Unsupervised Global Difference Modeling for Image Change Detection

Wenhua Zhang, *Member, IEEE*, Qihao Weng, *Fellow, IEEE*, Jia Liu, *Member, IEEE*, Zeng Qu, Yong Li, Jinjin Chai, and Liang Xiao, *Senior Member, IEEE*

***Abstract*—In change detection, impact of nonintrinsic changes such as those caused by illumination, season, and viewing angle variances are common in practice but also a great challenge for change detection methods. In this article, we propose a novel unsupervised image change detection method by modeling global difference information to deal with such nonintrinsic changes. Comparing global features can mitigate the impact of them due to the global consistency of them in the same scene at the same time. But global modeling for change detection also faces the challenges of feature learning with limited data and difficulty in generating pixelwise changed regions. To overcome the challenges, first, we use a backbone network to capture the global features of bitemporal images. Then, an energy function is designed with a masked difference between the two features and a margin-aware constraint in order to align the global features and meanwhile maintain detail information. To train the network with only two images, we propose an adversarial learning method by introducing a generalization network that consecutively generates two images that can minimize the energy. Then, a new loss function is derived to alternately train the feature learning network and generalization network. Second, after learning with bitemporal images, it is also important to generate the pixelwise changed regions. Then, we design a difference mapping method that maps the changed regions from global difference. Experiments on different types of data by comparing with both supervised and unsupervised methods demonstrate the effectiveness of the proposed method.**

***Index Terms*—Image change detection, neural networks, probabilistic model, unsupervised learning.**

Received 26 March 2025; revised 8 June 2025; accepted 27 June 2025. Date of publication 14 July 2025; date of current version 5 August 2025. This work was supported in part by the National Nature Science Foundation of China under Grant 62302219 and Grant 62276133 and Grant 62471236, and Grant 62471235, in part by the Nature Science Foundation of Jiangsu Province, China under Grant BK20220948, in part by the Frontier Technologies Research and Development Program of Jiangsu under Grant BF2024070, and in part by the Fundamental Research Funds for the Central Universities under Grant 30924010918, and Grant 4009002509. *(Corresponding authors: Jia Liu.)*

Wenhua Zhang, Jia Liu, and Liang Xiao are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: omegaliuj@gmail.com).

Qihao Weng is with the Research Centre for Artificial Intelligence in Geomatics, Hong Kong Polytechnic University, Hong Kong Hong Kong.

Zeng Qu is with the State Key Laboratory of Extreme Environment Optoelectronic Dynamic Testing Technology and Instrument, Taiyuan 030051, China, and also with the School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China.

Yong Li is with the School of Computer Science and Engineering, Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications, Southeast University, Nanjing 211189, China (e-mail: mysee1989@gmail.com).

Jinjin Chai is with the School of Air Defence and Antimissile, Air Force Engineering University, Xi'an 710051, China.

Digital Object Identifier 10.1109/JSTARS.2025.3588154

I. Introduction

IMAGE change detection is to detect regions of change in images of the same scene taken at different times [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. It is significant in many applications, including video surveillance [13], medical diagnosis and treatment [14], and especially in remote sensing monitoring and land use analysis [15], [16], [17]. Change detection methods can be divided into supervised or semisupervised and unsupervised ones according to whether they need manually annotated samples to learn. Supervised methods can adapt to different complex scenarios with the assistance of annotated training samples but require efforts to collect and annotate large scale dataset. Unsupervised ones are widely applied due to that they can detect the changes given only two multitemporal images but the accuracy of them highly depends on the effectiveness of preprocessing methods, such as geometric adjustments (coregistration) [18] and radiometric adjustments (denoising, atmospheric corrections, normalization, etc.) [19].

With the excellent performance of deep neural networks in various applications [20], [21] and large-scale annotated data, many deep learning based supervised change detection methods have been proposed [22], [23], [24], [25], [26]. We also proposed an effective self-weighted spatial-temporal attention neural network in [9] to fully explore the difference information. With meticulously designed architectures and objective functions, deep learning based changed detection methods have achieved excellent performance. But they largely depend on the large-scale annotated training datasets. For example, the widely used LEVIR-CD [23] and WHU[1] datasets were constructed for building changes. CDD dataset [27] was constructed for robustness to season changes of change detection networks. To reduce annotation efforts, recently, many weakly-supervised change detection methods have been proposed where only image-level labels (changed or unchanged within the pairs of images) are required [28], [29].

Different from other tasks such as the semantic segmentation and scene classification, in changed detection, only the changed and unchanged regions instead of various manually defined classes should be distinguished. As a consequence, unsupervised ones are still mainstream change detection methods. But due to the feature diversity within the same class, they suffer larger challenge in designing efficient methods than supervised ones.

[1] [Online]. Available: https://study.rsgis.whu.edu.cn/pages/download/building_dataset.html

Traditional unsupervised change detection methods usually generate the difference information via comparing pixel pairs or manually designed local feature pairs, such as the classical change vector analysis [30] and its derivatives [31], [32]. While such comparison is not robust to many nonintrinsic changes such as shadow variance caused by different illumination conditions, seasonal changes, view angle variances, etc. To align the widespread feature inconsistency such as seasonal changes and heterogeneous distributions, we proposed a symmetric network in [33] and improved it in [34] to extract features for comparison and train it via dynamic pseudolabels. The pseudolabels are trainable parameters and learned along with the network parameters. We also integrated it in a probabilistic model for change detection in [10]. Then, many methods have been proposed based on different feature learning methods, such as dictionary learning [35], autoencoder [36], slow feature analysis [37], and graph learning [38], [39]. Except to compare multitemporal feature pairs, structure or topological information can also be compared. For example, nonlocal patch-based graph is constructed as the structure in [39]. Topological structure of multitemporal images is explored by linking class relationships and change labels of pairwise superpixels in [40]. Recently, deep neural networks have been widely applied in unsupervised change detection due to their feature learning capability for reducing nonintrinsic changes. However, training a deep network is a great challenge with only unannotated bitemporal images. UCDFormer proposed in [41] trains a Transformer via regional patches with selected reliable pseudochanged and unchanged pixels. In [42], iterative refining modules are proposed to gradually refine pseudochange maps via cross-reconstruction and bidirectional contrastive constraints. Moreover, progressive learning is also used to train deep networks for unsupervised change detection [43], [44].

