SemiCDNet: A Semisupervised Convolutional Neural Network for Change Detection in High Resolution Remote-Sensing Images

Daifeng Peng, Lorenzo Bruzzone, *Fellow, IEEE*, Yongjun Zhang, *Member, IEEE*, Haiyan Guan, *Senior Member, IEEE*, Haiyong Ding, and Xu Huang

***Abstract*— Change detection (CD) is one of the main applications of remote sensing. With the increasing popularity of deep learning, most recent developments of CD methods have introduced the use of deep learning techniques to increase the accuracy and automation level over traditional methods. However, when using supervised CD methods, a large amount of labeled data is needed to train deep convolutional networks with millions of parameters. These labeled data are difficult to acquire for CD tasks. To address this limitation, a novel semisupervised convolutional network for CD (SemiCDNet) is proposed based on a generative adversarial network (GAN). First, both the labeled data and unlabeled data are input into the segmentation network to produce initial predictions and entropy maps. Then, to exploit the potential of unlabeled data, two discriminators are adopted to enforce the feature distribution consistency of segmentation maps and entropy maps between the labeled and unlabeled data. During the competitive training, the generator is continuously regularized by utilizing the unlabeled information, thus improving its generalization capability. The effectiveness and reliability of our proposed method are verified on two high-resolution remote sensing data sets. Extensive experimental results demonstrate the superiority of the proposed method against other state-of-the-art approaches.**

***Index Terms*— Change detection (CD), deep learning (DL), feature distribution, generative adversarial network (GAN), remote sensing (RS), semisupervised convolutional network.**

Manuscript received May 6, 2020; revised June 22, 2020 and July 17, 2020; accepted July 21, 2020. Date of publication August 6, 2020; date of current version June 24, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 41801386, Grant 41701540, Grant 41671454, Grant 41571350; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180797; in part by the Startup Project for Introducing Talent of Nanjing University of Information Science and Technology (NUIST) under Grant 2018r029; and in part by the China Scholarship Council under Grant 201908320183. *(Corresponding author: Daifeng Peng.)*

Daifeng Peng is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: daifeng@nuist.edu.cn).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

Yongjun Zhang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zhangyj@whu.edu.cn).

Haiyan Guan and Haiyong Ding are with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: guanhy.nj@nuist.edu.cn; hyongd@163.com).

Xu Huang is with Wuhan Engineering Science and Technology Institute, Wuhan 430019, China (e-mail: huangxurs@whu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2020.3011913

I. Introduction

RECENTLY, due to the implementation of increasingly growing Earth observation programs such as Sentinel, WorldView, GeoEye, ZY-3, and GF, large amounts of remote-sensing (RS) images with different resolutions and modalities are available from various sensors. Therefore, the problem of developing effective techniques to exploit meaningful knowledge from RS big data becomes increasingly popular. Among the possible techniques for extracting semantic information, change detection (CD) methods are crucial for an accurate understanding of land use/cover change over a certain period at a large scale [1]. Generally, for coregistered image pairs or sequence images of the same region acquired at different dates, CD can detect changes of interest related to vegetation, forests, land, buildings, roads, fires, floods, landslides, etc. This means CD plays an important role in the monitoring of urban expansion, deforestation, natural disasters, resources, and ecosystem [2]–[4].

An extensive summary and reviews of CD algorithms have been made in [5]–[7], where it has been shown that CD techniques evolved with the development of machine learning and pattern recognition. In the early stage, only middle- and low-resolution RS images were available, such as MODIS and Landsat, with pixels containing many ground objects. Pixel-based CD (PBCD) was employed by comparing pixel spectral or textual values of images acquired on different dates [8]. To extract robust features, image transformation methods, such as principal component analysis (PCA) [9] and multivariate alteration detection (MAD) [10], were utilized. Furthermore, to overcome the effect of "salt and pepper" noise, spatial-context information was introduced by using neighboring windows [11], and probability graph models (PGMs) [12], [13]. Later on, with the increase of spatial resolution, the ground could be observed at a finer scale. Image segmentation algorithms were often applied to generate image objects, which are more closer to human understanding of image data. Object-based CD (OBCD) hereby were widely employed to compare image difference by utilizing object features [14] or class memberships [15]. In OBCD, spatial-context information

can be included naturally, and many object features such as texture, shape, and geometry can be extracted to improve CD accuracy and reliability. Note that CD can be addressed in an unsupervised or supervised way. The former is intended to detect all the possible changes without any labeled data and is mostly employed for large-scale monitoring with low- and middle-resolution satellite images. The latter aims to detect specific changes through supervised models, which are learned from labeled data. In short, unsupervised CD methods usually detect the location and extension of changes, while supervised CD methods also identify the type of changes. Very high-resolution (VHR) RS images are extraordinarily rich of spatial details. Thus, it is almost impossible to focus on all kinds of changes. Furthermore, in VHR images, thematic changes, such as buildings and roads, which reflect main artificial changes due to human activities or natural disasters [16], [17], are more important than others. Through labeled information, such changes can be defined and learned more easily by using supervised learning models. Accordingly, in our case, the focus is on the supervised methods, as the interest is to extract specific changes (e.g., buildings) in complex VHR images through a training phase. In the literature, most building CD methods are based on generating difference images (DIs), and then extracting building changes through indicators such as morphological building index (MBI) [18] and height information generated by stereo images [19] or point clouds [20].

In the past few years, computing resources have undergone a revolutionary development, especially related to GPUs, which make it possible to process large amounts of image data in a short time. In this context, due to the powerful deep feature representation and nonlinear problems modeling abilities, deep learning (DL) methods have achieved dominant success against traditional machine learning methods in a variety of areas such as image recognition/segmentation, scene understanding, and natural language processing. This opened up a new era in the areas of artificial intelligence (AI). Meanwhile, DL methods have also been widely used to solve RS problems such as image classification [21], object detection [22], [23], image super-resolution and denoising [24], [25], scene interpretation and segmentation [26]. Many studies have also been made to address CD issues based on DL techniques in either unsupervised or supervised manner. Note that only specific changes can be detected based on supervised DL methods, where the change contents and types are determined by the labeled ground truth. In [27], we make a comprehensive review of DL-based CD (DLCD) methods, which can mainly be divided into: 1) feature-based DLCD (FB-DLCD), 2) patch-based DLCD (PB-DLCD), and 3) image-based DLCD (IB-DLCD). Contrary to hand-crafted features, which are designed carefully with expert knowledge and are scene-dependent, deep features are learned hierarchically from deep neural networks (DNNs) with available data sets. That means that deep features can be more robust and discriminative for distinguishing image changes. In FB-DLCD methods, deep features are used to generate a DI, then a threshold segmentation method is applied to obtain a final change map (CM) [28], [29]. Based on the characteristics of data sets, deep features can be extracted from pretrained convolutional neural networks (CNNs) or specifically designed DL models. However, the feature representation and similarity metric errors during the generation of DIs will inevitably be propagated into final CMs. In order to overcome these drawbacks, PB-DLCD are proposed, where pixel patches or super-pixels are constructed first. Then DIs are obtained to serve as pseudo-training sets, and a specific DNN model can be trained to learn the change type of the center pixel [30], [31]. To further overcome the error accumulation effect, pixel patches can also be utilized to train a DNN model from scratch using only the available data sets, where the change type of unknown pixels can be determined directly [32]. Nevertheless, obvious limitations exist for PB-DLCD methods: first, it is difficult to determine a proper size for the patch, which greatly influences the DNN performance; second, pixel patches contain too much redundant information, leading to overfitting effect and increasing computational burden. To overcome these limitations, IB-DLCD methods are investigated, where CD issues can be regarded as a binary semantic segmentation problem [27]. After image clips are constructed, a fully convolutional network (FCN) can be adopted to learn the semantic segmentation result of each pixel directly. Note that two periods of image clips can be stacked [33], [34] or act as independent branches [35] when they are fed into different FCN architectures. However, FCN models rely heavily on large amounts of annotated training samples, which are typically generated manually with expert knowledge and tedious work. For the RS CD task, very few open annotated data sets are available, thus limiting the practical applications of DL models, especially FCNs.

A possible solution to address this limitation is data augmentation, where the number of training data is enlarged with artificially synthesized operations such as translation, rotation, scaling, shifting, flipping, and cropping. It has to be mentioned that random noise sampled from a simple distribution can be mapped into realistic image-label pairs as augmented data by utilizing generative adversarial networks (GANs) [36]. However, none of the augmentation methods utilize unlabeled data, and the enlarged data can only be regarded as an interpolation of the existing labeled data. That makes the model performs poorly on unseen test images. In addition, FCN models require densely annotated images, which are difficult to generate through GANs and few work has been proposed to solve such a problem. Another solution to overcome the drawbacks is to use weakly supervised learning, where easier-to-obtain annotations, such as image-level labels [37], bounding boxes [38], or scribbles [39], are exploited to train the DNN models. The main idea is to generate pseudo labels so that supervised learning can be adopted, so are the saliency-guided methods [40], where pseudo labels are generated by saliency detection. Nevertheless, human interaction is still needed, which may be difficult to obtain in some cases, especially for RS images with ground objects that vary sharply in scale and size.

Instead of relying on weak annotations, semisupervised learning (SSL) is capable of exploiting the discriminative features from available unlabeled images, thereby improving the model generalization performance and making it possible to train the model using small labeled data sets. Due to

the extreme difficulty of obtaining large labeled data sets in RS problems, SSL methods have attracted increasing interest toward tasks like classification [41], dimensionality reduction [42], and CD [43]. For example, to overcome the shortage of training data in hyperspectral image classification, Hong *et al.* [44] proposed a semisupervised cross-modality learning framework, where a large amount of poorly discriminative multispectral data was utilized to improve classification performance.

In general, unlabeled pixels account for a large proportion in CD data sets. It is natural to improve CD results by exploiting the potential of unlabeled pixels. However, most of the existing semisupervised CD methods are focused on individual pair of hyperspectral images, SAR images, or multispectral images with medium resolution [45]–[48], where semisupervised classifiers or metric learning are applied to exploit the unlabeled information. With the advent of RS big data era, it is more convenient to obtain large amounts of image pairs for CD. Nevertheless, few work has been done to deal with semisupervised CD in such a context, which is of great importance to improve CD accuracy and efficiency, and promote CD practical applications.

To address the abovementioned issues, we propose a novel semisupervised CD network (SemiCDNet) for VHR RS images. First, a UNet++ model [49] is adopted to generate initial change results. To combine multilevel features effectively, attention mechanism is introduced to refine the UNet++ model, thus generating finer initial CMs for VHR images. Second, to exploit the potential of unlabeled data, two discriminators are used by enforcing the feature distribution consistency of CMs and entropy maps between the labeled data and the unlabeled data, thereby improving final CD results. The contributions of this article are twofold.

