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[2] X. Liu and R. G. Lathrop, "Urban change detection based on an artificial neural network," *Int. J. Remote Sens.*, vol. 23, no. 12, pp. 2513–2518, 2002.

[3] X. Huang, Y. Cao, and J. Li, "An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images," *Remote Sens. Environ.*, vol. 244, 2020, Art. no. 111802.

[4] S. Ye, J. Rogan, Z. Zhu, and J. R. Eastman, "A near-real-time approach for monitoring forest disturbance using Landsat time series: Stochastic continuous change detection," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112167.

[5] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5407–5423, Sep. 2017.

[6] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019.

[7] J. C. Trinder and M. Salah, "Aerial images and LIDAR data fusion for disaster change detection," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. I-4, pp. 227–232, 2012.

[8] J. Sublime and E. Kalinicheva, "Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1123.

[9] Z. Lv, P. Zhong, W. Wang, Z. You, J. A. Benediktsson, and C. Shi, "Novel piecewise distance based on adaptive region key-points extraction for LCCD with VHR remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5607709.

[10] X. Li, M. He, H. Li, and H. Shen, "A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8017505.

[11] X. Zhang, H. Su, C. Zhang, X. Gu, X. Tan, and P. M. Atkinson, "Robust unsupervised small area change detection from SAR imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 79–94, 2021.

[12] Z. Lv et al., "Land cover change detection with heterogeneous remote sensing images: Review, progress, and perspective," *Proc. IEEE*, vol. 110, no. 12, pp. 1976–1991, Dec. 2022.

[13] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.

[14] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.

[15] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model," *IEEE Trans. Image Process.*, vol. 29, pp. 757–767, 2020.

[16] W. Zhou, A. Troy, and M. Grove, "Object-based land cover classification and change analysis in the Baltimore metropolitan area using multitemporal high resolution remote sensing data," *Sensors*, vol. 8, no. 3, pp. 1613–1636, 2008.

[17] K. Mubea and G. Menz, "Monitoring land-use change in Nakuru (Kenya) using multi-sensor satellite data," *Adv. Remote Sens.*, vol. 1, pp. 74–84, 2012.

[18] L. Wan, Y. Xiang, and H. You, "A post-classification comparison method for SAR and optical images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.

[19] V. Alberga, "Similarity measures of remotely sensed multi-sensor images for change detection applications," *Remote Sens.*, vol. 1, no. 3, pp. 122–143, 2009.

[20] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700221.

[21] Y. Sun, L. Lei, X. Li, H. Sun, and G. Kuang, "Nonlocal patch similarity based heterogeneous remote sensing change detection," *Pattern Recognit.*, vol. 109, 2021, Art. no. 107598.

[22] L. Lei, Y. Sun, and G. Kuang, "Adaptive local structure consistency-based heterogeneous remote sensing change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8003905.

[23] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Patch similarity graph matrix-based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4841–4861, Jun. 2021.

[24] B. Ayhan and C. Kwan, "A new approach to change detection using heterogeneous images," in *Proc. IEEE 10th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf.*, 2019, pp. 192–197.

[25] C. Kwan, B. Ayhan, J. Larkin, L. Kwan, S. Bernabé, and A. Plaza, "Performance of change detection algorithms using heterogeneous images and extended multi-attribute profiles (EMAPs)," *Remote Sens.*, vol. 11, no. 20, 2019, Art. no. 2377.

[26] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.

[27] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.

[28] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.

[29] L. T. Luppino et al., "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4700422.