The above methods can avoid many nonintrinsic changes caused by sensor noise, illumination variation, nonuniform attenuation, atmospheric absorption, and even heterogeneous sensors. However, misalignment caused by view angle variance is still a great challenge for unsupervised change detection methods. As a consequence, in this article, we propose a new unsupervised change detection framework by modeling global features instead of comparing locally or regionally to generate pixelwise change information. Many variances are globally consistent due to that the ground objects in the same image are captured at the same time. Then, by aligning them many globally consistent nonintrinsic changes can be mitigated such as illumination, season, and even view-angle changes. However, global comparison faces the problems of network training and pixelwise difference generation. Then, the proposed method is designed and the contributions are as follows.

1) We propose a new unsupervised change detection framework by modeling global features to deal with global changes.
2) We propose a global difference based energy function with a margin-aware constraint. On the one hand, it is used to align the global features. On the other hand, the constraint is used to retain the margins and details of changed regions. Moreover, to train the backbone network, inspired from GAN, we propose an adversarial learning method to learn the features that can be well aligned and meanwhile capture the distribution of input bitemporal images.
3) To generate the pixelwise changes, we propose a difference mapping method that maps the global difference to local pixelwise difference. Experiments on images in difference scenarios by comparing with both unsupervised and supervised methods demonstrate the promising performance of the proposed method.

The rest of this article is organized as follows. Change detection, related work, and motivation are introduced in Section II. The detailed architecture, model, and learning process are described and analyzed in Section III. Experiments on various types of images are conducted in Section IV. Finally, Section V concludes this article.

II. Preliminaries

To clearly tell the whole story, in this section, we first introduce the change detection problem and motivation of the proposed method. Then, the motivation of proposing adversarial learning is discussed.

A. Change Detection and Motivation

In this article, we only consider change detection between two images. For unsupervised change detection, given a pair of preprocessed (the preprocessing mainly includes coregistration) images $I_1 \in \mathbf{R}^{W\times H\times N}$ and $I_2 \in \mathbf{R}^{W\times H\times N}$ that are captured at the same region while in different times, the aim of it is to generate a binary change map $Y \in \mathbf{B}^{W\times H}$ that labels the changed pixels. W, H, and N, respectively, denote the width, height, and number of channels of images. $\mathbf{R}$ denotes the real set and $\mathbf{B}$ denotes the binary set. To generate Y, unsupervised methods usually first generate a difference image $D \in \mathbf{R}^{W\times H}$ that indicates the pixelwise difference degree between I_1 and I_2 by comparing them. The aim of the proposed method is also to generate a difference image.

To train deep networks and generate pixelwise difference, existing unsupervised methods usually model multitemporal images with local or regional features. In this article, we innovatively propose to model global difference. Local or regional learning and comparison is difficult to distinguish intrinsic and nonintrinsic changes because they both may have large feature differences. As shown in Fig. 1, changed and unchanged regions show similar pairwise appearance. Moreover, the local location changes caused by view angle variance also influence the comparison of local features, as shown in Fig. 1, where the unchanged regions in the same location have distinct appearance. As introduced above, many variances are globally consistent due to that the ground objects in the same image are captured at the same time. Analyzing and aligning global features can mitigate many global changes in a larger feature extraction scale, especially for images of large areas or captured within 3-D scenes. For example, the change of season results in color variance of vegetation, as shown in 1. Such a change will significantly influence the detection of real intrinsic changes. But

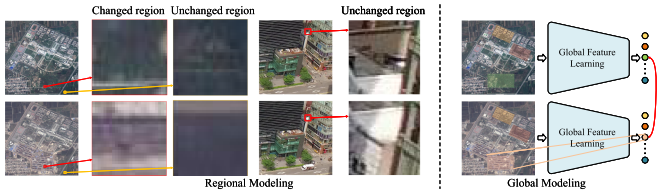


Fig. 1. Illustration of the motivation. Regional modeling is difficult to distinguish nonintrinsic and intrinsic changes. Moreover, it is not robust to view angle variances for images captured within 3-D scenes. Global modeling analyzes images in larger feature extraction scale which is robust to many nonintrinsic changes.

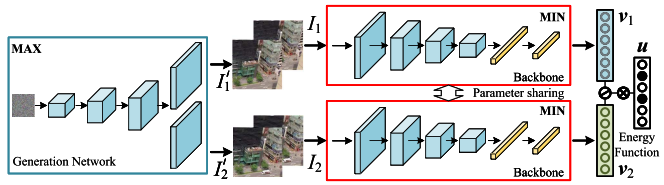


Fig. 2. Architecture of the proposed unsupervised change detection network. A backbone network is used to extract the global features of the bitemporal images, respectively. To train the backbone network for useful global features, we introduce the generation network and design an energy driven adversarial learning method to align the global features and meanwhile capture the distribution of bitemporal images.

it is usually pervasive over the acquired scene and its influence can be mitigated via global aligning. Viewing angle change results in local position variance between bitemporal images which can also be mitigated via global feature modeling.

B. GAN and Adversarial Learning

To align the global features, we design a global difference based energy function. But since all the parameters in networks are trainable, directly minimizing the difference will result in identical but useless features. So, except aligning the features, it is also expected that the learned features can well capture the distribution of input images. As a consequence, in this article, we propose an adversarial learning method inspired from the GAN.

GAN is composed of a generator $\mathcal{G}(\cdot)$ and a classifier $\mathcal{D}(\cdot)$. The generator is used to generate a data x' from a noise z: $x' = \mathcal{G}(z)$. The classifier is used to discriminate real data x from generated data x'. The two networks are trained via a min-max objective function as follows:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_x \log(\mathcal{D}(x)) + \mathbb{E}_z \log(1 - \mathcal{D}(\mathcal{G}(z))) \tag{1}$$

where $\mathbb{E}_x$ and $\mathbb{E}_z$, respectively, denote the expectation of training data and noise. θ_D and θ_G, respectively, denote the network parameters of classifier and generator. By maximizing the negative cross entropy loss, the classifier is trained to better classify x and x'. Then, by minimizing the loss of x', the generator is expected to generate a data that is more close to x. Then, by alternately training, the generator is able to capture the distribution of the training data and generate almost real data.