1) A novel end-to-end SemiCDNet is proposed for the semisupervised CD task. To improve segmentation performance, a light-weighted attention module is embedded into the UNet++ network. To fully exploit the potential of unlabeled data, a GAN is employed by combined usage of segmentation adversarial loss and entropy adversarial loss. Comprehensive comparisons and ablation studies are carried out to verify its effectiveness.
2) A challenging data set is constructed by employing VHR satellite images, which will be released publicly for the benefit of promoting RS CD research using DL techniques. The data set can be found at https://github.com/daifeng2016/Change-Detection-Dataset-for-High-Resolution-Satellite-Imagery.

The rest of the article is organized as follows. Section II describes the background and related work. The proposed SemiCDNet is illustrated in detail in Section III. Experimental results on the effectiveness of the proposed method are presented and discussed in Section IV. Finally, Section V draws the conclusions of this article.

II. Background and Related Work

In this section, the concepts of semantic segmentation and SSL, as well as the methods of semisupervised CD will be briefly illustrated.

A. Semantic Segmentation

Given a set of class labels, the task of semantic segmentation is to predict the label value for each pixel in the image, namely obtaining the dense annotations for the whole image. Binary CD can be seen as a binary semantic segmentation problem, where only two labels with regard to changed or unchanged class have to be detected. Therefore, the techniques and strategies of semantic segmentation can be naturally introduced to deal with CD task. Recent advancements in convolutional networks have shown great potential to solve such problems. Long *et al.* [50] first proposed an end-to-end image segmentation method based on FCN, which outperformed existing approaches on both accuracy and efficiency. However, the considered feature map resolution becomes poor after a series of pooling operations, leading to poor spatial accuracy. In order to overcome this limitation, atrous convolution [51] and encoder–decoder architecture [52] were proposed to produce fine-grained segmentation results. By adding skip connections between encoder and decoder layers, UNet [53] achieves even better segmentation results and has been widely used in image segmentation and object detection. To overcome the semantic gaps between encoder and decoder parts, many extensions of UNet have been proposed, such as MultiResUnet [54], MDU-UNet [55], and UNet++ [49], which have opened up new perspectives of improving semantic segmentation performance for RS community.

B. Semisupervised Learning

SSL was proposed for exploiting the potential of unlabeled data, thereby overcoming the limitations of few labeled data. In the setting of SSL, the data set $\mathcal{D}$ is split into the labeled part $\mathcal{D}_{\mathcal{L}} = \{(x_i^l, y_i^l)\}_{i=1}^{M}$, which contains M labeled images x_i^l and the corresponding ground-truth masks y_i^l, and unlabeled part $\mathcal{D}_{\mathcal{U}} = \{(x_i^u)\}_{i=1}^{N}$, which contains only N unlabeled images x_i^u, typically $N \gg M$. For a given neural network F, its parameters θ_F are learned by solving an optimization problem as follows:

$$\min_{\theta_F}\left\{\frac{1}{M}\sum_{i=1}^{M}\mathcal{L}_{\text{seg}}\left(x_i^l, y_i^l\right) + \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{\text{semi}}\left(x_i^u\right)\right\} \tag{1}$$

where the first term is the supervised loss calculated using the labeled data, while the second term is the semisupervised loss calculated based on the unlabeled data. Note that the semisupervised loss can be regarded as a regularization term, thereby introducing unsupervised regularization effect and improving model generalization ability.

With regard to semantic segmentation, SSL has achieved great success and paved the way for model training using few labeled data. Hung *et al.* [56] proposed a semisupervised semantic method using GAN. By coupling the adversarial loss with cross entropy loss, the semantic segmentation accuracy can be improved. Unlabeled images were leveraged through discovering the high-confidence regions of the predicted results, thereby enhancing the segmentation model. In [57], a semisupervised semantic segmentation method is proposed by using two network branches, which encourage high- and low-level consistency when training with

few labeled images. In order to improve the final performance, self-training procedure was employed through utilizing high-quality generator predictions of the unlabeled images. Mondal *et al.* [58] exploited the potential of CycleGAN for semisupervised segmentation. In addition to using adversarial learning, cycle consistency was also employed to learn a bidirectional mapping between unpaired images and segmentation tasks, which served as an unsupervised regularization and boost the segmentation performance. Chen *et al.* [59] proposed a novel semisupervised image segmentation method by using unsupervised reconstruction objectives. An attention mechanism was also applied to separate the reconstruction results into different classes, thus learning more discriminative features and improving segmentation performance.

C. Semisupervised CD

In the RS community, supervised CD has always been faced with challenges of limited labeled data set, which is even worse when it comes to RS big data. SSL is capable of introducing unsupervised regularization to improve supervised learning, thus making it possible to train a model using limited number of labeled data. Due to its powerful capacities to exploit unlabeled information, SSL has attracted increasingly interest to solve CD problems. Bovolo *et al.* [45] proposed a semisupervised CD for multispectral RS images by utilizing a defined binary semisupervised support vector machine (S^3VM) classifier, where the unlabeled patterns were gradually considered to define the decision boundary between the changed and unchanged pixels. Chen *et al.* [60] proposed a semisupervised CD method via a Gaussian process (GP) classifier and a Markov random field (MRF) model. DIs were generated first, then both the labeled and unlabeled data were exploited by a probabilistic GP classifier. To overcome the shortcomings of GP classifier and include the spatial contextual information, MRF regularization was employed by introducing edge information and high-order potentials. A modified self-organizing feature map (MSOFM) was proposed by Ghosh *et al.* [61], where only few labeled patterns were utilized to initialize the MSOFM network, and fuzzy set theory was then employed to determine the membership values of the unlabeled data. The main limitation is that DIs have to be generated to serve as input patterns.

Through metric learning, samples from the same class are mapped closely to each other, while as farther apart as possible for the samples from different classes. Thus, a more discriminative metric for measuring samples similarities can be learned, which can be effectively used for semisupervised CD. Yuan *et al.* [47] proposed a semisupervised CD method for hyperspectral images, where a semisupervised Laplacian regularized metric learning was employed to exploit the large amount of unlabeled data. Based on keep it simple and straightforward (KISS) metric learning, Zhang *et al.* [48] proposed a coarse-to-fine semisupervised CD for multispectral images. The contribution of the easy training samples was improved whereas that of the hard training samples was weakened. Then a coarse-to-fine strategy was applied on the testing samples by combining metric learning and neighborhood label information.

It is worth noting that GAN [36], which is capable of learning the feature distribution of training samples, has achieved great success on both supervised learning and semisupervised learning. With regard to semisupervised CD, GAN was also widely utilized. Gong *et al.* [62] proposed a generative discriminatory classified network (GDCN) for multispectral image CD. The fake data generated by random noise served as additional training samples, while the unlabeled data were used to estimate the appropriate prior information, thereby boosting the performance of discriminator. However, patch size is difficult to define. In addition, for VHR images with complex scenes, it is quite difficult to generate fake data from only random noise. In [63], a graph model with GAN was first proposed for semisupervised CD. Multitemporal images were first converted into a partially labeled graph with changed nodes, unchanged nodes, and unlabeled nodes. Then, semisupervised graph learning based on GAN was applied to generate certain labels for the unlabeled nodes, where both the labeled and unlabeled nodes information was exploited.

To sum up, in order to exploit unlabeled information, existing semisupervised CD methods are mostly implemented by using semisupervised classifiers, metric learning, or GAN. However, only individual image pairs are investigated, while few works have been focused on a population of images, which is more beneficial for large-scale real-world applications.

III. Proposed SemiCDNet

In this section, the architecture of the proposed SemiCDNet will be illustrated first. Then, we will present details on the improved UNet++ segmentation network and discriminator network. Finally, the loss functions of the segmentation network and discriminator network will be defined.

A. SemiCDNet Architecture

The architecture of the proposed SemiCDNet is illustrated in Fig. 1, which consists of one generator and two discriminators. Let us assume that the data set is divided into two parts: 1) labeled images x^l with their ground truth y^l and 2) unlabeled images x^u. First, both the labeled images x^l and unlabeled images x^u are stacked to feed into the segmentation network $\boldsymbol{G}$, where corresponding initial predictions of $\hat{y}^l$ and $\hat{y}^u$ can be generated. Based on the ground truth y^l and segmentation predictions $\hat{y}^l$, the segmentation network $\boldsymbol{G}$ can be optimized in a supervised manner by using a binary cross-entropy loss L_{bce}. However, for the unlabeled images, due to the lack of ground truth, the network $\boldsymbol{G}$ cannot be optimized by L_{bce}. To address this issue, segmentation adversarial learning is employed by using a discriminator $\boldsymbol{D_s}$, which is designed to infer whether the segmentation output is either from unlabeled images or from the ground truth. During the competing training of segmentation network $\boldsymbol{G}$ and discriminator $\boldsymbol{D_s}$, the feature distributions of the unlabeled predictions $\hat{y}^u$ will become similar to those of the ground truth y^l, thus improving the segmentation results for the unlabeled images. In addition, the predictions on the unlabeled images tend to be of high uncertainty, especially on the boundary areas. To suppress the uncertain predictions, entropy adversarial learning (EAL) is further adopted through a discriminator

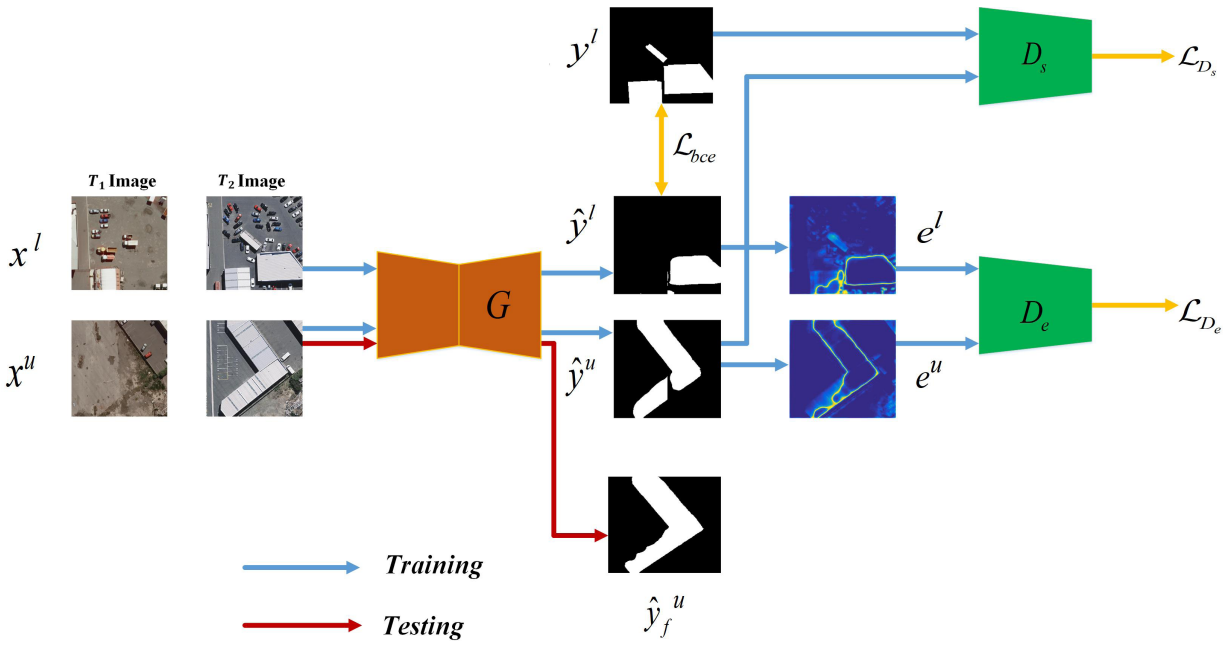


Fig. 1. Flowchart of the proposed method.