[30] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 14–34, 2021.
[31] R. Shao, C. Du, H. Chen, and J. Li, "SUNet: Change detection for heterogeneous remote sensing images from satellite and UAV using a dual-channel fully convolution network," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3750.
[32] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.
[33] Y. Ma, J. Wei, and X. Huang, "Balancing colors of nonoverlapping mosaicking images with generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8024405.
[34] Z. Liu, R. Feng, L. Wang, W. Han, and T. Zeng, "Dual learning-based graph neural network for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5628614.
[35] Z. Liu, R. Feng, L. Wang, and T. Zeng, "Gradient prior dilated convolution network for remote sensing image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3945–3958, Mar. 2023.
[36] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.
[37] R. Chen, W. Huang, B. Huang, F. Sun, and B. Fang, "Reusing discriminators for encoding: Towards unsupervised image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8165–8174.
[38] R. Hang, F. Zhou, Q. Liu, and P. Ghamisi, "Classification of hyperspectral images via multitask generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1424–1436, Feb. 2021.
[39] J. Oh and M. Kim, "PeaceGAN: A GAN-based multi-task learning method for SAR target image generation with a pose estimator and an auxiliary classifier," *Remote Sens.*, vol. 13, no. 19, 2021, Art. no. 3939.
[40] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8007805.
[41] C. Li et al., "Attention Unet++: A nested attention-aware U-Net for liver CT image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 345–349.
[42] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1382.
[43] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[44] M. Mignotte, "A fractal projection and Markovian segmentation-based approach for multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8046–8058, Nov. 2020.
[45] N. Longbotham et al., "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.
[46] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019.

Zhengshun Du received the B.S. degree in geographic information engineering from Sun Yat-Sen University, Guangzhou, China, in 2019, and the M.S. degree in resources and environmental sciences from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2022.

His research interests include deep learning, image processing, remote sensing, and change detection.

Xinghua Li (Senior Member, IEEE) received the B.S. degree in geographical information system and the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2011 and 2016, respectively.

In 2016, he joined the School of Remote Sensing and Information Engineering, Wuhan University, as an Associate Professor. He is the PI of two National Natural Science Foundation of China projects and two subprojects of the National Key R&D Program of China. He has authored or coauthored more than 70 peer-reviewed papers. His current research interests include multitemporal remote sensing image processing and application, and deep learning.

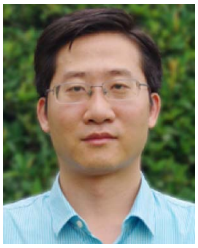
Dr. Li was a recipient of the Best Reviewer of IEEE JSTARS and JAG in 2021. He has served as a reviewer for more than 80 SCI-indexed journals. He is an Associate Editor for *IEEE* Journal *of Selected Topics in Applied Earth Observations and Remote Sensing* and the *Journal of Image Science and Technology*, an Editorial Member of *Remote Sensing* and *Intelligent Automation and Soft Computing*, and the Guest Editor for the *Journal of Applied Remote Sensing*.

Jianhao Miao received the B.S. degree in remote sensing science and technology in 2021 from Wuhan University, Wuhan, China, where he is currently working toward the M.S. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering.

His current research interests include remote sensing image normalization, change detection, and semantic segmentation.

Yanyuan Huang received the B.S. degree in remote sensing science and technology in 2021 from Wuhan University, Wuhan, China, where she is currently working toward the M.S. degree in remote sensing science and technology with the School of Remote Sensing and Information Engineering.

Her current research interests include deep learning, remote sensing image processing, and change detection.



Huanfeng Shen (Senior Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he joined the School of Resource and Environmental Sciences (SRES), Wuhan University, where he is currently a Luojia Distinguished Professor and the Dean of SRES. He was or is the PI of two projects supported by the National Key Research and Development Program of China and seven projects supported by the National Natural Science Foundation of China. He has authored more than 200 research papers published in peer-reviewed international journals. His research interests include remote sensing image processing, multisource data fusion, and intelligent environmental sensing.

Dr. Shen is a Council Member of the China Association of Remote Sensing Application, an Education Committee Member of the Chinese Society for Geodesy Photogrammetry and Cartography, and a Theory Committee Member of the Chinese Society for Geospatial Information Society. He is currently a member of the Editorial Board of the *Journal of Applied Remote Sensing* and *Geography and Geo-Information Science*.

Liangpei Zhang (Fellow, IEEE) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He was a Principal Scientist of the China State Key Basic Research Project, from 2011 to 2016, appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He is a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He has authored or coauthored more than 700 research papers and five books. He is the Highly Cited Author of the Institute for Scientific Information (ISI). He holds 30 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing, and the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics. He also serves as an Associate Editor or Editor for more than ten international journals. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is the Founding Chair of the IEEE Geoscience and Remote Sensing Society Wuhan Chapter.