In this article, we also consider to capture the distribution of input data in order to learn useful global features. So we introduce a generation network and train the networks in adversarial manner.

III. Global Difference Modeling and Change Detection

The proposed method is based on deep neural networks, which can extract global features for bitemporal images of a large scene. For unsupervised change detection, it is necessary to train the networks with only bitemporal images to capture the useful global features for comparison. We propose an energy driven adversarial learning method to train the networks. Following we detail the architectures, energy function, adversarial learning, and change detection using the trained networks, respectively.

A. Overall Structure of the Method

The architecture of the proposed method is shown in Fig. 2. The main backbone is based on a deep neural network with input of image and output of a global feature vector. Many existing network architectures can be used such as various convolutional neural networks (CNNs) or Transformers. As shown in Fig. 2, given an image I_t (for change detection $t \in \{1, 2\}$), with the feature extracting network, a feature vector $\boldsymbol{v}_t$ is generated via the multilayer architecture: $\boldsymbol{v}_t = f_\theta(I_t)$ where θ is the network parameter set. For change detection, we aim to compare the two images globally and then $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are compared to compute the difference of global features.

To train the backbone, we first define an energy function based on the difference of the two features in order to align the global features for global consistency. But the details are severely lost in the global feature, which significantly influences the margin of changed regions in pixelwise change map. As a consequence, we define a margin-aware constraint in the energy function to maintain the details of changed regions. However, since the backbone network is trainable, simply minimizing the energy function results in completely consistent while meaningless features (i.e., $\boldsymbol{v}_1 \equiv \boldsymbol{v}_2$). Then, we introduce a generation network, as shown in Fig. 2, and propose a adversarial learning method to meanwhile capture the distribution of bitemporal images. After optimization, the network is able to extract global features as well as represent global difference between the two input images. However, the aim of change detection is to generate pixelwise difference that indicates the difference of exact regions within the input images. Then, we propose a difference mapping method to generate the pixelwise difference via the trained network. To better describe the modeling, learning, and change detection processes, the notations used below are summarized in Table I according to the categories of them. Among them, θ and additionally introduced $\boldsymbol{u}$ are the parameters that should be learned. Then, we introduce the modeling process.

B. Energy Function

To mitigate nonintrinsic changes, the primal objective is to align the global features of bitemporal images. However, there are changed regions that should be avoided. Therefore, we

TABLE I
SUMMARY OF NOTATIONS

Categories	Notations	Descriptions
Images or matrixes	I_1, I_2	Input images
	I_1',I_2'	Generated images via the generation network
	δ_I	Gradient of the output of backbone with respect to I
	ξ	Matrix of random noise
	D	difference image
Functions	$f_{\theta/\phi}$	Backbone or generation network
	$L(\cdot)$	Feature alignment loss for energy function
	$C(\cdot)$	Margin-aware constraint for energy function
	$E(\cdot)$	Energy function
	$L_C(\cdot)$, $P_C(\cdot)$	Global difference for change detection
Vectors	$\boldsymbol{v}_1$, $\boldsymbol{v}_2$	Global features of input images
	$\boldsymbol{u}$	Unchanged probability between the features $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$
Others	θ	Parameter set of backbone including weights and biases
	ϕ	Parameter set of generation network
	i, j, i', j', k, l	Indexes
	$\Omega_{(i,j)}$	Pixel set within the neighborhood of pixel (i,j)
	$\omega_{(i,j)}$	Weight matrix within the neighborhood of (i,j)

introduce a new parameter $\boldsymbol{u}$ and define the alignment loss of the two images as follows:

$$L(I_1, I_2, \theta, \boldsymbol{u}) = \sum_i \boldsymbol{u}^i \left(\boldsymbol{v}_1^i \log \frac{\boldsymbol{v}_1^i}{\boldsymbol{v}_2^i} + \boldsymbol{v}_2^i \log \frac{\boldsymbol{v}_2^i}{\boldsymbol{v}_1^i} \right) \quad (2)$$

where i is the component index of the feature vectors. $\boldsymbol{u}$ is a pseudoprobability vector that indicates the unchanged probability of each component that is used to mask corresponding changed components. It is a trainable parameter and will be trained along with the network parameter set θ. In practice, we use the sigmoid activation function in the output layer of the backbone. Then, each component of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ can be taken as the probability that the input images contain certain objects, regions, or features. So we use the symmetric KL divergence to represent the feature alignment loss. Note that the energy function is not directly minimized but in an adversarial manner. $\boldsymbol{u}$ can be tuned to mask changed components during the learning process. The feasibility of the energy function will be analyzed later.

However, the global features extracted from the images lose many details such as margins and textures. So, simply aligning the global features is difficult to maintain the accurate margins of changed regions in the difference image. Therefore, we propose a margin-aware constraint given an image I as follows:

$$C(I, \theta) = \sum_{(i,j,k)} \sum_{(i',j') \in \Omega_{(i,j)}} \omega_{(i,j)}(i', j') [\delta_I(i', j', k) - \delta_I(i, j, k)]^2 \quad (3)$$

where (i, j, k) denotes the pixel of (i, j) in the kth channel. $\Omega_{(i,j)}$ denotes the square neighborhood of the pixel (i, j). δ_I denotes the differential coefficient of the output feature vector $\boldsymbol{v}$ with respect to input image I as follows:

$$\delta_I = \frac{\partial \boldsymbol{v}}{\partial I} = \frac{\partial f_\theta(I)}{\partial I}. \quad (4)$$

Similar to updating the network parameters, δ_I can be computed via back-propagation where the gradient of each component in output layer is 1. ω denotes the weight matrix within neighborhood of each pixel, which is computed by the difference between