$\boldsymbol{D_e}$, which aims to encourage the entropy map e^u to be similar to the entropy map e^l by aligning their feature distributions. Note that the discriminators $\boldsymbol{D_s}$ and $\boldsymbol{D_e}$ are optimized using the segmentation adversarial loss $\mathcal{L}_{D_s}$ and entropy adversarial loss $\mathcal{L}_{D_e}$, respectively. The segmentation network $\boldsymbol{G}$ and the two discrimination networks $\boldsymbol{D_s}$ and $\boldsymbol{D_e}$ are trained alternatively using the corresponding loss until the defined number of iterations is reached. Finally, during the testing stage, the unlabeled images are directly fed into the trained segmentation network $\boldsymbol{G}$ to generate the final predictions $\hat{y}_f^u$.

B. Segmentation Network

Contrary to UNet with simple short skip connections, Unet++ has nested dense skip connections, which greatly facilitates the multiscale feature extraction and enhancement, thereby alleviating the degeneration of spatial information. However, only direct concatenation is utilized to combine the high-level and low-level features in the decoding stage. We argue that this strategy brings two main limitations: 1) semantic gap exists for high-level and low-level features, which leads to some discrepancy and confusion for the network after fusing them directly and 2) redundant information will be generated as not all the combined features are useful for the network. Hence, an attention mechanism is adopted to reweight the features and integrate them effectively. The illustration of our proposed UNet++ network with attention mechanism is shown in Fig. 2(a). In the encoding stage, five convolution units are utilized to generate down-sampled feature maps, thereby extracting multiscale features. While in the decoding stage, dense skip connections are adopted, and an attention unit is embedded after every concatenation operation so as to combine different features effectively. Finally, a sigmoidal layer is adopted to generate the final segmentation map. In order to avoid the gradient vanishing problem during the network training, residual convolution strategy is employed in the convolution unit, as shown in Fig. 2(b). Fig. 2(c) presents the architecture of our proposed attention unit in detail. Assuming the input features $F_{\text{in}} = [F_1, F_2, \ldots, F_c] \in \mathbb{R}^{C\times H\times W}$, where C denotes the channel number of the feature maps, and H and W refer to the height and width, respectively. In order to capture different clues to produce finer channel-wise attention, both average-pooling and max-pooling are adopted to aggregate spatial information along the row and column dimensions, where $F_{\text{avg}} \in \mathbb{R}^{C\times 1\times 1}$ and $F_{\max} \in \mathbb{R}^{C\times 1\times 1}$ are generated. In addition, to learn the channel correlations and appropriate weight distribution, a 1×1 convolution layer is followed. Next, F_{avg} and $F_{\max}$ are merged through element-wise summation to produce channel weights, which are constrained between 0 and 1 by a sigmoid layer. Finally, the input features and channel weights are fused to generate final output features $F_{\text{out}} \in \mathbb{R}^{C\times H\times W}$, which can be computed as

$$F_{\text{out}} = \sigma[\text{Conv}(\text{AvgPool}(F_{\text{in}})) + \text{Conv}(\text{MaxPool}(F_{\text{in}}))] \otimes F_{\text{in}} \tag{2}$$

where σ denotes the sigmoid function and $\otimes$ denotes the element-wise multiplication.

C. Discriminator Network

Instead of using discriminator in a fully convolutional manner [48], we employ a discriminator in a fully connected way, which is proved to be more effective for GAN training. The illustration of the proposed discriminator network is shown in Fig. 3. It consists of an encoder module, a global average pooling (AvgPooling) layer, a fully connected (FC) layer, and a sigmoidal layer. The encoder module, which is composed of three convolutional units in sequence, contributes to feature

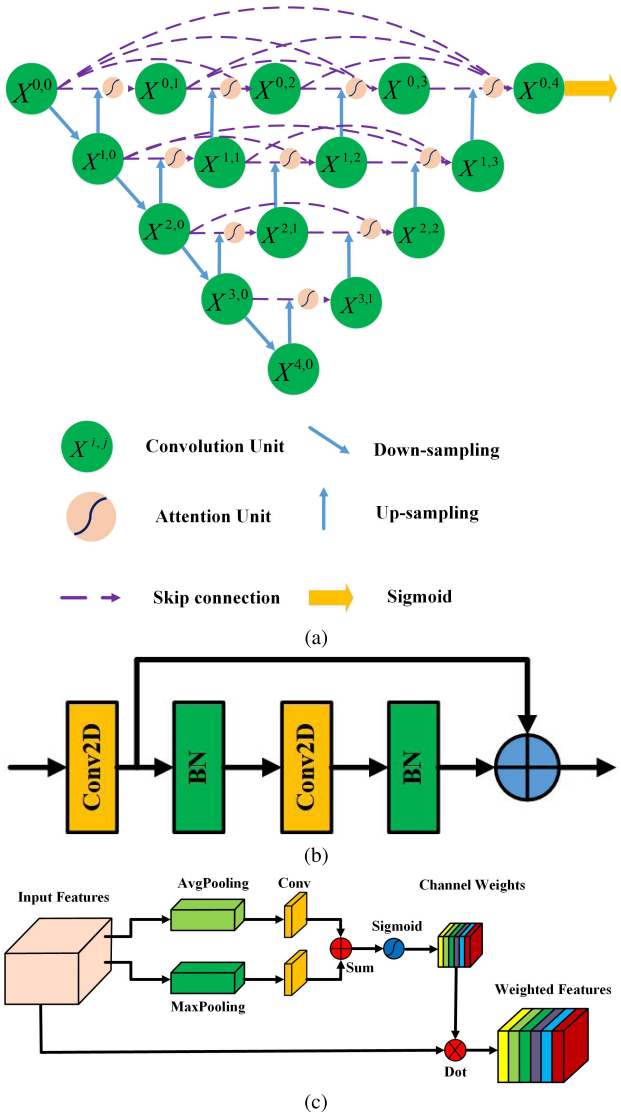


Fig. 2. Illustration of the proposed segmentation network. (a) UNet++ network with attention mechanism. (b) Convolution unit. (c) Attention unit.

extraction and abstraction. Each convolution unit consists of a convolution layer, a leaky-ReLU layer, and a dropout layer. Note that in the convolution layer, kernel size is set to 4, stride size is set to 2, and padding size is set to 1. Hence, max-pooling layer is not needed for feature map contraction. In addition, the dropout layer is adopted, which is crucial for stabilizing GAN training during our test.

D. Loss Functions

In our test, we assume that the data set $\mathcal{D}$ consists of: 1) labeled data $\mathcal{D}_{\mathcal{L}} = \{(x_i^l, y_i^l)\}_{i=1}^{M}$ with M labeled images x_i^l and their ground-truth masks y_i^l and 2) unlabeled data $\mathcal{D}_{\mathcal{U}} = \{(x_i^u)\}_{i=1}^{N}$, which contains only N unlabeled images x_i^u. For our proposed SemiCDNet, the segmentation network $\boldsymbol{G}$ is trained with three types of losses: weighted binary cross-entropy loss, segmentation adversarial loss, and entropy adversarial loss. Additionally, the discriminator network $\boldsymbol{D_s}$ is optimized by segmentation adversarial loss, while the discriminator network $\boldsymbol{D_e}$ is optimized through entropy adversarial loss.

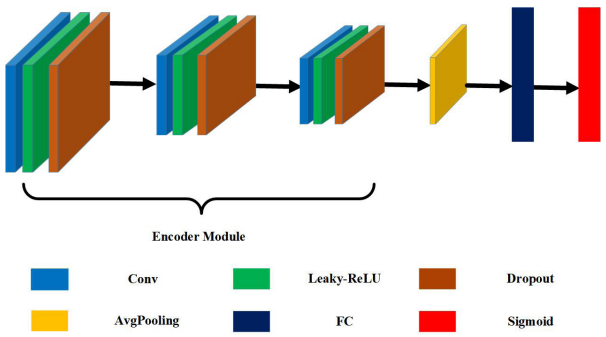


Fig. 3. Illustration of the proposed discriminator network.

1) Weighted Binary Cross-Entropy Loss: To overcome the issues of sample bias between changed pixels and unchanged pixels in CD, a simple weighted binary cross-entropy loss is adopted. This is a standard supervised pixel-wise segmentation loss imposed only on the labeled data to generate their segmentation predictions, which can be expressed as

$$\mathcal{L}_{\text{bce}} = -\frac{1}{M}\left[\beta \sum_{j \in y_+^l} \log(P(y_j = 1)) + (1-\beta) \sum_{j \in y_-^l} \log(P(y_j = 0))\right] \quad (3)$$

where $\beta = |y_-^l|/(|y_+^l| + |y_-^l|)$ and $1-\beta = |y_+^l|/(|y_+^l| + |y_-^l|)$, $|y_+^l|$ and $|y_-^l|$ represent the number of changed and unchanged pixels in the ground truth of the labeled images, respectively. $P(.)$ is the sigmoid output at pixel j.

2) Segmentation Adversarial Loss: Based on the assumption that the segmentation outputs of the labeled and unlabeled data share a similar semantic structure, a segmentation discriminator $\boldsymbol{D_s}$ is employed to figure out whether the segmentation map is from the unlabeled samples or from the ground truth, thereby aligning the feature distributions of the predicted $G(x^u)$ and the ground-truth maps of y^l. Inspired by conditional GAN (cGAN) [64], a stack of input images and label images are fed into $\boldsymbol{D_s}$, which facilitates the stability during the GAN training. Hence, the segmentation adversarial loss can be computed as

$$\mathcal{L}_{D_s} = \frac{1}{M} \sum_{x^l, y^l \in \mathcal{D}_{\mathcal{L}}} \mathcal{L}_D(y^l \oplus x^l, 1) + \frac{1}{N} \sum_{x^u \in \mathcal{D}_{\mathcal{U}}} \mathcal{L}_D(G(x^u) \oplus x^u, 0) \quad (4)$$

where $\oplus$ denotes the concatenation operation and $\mathcal{L}_{\mathcal{D}}$ is the binary cross-entropy loss, which aims to minimize the mean discrepancy between the distribution of the predicted $G(x^u)$ and ground truth y^l. Note that a general form of binary cross-entropy loss can be defined as

$$\mathcal{L}_{\mathcal{D}}(y^p, y^t) = -y^t \log(y^p) - (1-y^t)\log(1-y^p) \quad (5)$$

where y^p represents the prediction value and y^t denotes the target value.