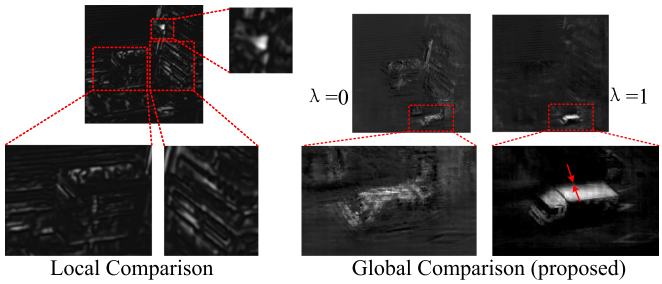


Fig. 3. Difference images obtained from I_1 and I_2 in Fig. 2 by the methods of local comparison [33] and global comparisons with and without the margin-aware constraint.

neighborhood pixels and center pixel as follows:

$$\omega_{(i,j)}(i', j') = \exp \left[\frac{-\|I(i, j) - I(i', j')\|_2^2}{2\sigma^2 \sqrt{(i - i')^2 + (j - j')^2}} \right] \quad (5)$$

where σ denotes the standard deviation of the pixels within the neighborhood. $\| \cdot \|_2^2$ is the square error between two pixels. The weight matrix defines the similarity between pixels within local regions. Larger weight value denotes that the corresponding pixel may belong to the same region with the center pixel. For the network, if two pixels belong to the same region, then the differential of the corresponding output code with respect to the two pixels should also be similar. Otherwise the differential between two pixels is different. Thus, the feature vector is able to well represent the margin between regions within the input images. If two feature vectors are different, the difference is respected to be generated by the difference of regions in input images instead of independent pixels.

Finally, the energy function is obtained by combining the two terms as follows:

$$E(I_1, I_2, \theta, \boldsymbol{u}) = L(I_1, I_2, \theta, \boldsymbol{u}) + \lambda [C(I_1, \theta) + C(I_2, \theta)] \quad (6)$$

where λ is a user defined parameter that controls the weight of the two terms. To intuitively illustrate the effect of global comparison and the margin-aware constraint, difference images generated from I_1 and I_2 in Figs. 1 and 2 with and without the constraint are compared in Fig. 3. Local comparison results in many false alarms due to the impact of viewing angle variance, which can be mitigated via the global comparison. Without the constraint, the margin between different regions are ambiguous. As analyzed above, many details are lost and thus the pixelwise difference reflected from the global difference cannot well maintain the margin between regions. With the constraint, the result shows clear margins and the details of the objects are well maintained. The pixels within changed regions are not independent and they are highlighted evenly. Note that the difference images are generated via trained network and the change detection process will be described later.

The energy function only defines the global difference and margin-aware constraint. With the trainable backbone network, only optimizing the energy function will result in identical global features. The global features should also can well represent the

input images. As a consequence, we design an energy driven adversarial learning method. In the following section, we describe the adversarial learning process.

C. Adversarial Learning

Directly minimizing the above energy function $E(I_1, I_2, \theta, \boldsymbol{u})$ may result in meaningless network parameters, i.e., $\theta = \mathbf{0}$ and $\boldsymbol{u} = \mathbf{0}$. In such a case, the minimal value 0 of the energy can be definitely achieved. So except to learning aligned features, we also expect the features can well represent the input images, i.e., capture the distribution. Inspired from the GAN that was proposed with the aim of capturing the distribution of training data. To implement adversarial learning, an additional generation network is introduced to generate two images I_1' and I_2' (similar to fake image in GAN) from a random noise: $[I_1', I_2'] = f_\phi(\xi)$ with ϕ as the parameter set of the generation network. Then, we define a new min-max loss function based on the energy as follows:

$$\begin{aligned} \min_{\theta, \boldsymbol{u}} \quad & \max_{\phi} \left[E(I_1, I_2, \theta, \boldsymbol{u}) - \mathbb{E}_\xi E(f_\phi(\xi), \theta, \boldsymbol{u}) \right] \\ \text{s.t.} \quad & 1 \geq \boldsymbol{u}_i \geq 0,\ i = 1, 2, \ldots \end{aligned} \tag{7}$$

The backbone and generation network are trained alternately by optimizing the loss function. The aim of the generator is to generate I_1' and I_2' that can minimize the energy function, i.e., the second term of the loss function in (7). While the backbone is used to minimize the energy of input images I_1 and I_2 while prevent other possible data to minimize the energy. So that the difference features learned are exclusive, i.e., they can well represent the input images.

To update the backbone network, the gradient of the learnable parameters are derived as follows:

$$\Delta\{\theta, \boldsymbol{u}\} = \frac{\partial[-E(I_1, I_2, \theta, \boldsymbol{u})]}{\partial \theta, \boldsymbol{u}} - \frac{\partial[-E(I_1', I_2', \theta, \boldsymbol{u})]}{\partial \theta, \boldsymbol{u}}. \tag{8}$$

There are two terms in the energy function, we derive the gradient of them respectively. The gradient of alignment loss is easy to be derived via back-propagation algorithm as follows:

$$\begin{aligned} \frac{\partial L(I_1, I_2, \theta, \boldsymbol{u})}{\partial \theta, \boldsymbol{u}} = & \frac{\partial L(I_1, I_2, \theta, \boldsymbol{u})}{\partial \boldsymbol{v}_1} \frac{\partial \boldsymbol{v}_1}{\partial \theta, \boldsymbol{u}} \\ & + \frac{\partial L(I_1, I_2, \theta, \boldsymbol{u})}{\partial \boldsymbol{v}_2} \frac{\partial \boldsymbol{v}_2}{\partial \theta, \boldsymbol{u}}. \end{aligned} \tag{9}$$

For the margin-aware constraint, the gradient can be derived similarly as follows:

$$\begin{aligned} \frac{\partial C(I, \theta)}{\partial \theta} &= \frac{\partial C(I, \theta)}{\partial \delta_I} \frac{\partial \delta_I}{\partial \theta} \\ &= \sum_{(i,j,k)} \sum_{(i',j') \in \Omega_{(i,j)}} \omega_{(i,j)}(i', j') [\delta_I(i, j, k) \\ &\quad - \delta_I(i', j', k)] \frac{\partial \delta_I}{\partial \theta} - \eta \delta I(i, j, k) \frac{\partial \delta_I}{\partial \theta} \end{aligned} \tag{10}$$

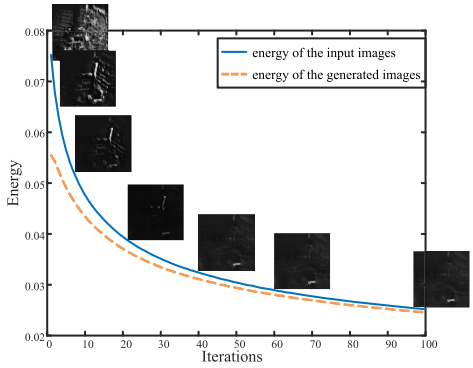


Fig. 4. Illustration of energies and intermediate results during the learning process.

where the difficulty lies in deriving the second derivative as follows:

$$\frac{\partial \delta_I}{\partial \theta} = \frac{\partial^2 \boldsymbol{v}}{\partial I \partial \theta} = \frac{\partial^2 \boldsymbol{v}}{\partial \theta \partial I} = \sum_k \frac{\partial \boldsymbol{v}}{\partial l_k} \frac{\partial^2 l_k}{\partial \theta \partial I} \tag{11}$$

where l_k denotes the kth layer of the network. This means that to compute the gradient, K back-propagation processes are necessary for a network with K layers. With the above gradients, the backbone is then updated according to (8). For the generation network, only the second term in (7) should be considered and the gradient can also be derived according to back-propagation as follows:

$$\Delta\phi = \frac{\partial E(f_\phi(\xi), \theta, \boldsymbol{u})}{\partial f_\phi(\xi)} \frac{\partial f_\phi(\xi)}{\partial \phi}. \tag{12}$$

In the following section, we analyze why the adversarial learning method can optimize the energy and meanwhile capture the changed regions.

D. Feasibility Analysis

To better illustrate the optimization process, we plot the value of the energies of input data $E(I_1, I_2, \theta, \boldsymbol{u})$ and generated data $E(I_1', I_2', \theta, \boldsymbol{u})$ during the learning process in Fig. 4. Both energies decrease with the increase of iterations. The two energies become closer during the learning process, which demonstrates the effect of the learning process. The intermediate difference images generated from I_1 and I_2 in Fig. 2 with different iterations are also exhibited in corresponding position. The difference images at early stages highlight many unchanged regions. After 40 iterations, the difference images tend to be stable and the background is well restrained.

In the energy function, the pseudoprobability $\boldsymbol{u}$ plays the critical rule that avoids the backbone network neglecting changed regions. With the help of min-max adversarial learning, $\boldsymbol{u}$ can be trained to mask components of changed regions. If $\boldsymbol{u}$ masks components that cannot represent the changed regions, then the KL divergence between unmasked components cannot be minimized, which is not the optimal case of (7). While if $\boldsymbol{u}$ masks most unchanged components or changed regions are neglected by the backbone network, the energy of I_1' and I_2' generated by the generation network is not much higher than that of I_1 and I_2

(less unmasked components result in equivalent energy value), which is also not the optimal case of (7). So with the min-max loss function, $\boldsymbol{u}$ can be trained to reduce the influence of changed regions for feature alignment.

However, the aim of change detection is to generate the pixelwise difference information. As a consequence, with the trained backbone network that captures the global difference and maintains the margin detail, a change region mapping method is designed to generate the pixelwise difference.

E. Detecting Changed Regions

After optimization, the feature vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ have been aligned and is able to well represent the global features of input images. As analyzed above, $\boldsymbol{u}$ is able to mask the changed components in the global features. Then, without $\boldsymbol{u}$, the difference between global features can well represent the changes as follows:

$$L_C(I_1, I_2) = \sum_i \left(\boldsymbol{v}_1^i \log \frac{\boldsymbol{v}_1^i}{\boldsymbol{v}_2^i} + \boldsymbol{v}_2^i \log \frac{\boldsymbol{v}_2^i}{\boldsymbol{v}_1^i} \right). \tag{13}$$

So we map the difference through the backbone network via gradients as follows:

$$M_1 = \frac{\partial L_C(I_1, I_2)}{\partial I_1},\ M_2 = \frac{\partial L_C(I_1, I_2)}{\partial I_2}. \tag{14}$$

Finally, we combine the two gradient maps and generate the difference image via moralization as follows:

$$D = \frac{M_1 + M_2 - \max\{M_1(i,j) + M_2(i,j)\}}{\max\{M_1(i,j) + M_2(i,j)\} - \min\{M_1(i,j) + M_2(i,j)\}}. \tag{15}$$

Then, a simple image segmentation method can be used to classify the pixels into changed and unchanged ones. Here, we use an image clustering algorithm called FLICM [45] to generate the final change map. As derived in (14), the difference image is generated via the back-propagated error from the global feature difference. Then the effect of the margin-aware constraint is explicit. Without the constraint, the back-propagated gradient of pixels within the same region cannot be kept coincident, which results in the ambiguous margins in Fig. 3. With the constraint, the details of changed region can be well maintained. The whole change detection process including learning and extracting changes is summarized in Algorithm 1, given two images I_1 and I_2, is summarized in Algorithm 1.

Even though derivatives seem complex, the learning process is concise, which is composed of two main steps including learning and detecting changes. In learning, there are two main steps in each iteration. The whole process seems time-consuming, but since the training only needs the two input multitemporal images, the model converges rapidly. Moreover, the parallel computation on GPU significantly increases the computational efficiency.

Algorithm 1: Workflow of the Proposed Change Detection Method.