Additionally, to fool the discriminator $\boldsymbol{D_s}$, the segmentation network $\boldsymbol{G}$ is optimized using the following adversarial loss:

$$\mathcal{L}_{\text{adv}}^{D_s} = \frac{1}{N} \sum_{x^u \in \mathcal{D}_\mathcal{U}} \mathcal{L}_D(G(x^u) \oplus x^u, 1). \tag{6}$$

3) Entropy Adversarial Loss: In general, the generator $\boldsymbol{G}$ is prone to produce low-entropy predictions with high certainty on the labeled data, whereas the predictions on the unlabeled data have high-entropy with low certainty. Accordingly, it is naturally to improve segmentation results for the unlabeled data by enforcing low-entropy constraints. In this work, Shannon entropy was applied to calculate the entropy maps, which is defined as

$$E(x) = G(x) \bullet \log[G(x)] \tag{7}$$

where $\bullet$ denotes the dot product operation. Based on the entropy maps, a discriminator $\boldsymbol{D_e}$ is introduced to infer whether the entropy maps is from the labeled data or from the unlabeled data, thereby aligning the feature distribution between $E(x^u)$ and $E(x^l)$, as well as suppressing the high uncertain regions in the unlabeled entropy maps. Similarly, during the training of the discriminator $\boldsymbol{D_e}$, the entropy maps are stacked with their corresponding images to improve GAN performance. Hence, entropy adversarial loss can be defined as

$$\mathcal{L}_{D_e} = \frac{1}{M} \sum_{x^l, y^l \in \mathcal{D}_\mathcal{L}} \mathcal{L}_D(E(x^l) \oplus x^l, 1) + \frac{1}{N} \sum_{x^u \in \mathcal{D}_\mathcal{U}} \mathcal{L}_D(E(x^u) \oplus x^u, 0). \tag{8}$$

Meanwhile, the segmentation network $\boldsymbol{G}$ is optimized to fool the discriminator $\boldsymbol{D_e}$ by utilizing the following adversarial loss:

$$\mathcal{L}_{\text{adv}}^{D_e} = \frac{1}{N} \sum_{x^u \in \mathcal{D}_\mathcal{U}} \mathcal{L}_D(E(x^u) \oplus x^u, 1). \tag{9}$$

Finally, the total loss function for the segmentation network $\boldsymbol{G}$ can be read as

$$\mathcal{L}_G = \mathcal{L}_{\text{bce}} + \lambda_s \mathcal{L}_{\text{adv}}^{D_s} + \lambda_e \mathcal{L}_{\text{adv}}^{D_e} \tag{10}$$

where λ_s and λ_e denote the weights of segmentation adversarial loss and entropy adversarial loss, respectively.

IV. Experimental Results and Discussion

In this section, the results of extensive experiments are presented to verify the effectiveness of the proposed approach. First, a detailed illustration of the data sets is presented. Next, we give a brief description of the training details, methods used for comparisons and the evaluation metrics. Finally, experimental settings and results are analyzed and discussed in detail.

A. Descriptions of Data Sets

To verify the effectiveness of the proposed method, two VHR RS image data sets are employed: WHU building data set [23] and Google data set, both of which include a large amount of VHR image pairs for CD task.

1) WHU Building Data Set: The data set consists of two coregistered aerial images and the change masks captured on the same area of Christchurch, New Zealand, in 2012 and 2016, respectively. The size of each image is 32 507×15 345 pixels with a resolution of 0.075 m, where the main changes are the buildings, as illustrated in Fig. 4. Due to the limitation of the GPU memory, the image pairs are cropped into 256×256 nonoverlapping image blocks with at least a fraction of changed pixels, where 1922 pairs of image clips were generated. Then, the training data set and testing data set were generated by random sampling. To avoid over-fitting, we enlarge the training data by randomly shifting and scaling, rotating by 90°, 180°, and 270°, and flipping in horizontal and vertical directions.

2) Google Data Set: As there have been few publicly available data set for CD, we construct a large-scale VHR multispectral satellite image data set specially for CD research. The images were acquired during the periods between 2006 and 2019, covering the suburb areas of Guangzhou City, China, where the urbanization process was rapid in the past decade. To facilitate image pair generation, Google Earth service through the BIGEMAP software (http://www.bigemap.com) was adopted to collect 19 season-varying VHR images pairs with three bands of red, green, and blue, a spatial resolution of 0.55 m, and the size ranging from 1006×1168 pixels to 4936×5224 pixels. The image changes include waters, roads, farmland, bare land, forests, buildings, ships, etc. However, due to the high complexity of the VHR image scenes, it is almost impossible to annotate all kinds of changes above. Buildings, which denote the main driving forces caused by urbanization, make up the main changes. Therefore, our annotation is focused on buildings, which represent the most significant changes. It is noteworthy that ArcGIS and eCognition software are employed for manual labeling process. The former is used for detecting building changes by careful visual interpretation, while the latter is aimed to label the changes by using object-based analysis. After initial annotation, we carry out careful inspection to make sure the building changes are complete and accurate, where some typical samples are presented in Fig. 5. For the benefit of GPU training, the image pairs are cropped into 256×256 nonoverlapping image blocks with at least a fraction of changed pixels, where 1067 pairs of image clips were generated. After the training data set and testing data set were divided by random sampling, the training data were also augmented to avoid over-fitting by adopting the same strategy as WHU building data set. During the augmentation stage the shifting ratio was set to a larger value to compensate residual coregistration errors. Note that compared with WHU building data set, this data set is more challenging due to: 1) large building shape and size changes, where the buildings are more complex and diverse, ranging from large industry and residence houses to small portable dwellings and 2) large displacement caused by perspective projection of high-rise buildings.

B. Training Details

The proposed method is implemented by Pytorch framework, which is installed on a workstation with Intel Xeon CPU

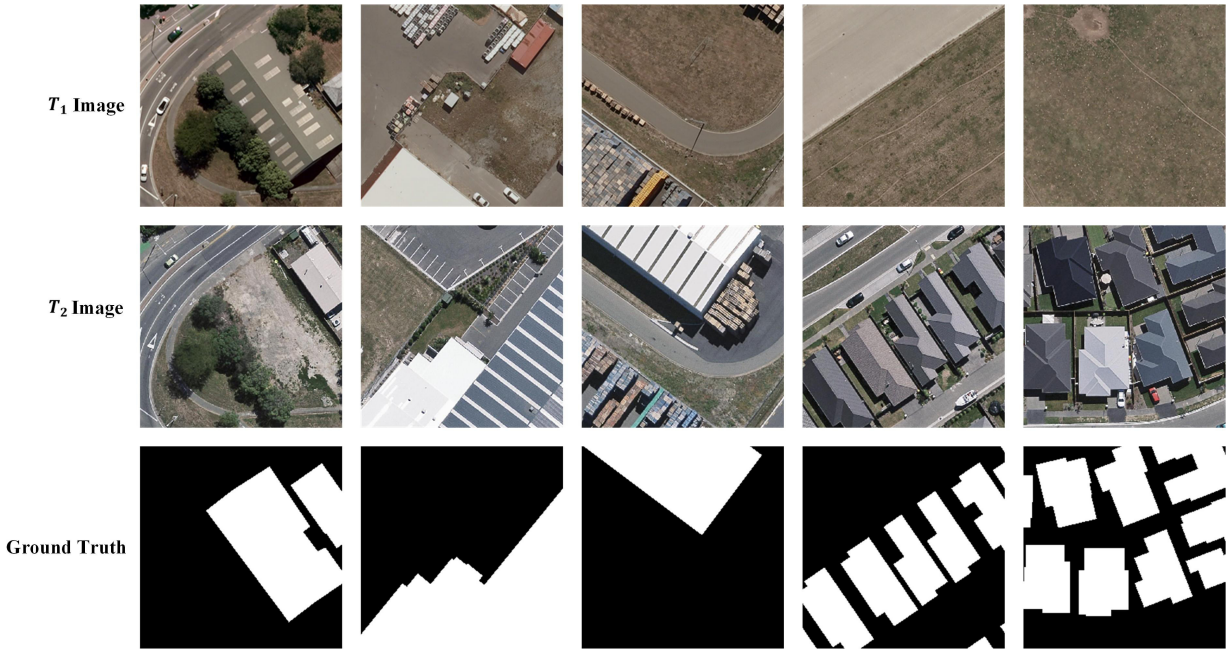


Fig. 4. Example images of WHU building data set.

Fig. 5. Example images of Google data set.

W-2123 (3.6 GHz, 8 cores, and 32GB RAM) and a single NVIDIA GTX 1080 Ti GPU with 11GB RAM. During the training, Adam optimizer is adopted for both the generator and the two discriminators, with the base learning rate set to 2.5e−4 and 1e−4, respectively. To better train the model, a poly-learning policy is utilized, where the initial learning rate was decayed by a factor of $(1 - (\text{iter}/\max\_\text{iter}))^{\text{power}}$ for every iteration, and the power is set to 0.9 for all of our tests. The proposed SemiCDNet was trained for $20K$ iterations for the WHU building data set and $30K$ iterations for the Google data set. The hyperparameters of λ_s and λ_e were set based on cross-validation. Furthermore, the batch size is set to four for all the data sets. In the SSL setting, the SemiCDNet was trained from scratch using randomly sampled labeled data from the nonoverlapping data sets. The sampling ratios for the WHU building data set were set to {5%, 10%, 20%, 50%}, while they were set to {10%, 20%, 40%, 60%} for the Google data set which is more challenging. In addition, for the segmentation network $\boldsymbol{G}$, the size of convolution kernel is set to 3×3, and the number of kernels in each convolution

unit are set to {16, 32, 64, 128, 256} for both WHU building data set and Google data set.

During training, the segmentation network $\boldsymbol{G}$ and the two discriminator networks ($\boldsymbol{D_s}$ and $\boldsymbol{D_e}$) were updated alternatively, with the $\boldsymbol{D_s}$ and $\boldsymbol{D_e}$ fixed first for updating $\boldsymbol{G}$ and then updating $\boldsymbol{D_s}$ and $\boldsymbol{D_e}$ by fixing $\boldsymbol{G}$. Note that both the labeled data and the unlabeled data were utilized for optimizing the network parameters during the training stage, while only the unlabeled data were used for predicting the change map.

C. Comparative Methods and Evaluation Metrics

To verify the effectiveness of our proposed approach, some state-of-the-art (SOTA) CD and SSL methods are compared and analyzed, which are as follows.