Input:
Input I_1 and I_2 (only two images).
Initialization:
Randomly initialize θ and $\boldsymbol{u}$.
Learning:
while The energy $E(I_1, I_2, \theta, \boldsymbol{u})$ is not stable and the number of iterations is under maximum value **do**
Updating generation network:
Generate noise ξ and compute the energy.
Compute the gradients $\Delta\phi$ via (12) and update ϕ.
Generate images I_1' and I_2'.
Updating backbone network:
Compute the energies in (7).
Compute the gradients $\Delta\{\theta, \boldsymbol{u}\}$ as in (8) and update.
end while
Detecting Changes:
Compute global difference via (13).
Map global difference into input images via (14).
Compute difference image D via (15)
Output:
Output the difference image D and segment it by FLICM.

IV. Experimental Study

In this section, we verify effectiveness of the proposed change detection framework via different types of images. All of them are misaligned. We first exhibit the datasets and corresponding experimental settings. Then, the results are compared with both unsupervised and supervised change detection methods.

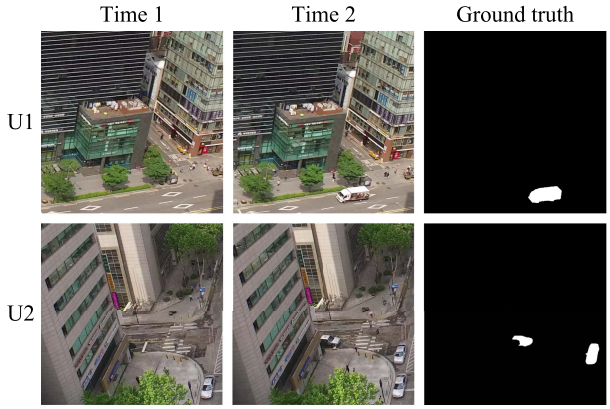


Fig. 5. Datasets generated from UAV.

A. Datasets and Experimental Settings

We use eight datasets with each dataset contains two multitemporal images and a ground truth to test the proposed method. Note that the ground truth is only used for result evaluation. The learning and change detection processes use only the two multitemporal images for each dataset. To fully evaluate the proposed method, images captured from different acquisition platforms and scenarios are used.

The first two datasets show the scene of urban street captured by unmanned aerial vehicle (UAV) which are exhibited in Fig.5 (named as Datasets U1 and U2, respectively). The size of them is 512×512 pixels. Since UAV moves persistently,

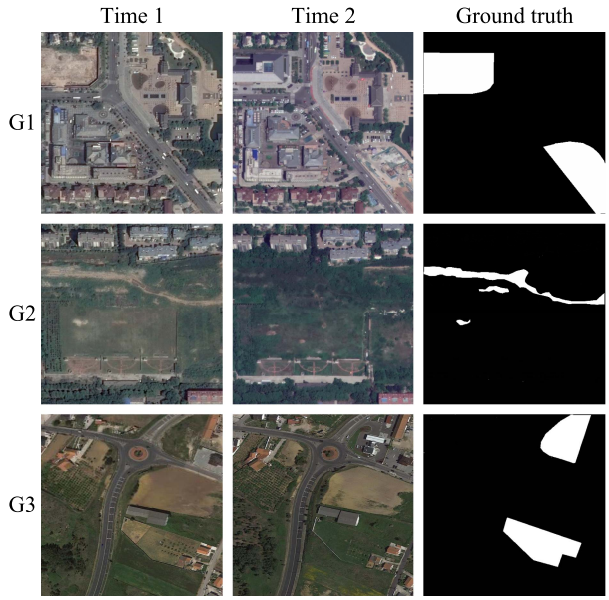


Fig. 6. Datasets generated from the Gaofen Challenge.

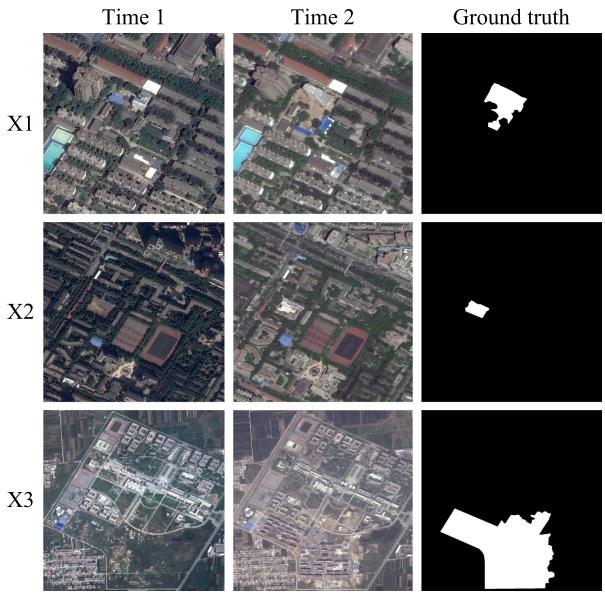


Fig. 7. Datasets generated from Google Map on the area of Xidian University.

even though it captures the same scene, the view position and angle are different, which results in the difficulty of accurate coregistration. Since the images are captured within a 3-D scene, many details such as lines, edges, and textures are not strictly aligned regardless of the seeming coregistration [46], as shown in Fig. 1, where the regions at the same location show distinct appearance.

The following three datasets are from the 2021 Gaofen Challenge on Automated High-Resolution Earth Observation Image Interpretation,[2] which are exhibited in Fig. 6 (named as Datasets G1, G2, and G3, respectively). The images are Gaofen-2 or Jilin-1 multitemporal optical images with resolution better than 1 m. The size of them is 512×512 pixels. Due to the illumination and seasonal variances, there are color deviations between bitemporal images.

The last three datasets show the changes of land use within Xidian University. They are exhibited in Fig. 7 (named as Datasets X1, X2, and X3, respectively). The size of them is 1024×1024 pixels. We capture the images from the Google Map according to the coordinate of latitude and longitude. So the images are strictly accordant from the geographic position. But since the images are captured from different times and view angles, the illumination and season are different. Moreover, many buildings are with different morphology, as shown in Fig. 8. The three datasets show large scenes from the Google Map without any prior information in order to demonstrate the learning capability of the proposed method in capturing differences for open large scale unsupervised scenarios.

To demonstrate the adaptation of the proposed method, we compare with both unsupervised and supervised change detection methods. Unsupervised methods are able to generate the results with only two multitemporal images and the compared methods include deep slow feature analysis [37], stacked

[2] [Online]. Available: http://sw.chreos.org/challenge

Fig. 8. Illustration of different view angles on the Xidian University datasets.

denoising autoencoder (SDAE) based method [47], the deep convolutional coupling network (DCCN) [33], our previous work BCNN [10] and TVRBN [34], and recently proposed unsupervised methods INLPG [39] and LPEM [40]. Moreover, since the compared methods generate difference images, we use the same image segmentation method FLICM [45] to generate the final result. In practice, since the prior information about the scenarios are not always available, for fair comparison, we use a public dataset, i.e., the widely used LEVIR-CD [23] dataset to train supervised methods including BiT [26], P2V [48], and SNUNet [49]. Then, the trained models are used to implement the change detection on the above test datasets.

We evaluate the difference images via two criteria, including areas under precision-recall (PR) curve [50] and receiver operating characteristic (ROC) curve [51], are called AP and AUR, respectively. These criteria are widely used in evaluating the separability of data. The curves are plotted based on a set of results with consecutive thresholds (0,1,2,...,255). Those results are compared with the ground truth to generate the values of true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The precision, true positive rate (TPR, a.k.a. recall rate), and false positive rate (FPR) are then computed as

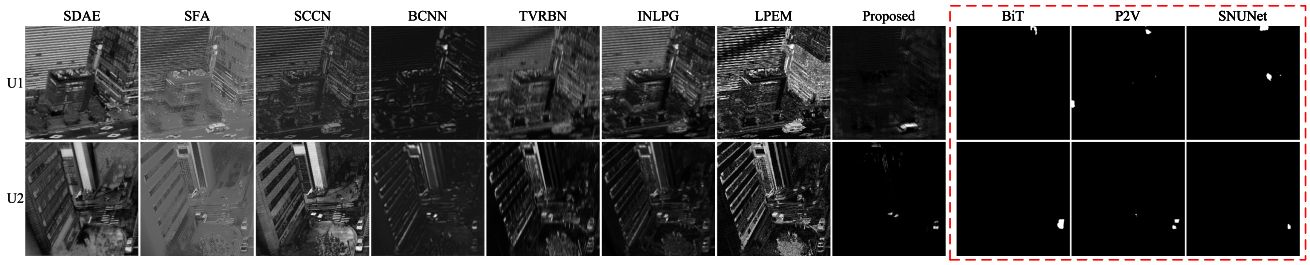


Fig. 9. Difference images or final results generated via different methods on UAV datasets. Final results of supervised methods are surrounded by red dashed box.

follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$
$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$
$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}. \tag{16}$$

Finally the PR curve is plotted by using the precision versus the recall rate and the ROC curve is plotted by using the TPR versus the FPR. Those criteria are used to evaluate the quality of the difference images. If a difference image is closer to the ground truth than others, its criteria are better. Then, accuracy (Acc.) and Kappa coefficient are use to evaluate the final result generated by FLICM. The two criteria are widely used in change detection [52].

To implement the proposed method, we use a CNN as the backbone network and set 16 layers for the backbone network including 1 input layer, 7 convolutional layers with each one followed by a 2×2 pooling layer, and 1 fully connected layer as the output layer. The number of feature maps is assigned as 3 (input layer), 16×2 (one convolutional layer and one pooling layer), $32 \times 2, 64 \times 4, 128 \times 6$, and 512 (output feature vector). For the generation network, we set 8 convolutional layers and 7 upsampling layers alternately. We set $\lambda = 1$, $\Omega_{(i,j)} = 5 \times 5$, and the learning rate as 0.1. We use softplus [53] as the activation function for hidden layers, which is the continuous version of the widely used ReLU in order to compute the second-order derivatives conveniently and meanwhile mitigate the vanished gradients. We use the sigmoid function as the activation function for the output layer in order to encode the images.

B. Experiments on UAV Datasets

The difference images generated by unsupervised methods on the two UAV datasets are exhibited in Fig. 9. Supervised methods directly output the final results, which are also illustrated in Fig. 9. From the datasets, it seems that they are coregistered. But since UAV moves frequently, the view angles are different. The different view angles can be reflected from the results of compared methods where the position bias of the same object results in the highlight of unexpected objects such as the textured wall. On the dataset U1, the car is prominent against the background. Therefore, the compared methods can highlight it. However, most of the existing methods generate the pixelwise difference via local or regional comparison. As analyzed above, the view angle variance results in the distinct features at the same position. As a consequence, compared methods cannot well restrain the unchanged region, which generate many false detections. While for the proposed method, global comparison successfully avoids the influence of view angle variance and generate clear unchanged region. For supervised methods, if there are few changed cars in the training dataset, the model cannot be trained to well detect the cars. As a consequence, they wrongly highlight the change objects. On the dataset U2, there are two changed cars and one unchanged car. SDAE highlights the unchanged car while ignores the changed cars. SFA can highlight the changed cars and restrain the unchanged car. But the building is also highlighted. DCCN highlights all the three cars. With the adversarial learning, BCNN highlights the changed objects but as well as some margins and textures due to the comparison of local features. Different from them, the proposed change detection compares the images according to the global features. Therefore, the changed cars can be highlighted and unchanged objects can be restrained. However, without any prior information, the car windows are recognized as independent objects and unchanged ones. For supervised methods, only one car can be detected.

The values of the evaluation criteria are listed in Table II. The proposed method significantly outperforms the compared methods. Unsupervised methods are not able to restrain the background objects while supervised methods heavily depend on training samples. Even though the proposed method can detect the changed objects, the completeness cannot be guaranteed.