1) A fully convolutional network with pyramid pooling (FCN-PP) [33], which has been applied to CD of landslide. It consists of a U-shape architecture and a pyramid pooling layer to capture wider receptive field.
2) A fully convolutional-early fusion with residual blocks (FC-EF-Res) [65], which has been employed for semantic CD in high-resolution satellite image. Residual blocks with skip connections are used to improve the spatial accuracy of change map.
3) Semisupervised training with adversarial network (AdvNet) [56], where a fully convolutional discriminator was proposed to encourage feature distribution consistency between labeled predictions and ground-truth, and a self-training strategy was adopted by exploiting the high-confidence regions of the unlabeled predictions.
4) Semisupervised semantic segmentation GAN (s4GAN) [57]. To facilitate GAN training, feature matching loss was utilized to minimize the discrepancy between the predicted segmentation maps and the ground-truth masks. A self-training loss was further applied to balance the generator and discriminator.
5) CycleGAN for semisupervised segmentation (CycleGAN) [58], where CycleGAN was first introduced for SSL by enforcing cycle consistency, so that a bidirectional mapping between unpaired images and segmentation maps can be learned.

In addition, UNet++ with attention mechanism (UNet++_att) is compared as the fully supervised baseline, which is trained only on the labeled data.

In order to evaluate the effectiveness of the proposed method quantitatively, F1-score (F1), overall accuracy (OA) and Kappa coefficient (Kappa) are utilized by comparing the ground-truth and prediction maps, which can be defined as

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{11}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{12}$$

$$F1 = \frac{2 \times P \times R}{P+R} \tag{13}$$

$$\mathrm{OA} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{14}$$

$$\mathrm{Kappa} = \frac{\mathrm{OA}-\mathrm{PRE}}{1-\mathrm{PRE}} \tag{15}$$

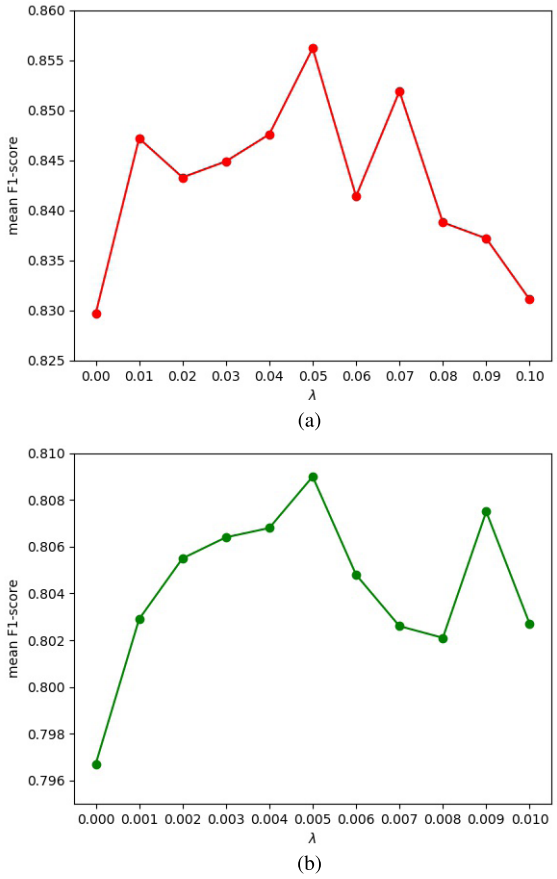


Fig. 6. Effect of parameter λ on the accuracy of the proposed method. (a) WHU building data set. (b) Google data set.

$$\mathrm{PRE} = \frac{(\mathrm{TP}+\mathrm{FN}) \times (\mathrm{TP}+\mathrm{FP})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})^2} + \frac{(\mathrm{TN}+\mathrm{FP}) \times (\mathrm{TN}+\mathrm{FN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})^2} \tag{16}$$

where TP denotes the number of true positives, FP denotes the number of false positives, TN denotes the number of true negatives, and FN denotes the number of false negatives. Note that higher F1-score, OA and Kappa point out better overall performance.

D. Results

1) Parameters Setting: In the loss function of (10), λ_s and λ_e, which balance the weight of the binary cross-entropy loss and adversarial loss, respectively, play an important role on the final loss. In our test, segmentation adversarial loss and entropy adversarial loss are treated equally, i.e., $\lambda_s = \lambda_e = \lambda$. To verify the sensitivity of the parameter, we vary λ from 0.01 to 0.1 for the WHU building data set, and computed the corresponding mean F1-score based on different SSL settings, as illustrated in Fig. 6(a). With regard to the Google data set, as it is more challenging for the generator to produce initial predictions, λ is varied from 0.001 to 0.01, obtaining the results presented in Fig. 6(b). When λ is set to 0, only binary cross-entropy loss is adopted, and the F1-score is low. Then, by increasing λ, the value of F1-score grows gradually, demonstrating the

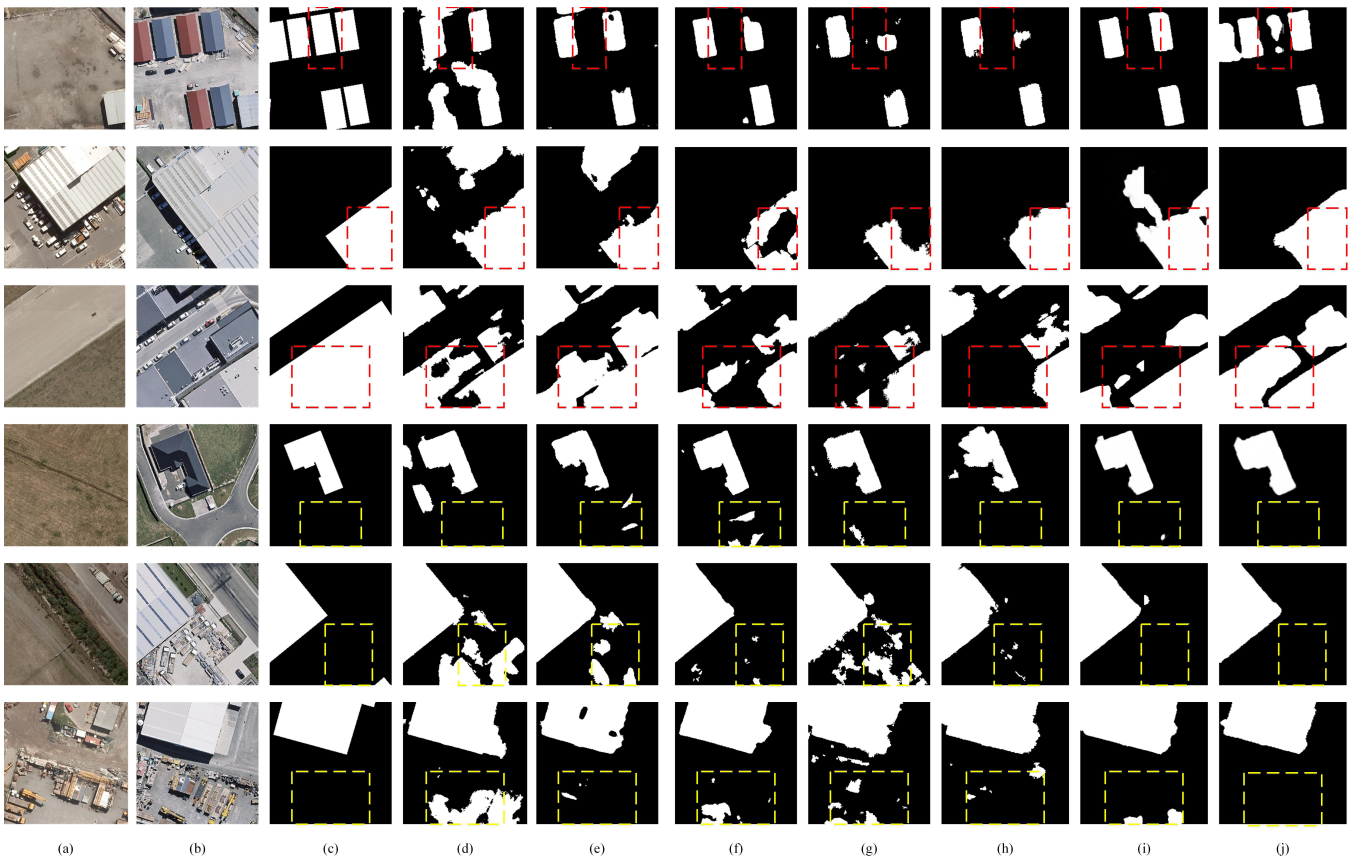


Fig. 7. Visual comparisons of CD maps by different approaches on WHU building data set using 10% labeled images for training. (a) Image T_1. (b) Image T_2. (c) Reference change map. (d) FC-EF-Res. (e) FCN-PP. (f) UNet++_att. (g) AdvNet. (h) CycleGAN. (i) s4GAN. (j) Proposed SemiCDNet. The changed areas are marked in white, while the unchanged parts are in black.

effectiveness of incorporating the segmentation adversarial loss and entropy adversarial loss. Note that the F1-score reaches the peak when the λ value is 0.05 and 0.005 for the WHU building data set and the Google data set, respectively. After that, the F1-score shows a fluctuation downward trend by further increasing λ, denoting that too much adversarial loss can over-correct the initial results and lead to worse performance. Therefore, on the basis of the above analysis, λ is set to 0.05 and 0.005 for the two data sets, respectively.

2) Performance Analysis: To assess the effectiveness of the proposed SemiCDNet, extensive experimental results are summarized and analyzed for both data sets.

a) WHU Building Data Set: For a visual comparison, some typical CD results on testing samples are presented in Fig. 7. One can observe that there exist many missed detections and false alarms in the comparative methods. Whereas the proposed SemiCDNet achieves the best visual performance, as the change maps are more consistent with the ground truth. To be specific, compared with the baseline method of UNet++_att, missed detections such as missed buildings or holes are largely reduced by SemiCDNet, as shown in the first three rows of Fig. 7. In addition, compared with other methods, SemiCDNet generated change maps with more accurate boundaries, and reduced false alarms and uncertain areas, as presented in the last three rows of Fig. 7.

To quantitatively analyze the results, three evaluation metrics, F1, OA, and Kappa, were calculated and summarized in Table I based on different SSL settings. We can conclude that the quantitative results are consistent with the visual performance, and the proposed SemiCDNet achieves the best accuracies against the literature methods for all the SSL settings. The improvement is more remarkable when the labeled ratio is low. In particular, compared with the baseline method of UNet++_att, the proposed method obtains an increase of 5.41%, 1.37%, and 6.54% of F1, OA, and Kappa, respectively, when utilizing only 5% labeled data for training. These numbers are 2.63%, 0.69%, and 3.02% when employing 10% labeled data for training. Then, when the labeled ratio increases, the gap gradually decreases, with a gain of 1.01%, 0.18%, and 0.8% for F1, OA, and Kappa, respectively, when using 50% labeled data for training. The reasons of this behavior are: 1) the unlabeled information was fully exploited by enforcing the distribution consistency constraints of segmentation maps and entropy maps, whereby improving the capabilities of detecting real changes and suppressing uncertain changes and 2) the generalization and robustness of the generator was improved through competitive training between the generator and discriminators. Note that, compared with FC-EF-Res and FCN-PP, the baseline method of UNet++_att obtains better overall performance, with a mean increase of 1.37%, 1.08%, and 2.10% of F1, OA, and

TABLE I
SUMMARY OF THE QUANTITATIVE ACCURACY RESULTS FOR DIFFERENT METHODS WITH DIFFERENT LEVELS OF LABELED RATIOS ON WHU BUILDING DATA SET

Method	Labeled Ratio											
	5%			10%			20%			50%		
	F1	OA	Kappa	F1	OA	Kappa	F1	OA	Kappa	F1	OA	Kappa
FC-EF-Res	0.7715	0.9116	0.7177	0.8145	0.9328	0.7737	0.8451	0.9467	0.8129	0.8561	0.9513	0.8268
FCN-PP	0.7694	0.9212	0.7219	0.8110	0.9351	0.7719	0.8475	0.9512	0.8186	0.8539	0.9504	0.8240
UNet++_att	0.7749	0.9297	0.7308	0.8265	0.9448	0.7938	0.8499	0.9502	0.8200	0.8673	0.9577	0.8458
AdvNet	0.7599	0.9238	0.7149	0.8159	0.9404	0.7839	0.8400	0.9483	0.8093	0.8760	0.9584	0.8510
CycleGAN	0.7255	0.9191	0.6794	0.7807	0.9315	0.7405	0.8017	0.9338	0.7620	0.8330	0.9450	0.8000
s4GAN	0.8174	0.9425	0.7836	0.8493	0.9501	0.8194	0.8557	0.9523	0.8272	0.8772	0.9593	0.8528
SemiCDNet	**0.8290**	**0.9434**	**0.7962**	**0.8528**	**0.9517**	**0.8240**	**0.8657**	**0.9559**	**0.8403**	**0.8774**	**0.9595**	**0.8538**

Kappa, respectively, when using 10% labeled data for training. This demonstrates the advantages of dense skip connections in the baseline method. Furthermore, AdvNet achieves slightly worse results than the baseline method. The reason lies in the fact that instead of utilizing pretrained weights from the original literature, both the generator and the discriminator were trained from scratch without pretraining in our test. In addition, a two-stage GAN training strategy was adopted in AdvNet, where the generator was updated by both the labeled data and unlabeled data, while the discriminator was only updated using the labeled data, which may lead to an over-fitting phenomenon. Furthermore, a hard threshold has to be chosen for the self-training strategy, which is difficult to define. With regard to CycleGAN, it obtains even worse performance than AdvNet. The reason may be that two GANs were employed to enforce consistency between unpaired images and their segmentation maps, which are more prone to result in mode collapses problems during GAN training, thus greatly reducing the stability and leading to lower accuracies. On the contrary, s4GAN improved the baseline performance with an increase of 4.25%, 1.28%, and 5.28% for F1, OA, and Kappa, respectively, when utilizing 5% labeled data. This is due to the usage of a feature matching loss to stabilize GAN training, and the further incorporation of a self-training loss to balance generator and discriminator networks. Note that compared with s4GAN, the proposed SemiCDNet achieves better performance. The reason lies in the difference in exploiting unlabeled information: in s4GAN only high-confidence regions of unlabeled predictions are exploited by utilizing self-training strategy, whereas in SemiCDNet through enforcing entropy adversarial loss, both high-confidence and low-confidence regions are utilized by highlighting high-confidence regions and suppressing low-confidence regions. In addition, a hard threshold has to be defined to produce high-confidence regions in the self-training strategy, which requires troublesome trial and error procedure.

b) Google Data Set: For the benefit of visual performance comparison of different approaches, the results obtained on six typical test samples are illustrated in Fig. 8, where 20% labeled data are utilized for training. As one can see, the proposed SemiCDNet achieves the best performance against other comparative methods by reducing the missed detections and false alarms, thus generating finer change maps. Particularly, compared with the baseline method, some missed detections like holes in the buildings can be better detected with SemiCDNet, as shown the first three rows of Fig. 8. Furthermore, some false alarms can also be better removed, thus producing more accurate change maps, as presented in the last three rows of Fig. 8. Notably, compared with the baseline, s4GAN can also generate better change maps, whereas the results of AdvNet and CycleGAN are not stable or even worse.

Table II reports the quantitative evaluation results of different methods based on the defined SSL settings. We can conclude that the proposed SemiCDNet achieves the best performance against other comparative methods, exhibiting the highest F1, OA, and Kappa values in all the SSL settings. In particular, compared with the UNet++_att baseline, SemiCDNet yields an improvement of 1.86%, 0.71%, and 1.95% for F1, OA, and Kappa, respectively, when using 10% labeled data for training. The improvement is of 1.24%, 0.68%, and 1.62%, respectively, when employing 20% labeled data for training. This denotes the effectiveness of exploiting unlabeled information for improving the CD results by combined use of segmentation adversarial loss and entropy adversarial loss. Note that by further increasing the labeled ratio, the gap gradually decreases, with a gain of 0.76%, 0.37%, and 0.88% for F1, OA, and Kappa, respectively, when utilizing 40% labeled data for training. However, compared with the WHU building data set, the gains are not as prominent, with the largest F1 gain smaller than 2%. We argue this can mainly be attributed to the less reliable initial predictions due to the complex and heterogeneous scenes in the Google data set, which increase the difficulty of matching the feature distribution of the labeled and unlabeled data. Note that, when compared with FC-EF-Res and FCN-PP, the baseline method of UNet++_att also obtains better overall performance, with a mean increase of 0.84%, 0.70%, and 1.34% of F1, OA, and Kappa, respectively, when using 20% labeled data for training. On the contrary, AdvNet still achieves worse performance than the baseline due to the drawbacks of the two-stage GAN training. Nevertheless, CycleGAN obtains an even inferior performance, as mode collapse problems are more prone to happen due to the incorporation of unpaired images during GAN training. On the contrary, s4GAN, which adopts feature

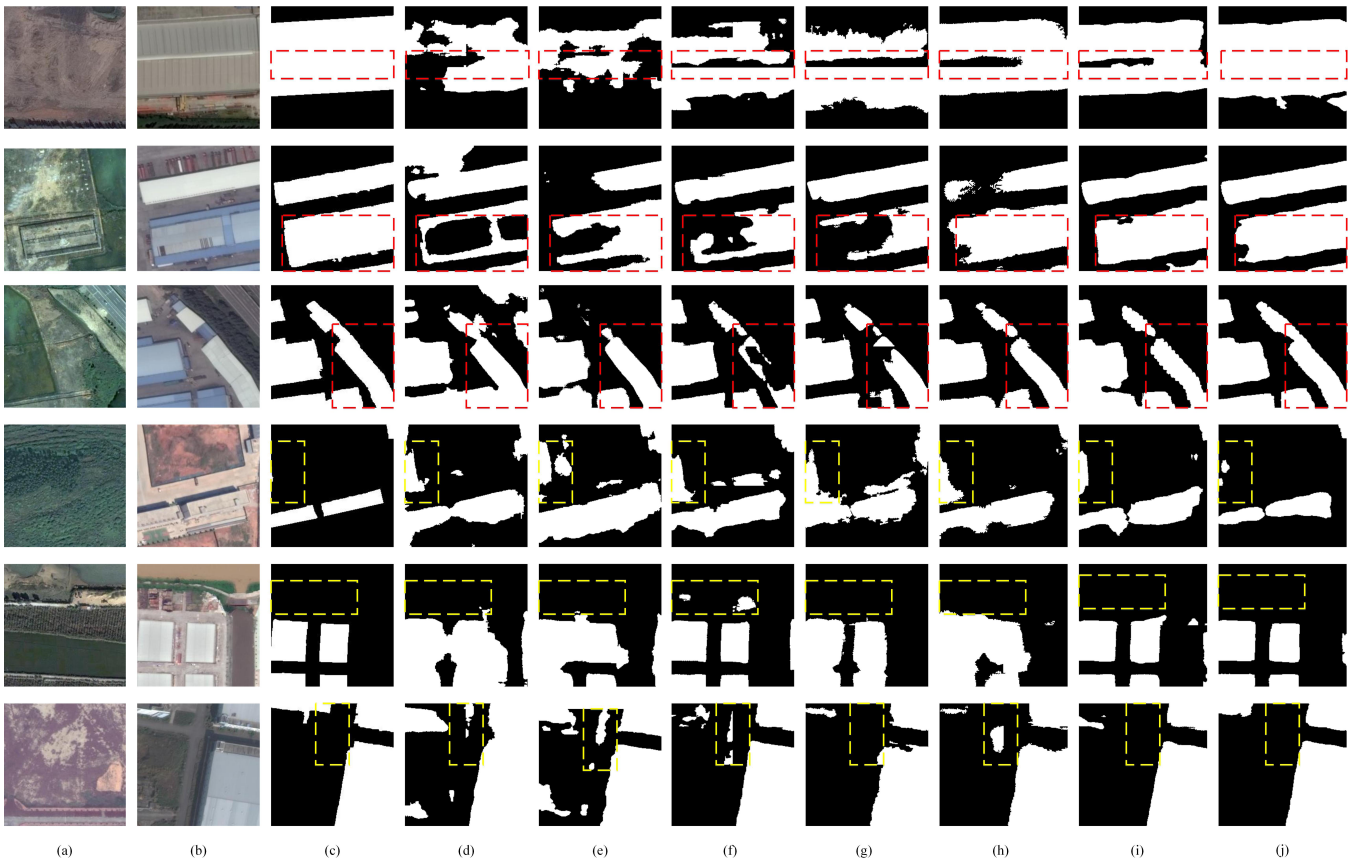


Fig. 8. Visual comparisons of CD maps by different approaches on Google data set using 20% labeled images for training. (a) Image T_1. (b) Image T_2. (c) Reference change map. (d) FC-EF-Res. (e) FCN-PP. (f) UNet++_att. (g) AdvNet. (h) CycleGAN. (i) s4GAN. (j) Proposed SemiCDNet. The changed areas are marked in white, while the unchanged parts are in black.

TABLE II

SUMMARY OF THE QUANTITATIVE ACCURACY RESULTS FOR DIFFERENT METHODS WITH DIFFERENT LEVELS OF LABELED RATIOS ON GOOGLE DATA SET

Method	Labeled Ratio											
	10%			20%			40%			60%		
	F1	OA	Kappa	F1	OA	Kappa	F1	OA	Kappa	F1	OA	Kappa
FC-EF-Res	0.7381	0.8460	0.6261	0.7810	0.8710	0.6822	0.7870	0.8613	0.6781	0.8098	0.8747	0.7078
FCN-PP	0.7295	0.8385	0.6144	0.7977	0.8743	0.7001	0.8014	0.8745	0.7029	0.8163	0.8860	0.7253
UNet++_att	0.7441	0.8550	0.6415	0.7978	0.8796	0.7045	0.8167	0.8868	0.7258	0.8281	0.8898	0.7384
AdvNet	0.7173	0.8273	0.5939	0.7882	0.8657	0.6834	0.8034	0.8742	0.7022	0.8267	0.8795	0.7258
CycleGAN	0.7145	0.8483	0.6131	0.7641	0.8545	0.6562	0.7780	0.8625	0.6740	0.7923	0.8656	0.6877
s4GAN	0.7397	0.8369	0.6214	0.8070	0.8829	0.7152	0.8224	0.8895	0.7320	0.8351	0.8935	0.7466
SemiCDNet	**0.7627**	**0.8621**	**0.6610**	**0.8102**	**0.8864**	**0.7207**	**0.8243**	**0.8905**	**0.7346**	**0.8389**	**0.8964**	**0.7525**

match loss and self-training strategy, achieves better performance than the baseline, with a gain of 0.92%, 0.33%, and 1.07% for F1, OA, and Kappa, respectively, when employing 20% labeled data for training.

Fig. 9 presents the training time of different methods on the two data sets. We can conclude that FC-EF-Res and FCN-PP methods, which exploit only one simple generator based on UNet, require the lowest training time. On the contrary, the proposed baseline method of UNet++_att requires a longer training time due to the higher complexity of the network architecture by including dense skip connections and attention units. Furthermore, after including adversarial learning, the training time is clearly increased for both AdvNet and s4GAN. In particular, the training time of CycleGAN is sharply increased, due to the fact that the amount of labeled and unlabeled data has to be enlarged to the same level by random replication so that CycleGAN model can be trained through unpaired images. In addition, two generators and two discriminators have to be trained by using six kinds of losses, which greatly increases the training burden. Note that,

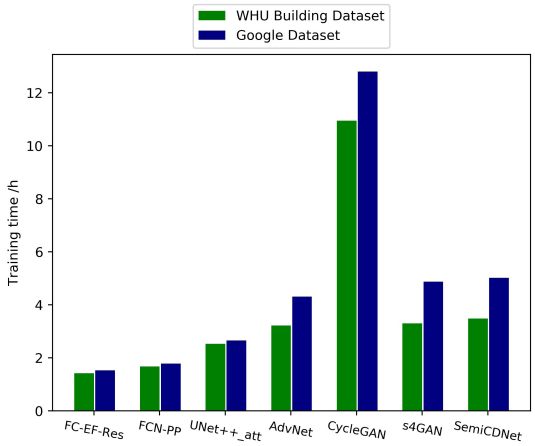


Fig. 9. Comparison of the training time required by different methods.

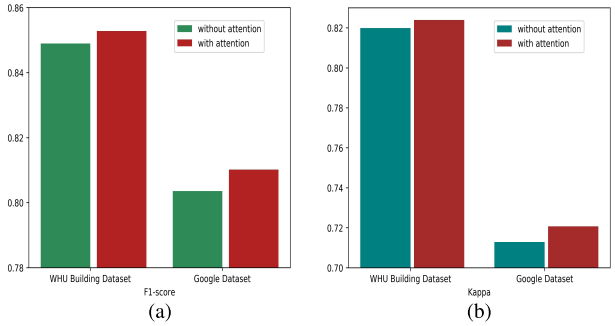


Fig. 10. Effects of the attention mechanism on the accuracy of the proposed method. (a) F1-score. (b) Kappa.

compared with AdvNet and s4GAN, SemiCDNet, which is based on one generator and two discriminators, only requires slightly more time to be trained. Therefore, the proposed SemiCDNet achieves a better balance between accuracy and computational efficiency.

E. Discussion

1) Effect of Attention Mechanism: In order to combine multilevel features effectively, an attention mechanism is proposed. Fig. 10 presents the influence of attention mechanism on the performance of the model trained with 10% labeled data for the WHU building data set and 20% labeled data for the Google data set. We can conclude that the model performance can be improved for both data sets by utilizing the proposed attention mechanism, with a F1-score increase of 0.38% and 0.66%, and a Kappa increase of 0.41% and 0.78% for the WHU building data set and the Google data set, respectively. That is due to the fact that the weights of features maps from different levels are learned through the attention mechanism, thus highlighting useful features and suppressing redundant information. Therefore, the proposed attention mechanism is effective to improve CD performance.

2) Effect of Adversarial Loss: To validate the effectiveness of different adversarial losses, an ablation study was carried out, where the model was trained without segmentation adversarial loss or entropy adversarial loss. On such experiments, we investigated the performance of the SemiCDNet trained with 10% labeled data for the WHU building data set and 20% labeled data for the Google data set. Table III summarizes

TABLE III

ABLATION STUDY OF DISCRIMINATORS ON THE WHU BUILDING DATA SET (WITH 10% LABELED TRAINING DATA) AND THE GOOGLE DATA SET (WITH 20% LABELED TRAINING DATA)

Method	WHU Building Dataset	Google Dataset
w/o D_s D_e	0.8265	0.7978
w/o D_s	0.8490	0.8061
w/o D_e	0.8456	0.8003
Proposed	**0.8528**	**0.8102**

the quantitative results of the ablation study in terms of F1-score. One can observe that the proposed method achieves the best performance on both data sets, with the F1 values reaching 0.8528 and 0.8102, respectively. When removing $\boldsymbol{D_s}$, segmentation adversarial loss was removed during generator training, and the F1 values decreased to 0.8490 and 0.8061 for the two data sets, respectively. When removing $\boldsymbol{D_e}$, entropy adversarial loss was ignored during generator training and the F1 values reduced to 0.8456 and 0.8003 for the two data sets, respectively. This confirms that both $\boldsymbol{D_s}$ and $\boldsymbol{D_e}$ are beneficial to improve the generalization abilities of the generator, and the best performance can be achieved by combining them. Note that when both discriminators $\boldsymbol{D_s}$ and $\boldsymbol{D_e}$ were removed, only labeled data were used for training, and the F1 values further reduced to 0.8265 and 0.7978, respectively. This confirms the significance of exploiting unlabeled information to regularize the generation model.

3) Direct Entropy Minimization Versus Entropy Adversarial Learning: In dealing with low-entropy constraints, two main strategies can be adopted: 1) direct entropy minimization (DEM) and 2) entropy adversarial learning (EAL). The former is employed by minimizing the entropy of the unlabeled entropy maps directly through a pixel-wise entropy loss. Instead, the latter introduces an adversarial loss to align the feature distributions between the labeled and unlabeled entropy maps. To verify the effectiveness of the proposed entropy adversarial loss, we carry out a comparative study with 10% labeled data for the WHU buildling data set and 20% labeled data for the Google data set. Fig. 11 illustrates the quantitative performance of different approaches in terms of F1-score and Kappa. One can observe that compared with DEM, EAL obtains a better performance, with an increase of F1-score of 0.89% and 0.52%, and a gain of Kappa of 1.02% and 0.62% for WHU building data set and Google data set, respectively. The reason for these phenomena is that DEM treats each pixel independently when calculating the entropy loss, thus ignoring local structural information. On the contrary, in EAL, by enforcing the feature distribution consistency between unlabeled and labeled entropy maps, local structural consistency is well considered and the entropy is minimized indirectly.

4) FCN Discriminator Versus FC Discriminator: GAN has a high ability in distribution modeling. However, it is difficult to train the GAN due to mode collapse problems, where the balance of discriminator plays a significant role. For our CD task based on GAN, two kinds of discriminators can be chosen, namely FCN discriminator (FCN-Dis) and FC

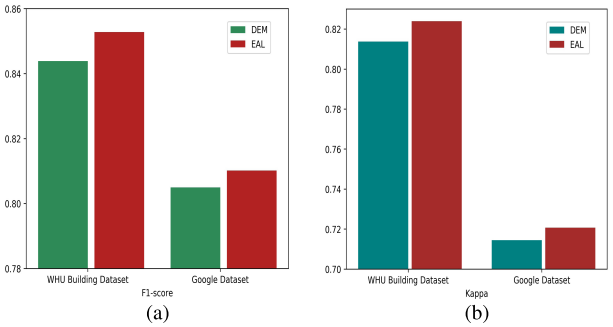


Fig. 11. Effects of DEM and EAL on the accuracy of the proposed method. (a) F1-score. (b) Kappa.

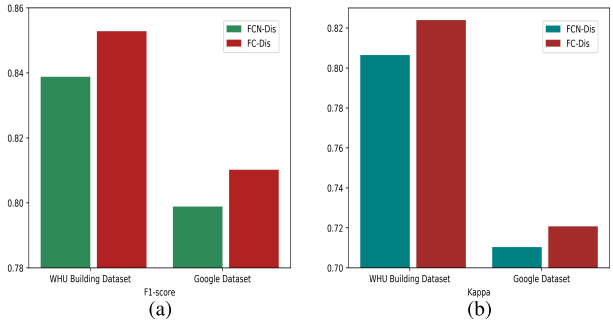


Fig. 12. Effects of FCN-Dis and FC-Dis on the accuracy of the proposed method. (a) F1-score. (b) Kappa.

discriminator (FC-Dis). The FCN is utilized for the FCN-Dis, which can produce a single-channel map of the same size as the input image. The outputs are the probability values that denote if each pixel is real or fake. Differently, in the FC-Dis, only a feature vector is produced, which denotes if the input image is real or fake. In order to verify the effect of different discriminators on our proposed approach, a comparative study was done with 10% labeled data for the WHU buildling data set and 20% labeled data for the Google data set. Fig. 12 presents the quantitative performance of the different approaches in terms of F1-score and Kappa. As one can see, FC-Dis achieves a better performance than the FCN-Dis on both data sets, with an increase of F1-score of 1.67% and 1.41%, and a Kappa gain of 2.17% and 1.45% for the WHU building data set and Google data set, respectively. This may be explained by the fact that the FCN discriminator produces much more information in output space than the FC discriminator, leading to a severe imbalance between the generator and the discriminator. Consequently, the performance of the generator degrades due to unreasonable guidance from the discriminator.

V. Conclusion

A large amount of labeled data is needed in the recently developed supervised FCN-based CD methods. These data are difficult and tedious to obtain in real application scenarios. To address such an issue in this article, we have proposed a novel semisupervised CD Network (SemiCDNet) based on FCN and GANs. Instead of concatenating multilevel features directly, an improved UNet++ network with attention mechanism has been proposed to serve as the generator, thus producing reliable initial segmentation maps. To exploit the potential of the unlabeled data, two discriminators have been introduced, one for encouraging segmentation output feature distribution consistency and the other for suppressing uncertain areas of the change map for the unlabeled samples. Through competing training, the generator gradually learned the information from both the labeled and the unlabeled samples, thus generating finer change maps. The effectiveness and reliability of the proposed approach have been verified on both VHR aerial and satellite data sets. The experimental results demonstrated the superiority against some SOTA methods. Note that, in addition to SSL, other small sample learning techniques such as few-shot learning and self-supervised learning, have opened up new perspectives of improving model generalization ability. In the future, more recent development of such techniques will be exploited to further improve the CD performance.

References

[1] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[2] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, "A comprehensive change detection method for updating the national land cover database to circa 2011," *Remote Sens. Environ.*, vol. 132, pp. 159–175, May 2013.

[3] G. Chen and G. J. Hay, "An airborne lidar sampling strategy to model forest canopy height from Quickbird imagery and GEOBIA," *Remote Sens. Environ.*, vol. 115, no. 6, pp. 1532–1542, Jun. 2011.

[4] S. Ji, Y. Shen, M. Lu, and Y. Zhang, "Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples," *Remote Sens.*, vol. 11, no. 11, p. 1343, Jun. 2019.

[5] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.

[6] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, Sep. 2015.

[7] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, Apr. 2015.

[8] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[9] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008.

[10] P. R. Marpu, P. Gamba, and M. J. Canty, "Improving change detection results of IR-MAD by eliminating strong changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 799–803, Jul. 2011.

[11] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.

[12] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, Oct. 2009.

[13] P. Lv, Y. Zhong, J. Zhao, and L. Zhang, "Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4002–4015, Jul. 2018.

[14] Y. Zhang, D. Peng, and X. Huang, "Object-based change detection for VHR images based on multiscale uncertainty analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 13–17, Jan. 2018.

[15] G. Xian and C. Homer, "Updating the 2001 national land cover database impervious surface products to 2006 using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 114, no. 8, pp. 1676–1686, Aug. 2010.

[16] D. Wen, X. Huang, A. Zhang, and X. Ke, "Monitoring 3D building change and urban redevelopment patterns in inner city areas of Chinese megacities using multi-view satellite imagery," *Remote Sens.*, vol. 11, no. 7, p. 763, Mar. 2019.

[17] M. Janalipour and M. Taleai, "Building change detection after earthquake using multi-criteria decision analysis based on extracted information from high spatial resolution satellite images," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 82–99, Jan. 2017.

[18] P. Xiao, M. Yuan, X. Zhang, X. Feng, and Y. Guo, "Cosegmentation for object-based building change detection from high-resolution remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1587–1603, Mar. 2017.

[19] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.

[20] S. Du *et al.*, "Building change detection using old aerial images and new LiDAR data," *Remote Sens.*, vol. 8, no. 12, p. 1030, Dec. 2016.

[21] Y. Hua, L. Mou, and X. X. Zhu, "Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 149, pp. 188–199, Mar. 2019.

[22] C. Tao, J. Qi, Y. Li, H. Wang, and H. Li, "Spatial information inference net: Road extraction using road-specific contextual information," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 155–166, Dec. 2019.

[23] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.

[24] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 305–319, Dec. 2018.

[25] Z. Huang *et al.*, "Unidirectional variation and deep CNN denoiser priors for simultaneously destriping and denoising optical remote sensing images," *Int. J. Remote Sens.*, vol. 40, no. 15, pp. 5737–5748, Aug. 2019.

[26] C. Peng, Y. Li, L. Jiao, Y. Chen, and R. Shang, "Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2612–2626, Aug. 2019.

[27] 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, p. 1382, Jun. 2019.

[28] B. Hou, Y. Wang, and Q. Liu, "Change detection based on deep features and low rank," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2418–2422, Dec. 2017.

[29] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[30] Y. Lei, X. Liu, J. Shi, C. Lei, and J. Wang, "Multiscale superpixel segmentation with deep features for change detection," *IEEE Access*, vol. 7, pp. 36600–36616, 2019.

[31] M. Gong, T. Zhan, P. Zhang, and Q. Miao, "Superpixel-based difference representation learning for change detection in multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658–2673, May 2017.

[32] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2019.

[33] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 982–986, Jun. 2019.

[34] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2, pp. 565–571, May 2018.

[35] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 4063–4067.

[36] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[37] S. H Khan, X. He, F. Porikli, M. Bennamoun, F. Sohel, and R. Togneri, "Learning deep structured network for weakly supervised change detection," 2016, *arXiv:1606.02009*. [Online]. Available: http://arxiv.org/abs/1606.02009

[38] J. Dai, K. He, and J. Sun, "BoxSup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1635–1643.

[39] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3159–3167.

[40] D. Peng and H. Guan, "Unsupervised change detection method based on saliency analysis and convolutional neural network," *J. Appl. Remote Sens.*, vol. 13, no. 2, 2019, Art. no. 024512.

[41] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.

[42] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35–49, Dec. 2019.

[43] B. Hou, Y. Wang, and Q. Liu, "A saliency guided semi-supervised building change detection method for high resolution remote sensing images," *Sensors*, vol. 16, no. 9, p. 1377, Aug. 2016.

[44] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, Jan. 2019.

[45] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.

[46] H.-C. Li, G. Yang, W. Yang, Q. Du, and W. J. Emery, "Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 167–179, Feb. 2020.

[47] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363–375, Jan. 2015.

[48] W. Zhang, X. Lu, and X. Li, "A coarse-to-fine semi-supervised change detection for multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3587–3599, Jun. 2018.

[49] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[50] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[51] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[52] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[54] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[55] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-net: Multi-scale densely connected U-net for biomedical image segmentation," 2018, *arXiv:1812.00352*. [Online]. Available: http://arxiv.org/abs/1812.00352

[56] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," 2018, *arXiv:1802.07934*. [Online]. Available: http://arxiv.org/abs/1802.07934

[57] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-supervised semantic segmentation with high- and low-level consistency," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Dec. 17, 2020, doi: 10.1109/TPAMI.2019.2960224.

[58] A. K. Mondal, A. Agarwal, J. Dolz, and C. Desrosiers, "Revisiting CycleGAN for semi-supervised segmentation," 2019, *arXiv:1908.11569*. [Online]. Available: http://arxiv.org/abs/1908.11569

[59] S. Chen, G. Bortsova, A. G.-U. Juárez, G. van Tulder, and M. de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2019, pp. 457–465.

[60] K. Chen, Z. Zhou, C. Huo, X. Sun, and K. Fu, "A semisupervised context-sensitive change detection technique via Gaussian process," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 236–240, Mar. 2013.
[61] S. Ghosh, M. Roy, and A. Ghosh, "Semi-supervised change detection using modified self-organizing feature map neural network," *Appl. Soft Comput.*, vol. 15, pp. 1–20, Feb. 2014.
[62] M. Gong, Y. Yang, T. Zhan, X. Niu, and S. Li, "A generative discriminatory classified network for change detection in multispectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 321–333, Jan. 2019.
[63] J. Liu *et al.*, "Semi-supervised change detection based on graphs with generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 74–77.
[64] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784
[65] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understand.*, vol. 187, Oct. 2019, Art. no. 102783.

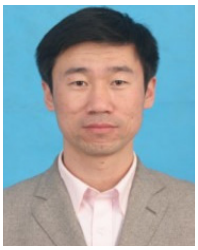
Daifeng Peng received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017.

He is a Lecturer with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China. He is also a Post-Doctoral Fellow with Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. His research interests include machine learning, deep learning, multitemporal image analysis, intelligent interpretation and information extraction from high resolution remote-sensing imagery.

Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree (*summa cum laude*) in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is the Founder and the Director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His research interests are in the areas of remote sensing, radar and SAR, signal processing, machine learning and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the Principal Investigator of many research projects. Among the others, he is the Principal Investigator of the *Radar for icy Moon exploration* (RIME) instrument in the framework of the *JUpiter ICy moons Explorer* (JUICE) mission of the European Space Agency. He is the author (or coauthor) of 276 scientific publications in referred international journals (209 in IEEE journals), more than 330 papers in conference proceedings, and 22 book chapters. He is editor/coeditor of 18 books/conference proceedings and one scientific book. He was invited as keynote speaker in more than 40 international conferences and workshops. Since 2009, he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS).

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since that he was a recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award and the 2017 IEEE IGARSS Symposium Prize Paper Award. He was a Guest Coeditor of many special issues of international journals. He is the Cofounder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is a member of the Permanent Steering Committee of this series of workshops. Since 2003, He has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the Founder of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE for which he has been Editor-in-Chief between 2013 and 2017. Currently, he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has been Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society between 2012 and 2016. His papers are highly cited, as proven form the total number of citations (more than 34 000) and the value of the h-index (87) (source: Google Scholar).



Yongjun Zhang (Member, IEEE) received the B.S. degree in Geodesy, the M.S. degree in Geodesy and Surveying Engineering, and the Ph.D. degree in Geodesy and Photography from Wuhan University, Wuhan, China, in 1997, 2000, and 2002, respectively.

He is the Vice Dean of the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China. Since 2006, he has been the Full Professor of the School of Remote Sensing and Information Engineering, Wuhan University. From 2014 to 2015, he was a Senior Visiting Fellow with the Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada. From 2015 to 2018, he was a Senior Scientist with Environmental Systems Research Institute, Inc. (ESRI), Redlands, CA, USA. He has published more than 150 research articles and one book. He holds 25 Chinese patents and 26 copyright registered computer software. His research interests include aerospace and low-attitude photogrammetry, image matching, combined block adjustment with multisource data sets, object information extraction and modeling with artificial intelligence technologies, integration of LiDAR point clouds and images, and 3-D city model reconstruction.

Dr. Zhang is the PI Winner of the Second-Class National Science and Technology Progress Award in 2017, and the PI Winner of the Outstanding-Class Science and Technology Progress Award in Surveying and Mapping in 2015. In recent years, he has also served as the session chair of above 20 international workshops or conferences. He has been frequently serving as a referee for over 20 international journals.

Haiyan Guan (Senior Member, IEEE) received the Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, and in geomatics from the University of Waterloo, Waterloo, ON, Canada, respectively.

She is a Professor with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, China. Her research interests include information extraction from LiDAR point clouds and from earth observation images. She has published more than 40 research papers in refereed journals, books, and proceedings, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (IEEE-TITS), IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *ISPRS Journal of Photogrammetry and Remote Sensing*, and IGARSS & ISPRS proceedings.

Haiyong Ding received the Ph.D. degree in cartography and geography information engineering from China Mining University, Xuzhou, China, in 2008.

He is a Professor of remote sensing with the School of Remote Sensing and Geomatics Engine, Nanjing University of Information Science and Technology, Nanjing, China. His research interests focus on urbanization, urban heat island, land use/land cover change detection, as well as the time series remote sensing data analysis.

Xu Huang received the B.S. degree in remote sensing science and technology and the Ph.D. degree in photogrammetry from Wuhan University, Wuhan, China, in 2010 and 2016, respectively.

He is an Engineer with Wuhan Engineering Science and Technology Institute, Wuhan, China. From March/2018 to March/2020, he was a Post-Doc with the Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA. He has published more than 30 research articles. He holds 13 Chinese patents. His research interests include aerospace and low-attitude photogrammetry, image matching, 3-D reconstruction, and LiDAR point processing.

Dr. Huang is the Winner of 2019 IEEE GRSS Data Fusion Contest, and the Winner of the best youth paper in the 2017 ISPRS Geospatial Week Workshop.