C. Experiments on Gaofen Challenge Datasets

The difference images or final results generated by different methods on the three Gaofen Challenge datasets are exhibited in Fig. 10. There are illumination difference and slight misalignment in the three datasets. Therefore, the difference images of the existing methods show many false alarms. Even though the changed regions can be highlighted, the background is not well restrained. The proposed method is able to recognize the changed regions, but part of changed regions are also missed and the complete changed region cannot be accurately detected. Especially in the G3 dataset where only a building is detected and actually the whole region around the building is changed. But the proposed method is able to reduce many false alarms caused by the misalignment, which demonstrate the superiority

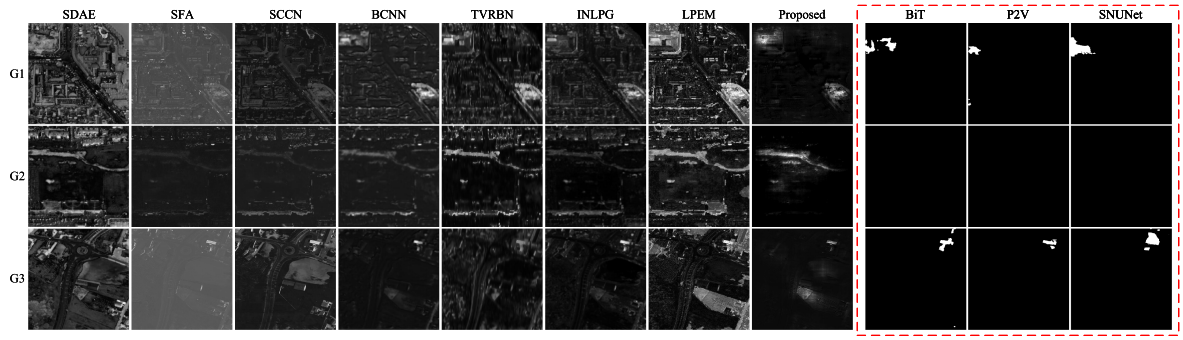


Fig. 10. Difference images or final results generated via different methods on Gaofen Challenge datasets. Final results of supervised methods are surrounded by red dashed box.

TABLE II
VALUES OF EVALUATION CRITERIA ON THE UAV DATASETS.

datasets	methods	AP	AUR	Acc.(%)	Kappa
U1	SDAE	0.1166	0.4811	79.50	0.0018
	SFA	0.1740	0.7523	98.26	0.3059
	DCCN	0.0794	0.7982	97.94	0.1818
	BCNN	0.0822	0.7846	97.35	0.1688
	TVRBN	0.0828	0.8896	96.01	0.1695
	INLPG	0.0222	0.6773	56.86	0.0220
	LPEM	0.0508	0.8120	92.60	0.0974
	BiT	-	-	98.23	0.0035
	P2V	-	-	98.11	0.0053
	SNUNet	-	-	97.92	0.0077
	proposed	**0.6039**	**0.9143**	**99.21**	**0.6055**
U2	SDAE	0.0075	0.3745	9.02	0.0015
	SFA	0.1082	0.7275	98.23	0.1947
	DCCN	0.0745	0.6880	98.39	0.1630
	BCNN	0.1020	0.7129	98.41	0.1795
	TVRBN	0.0779	**0.9070**	96.98	0.1516
	INLPG	0.0879	0.8640	96.46	0.1538
	LPEM	0.0804	0.8358	98.01	0.1468
	BiT	-	-	99.13	0.4703
	P2V	-	-	98.97	0.2971
	SNUNet	-	-	98.85	0.1451
	proposed	**0.4073**	0.7318	**99.20**	**0.4900**

The bold entities denote the best values among the compared methods.

TABLE III
VALUES OF EVALUATION CRITERIA ON THE GAOFEN CHALLENGE DATASETS.

datasets	methods	AP	AUR	Acc.(%)	Kappa
G1	SDAE	0.2284	0.6166	70.98	0.1575
	SFA	0.3322	0.7444	78.25	0.2768
	DCCN	0.3088	0.6752	78.44	0.2403
	BCNN	0.4849	0.7107	85.34	0.3857
	TVRBN	0.5097	0.7671	85.96	0.3694
	INLPG	0.4653	0.7701	83.79	0.3580
	LPEM	0.4674	**0.8103**	83.85	0.4332
	BiT	-	-	85.67	0.1559
	P2V	-	-	84.68	0.0637
	SNUNet	-	-	86.10	0.1959
	proposed	**0.6010**	0.7903	**88.02**	**0.4991**
G2	SDAE	0.1517	0.8582	89.06	0.2420
	SFA	0.2643	0.9396	92.84	0.4065
	DCCN	0.1367	0.8701	85.05	0.2569
	BCNN	0.3310	0.9520	93.61	0.4413
	TVRBN	0.2991	0.9222	95.33	0.4806
	INLPG	0.1428	0.8718	86.21	0.2348
	LPEM	0.4209	**0.9667**	95.36	0.5289
	BiT	-	-	95.97	0.0
	P2V	-	-	95.97	0.0
	SNUNet	-	-	95.97	0.0
	proposed	**0.7433**	0.9456	**97.82**	**0.7012**
G3	SDAE	0.0829	0.4003	10.69	0.0001
	SFA	0.1774	0.6620	79.33	0.1019
	DCCN	0.1143	0.4721	88.08	0.0618
	BCNN	0.3261	0.7500	83.44	0.2873
	TVRBN	0.2298	0.6189	82.66	0.1713
	INLPG	0.3069	0.8050	81.15	0.2606
	LPEM	0.3477	**0.8229**	85.90	0.4135
	BiT	-	-	90.42	0.1425
	P2V	-	-	90.12	0.0928
	SNUNet	-	-	**90.92**	0.2135
	proposed	**0.4716**	0.7156	90.39	**0.3954**

The bold entities denote the best values among the compared methods.

of comparing global features. Supervised methods highly depend on the training data. Since the LEVIR-CD dataset mainly focuses on the changes of buildings, the changes of vegetation are missed.

The values of the evaluation criteria are listed in Table III. From those evaluation criteria, the proposed method is superior over existing ones. But on the datasets G2 and G3, the proposed method cannot outperform BCNN in terms of AUR. For the proposed method, even though part of the changed regions can be highlighted, part of them is restrained. Therefore, when FPR is high, TPR becomes lower. Even though supervised methods may miss some changed regions, the detected changed regions are more complete, which results in larger Acc on the G3 dataset. However, in terms of other criteria, the proposed method significantly outperforms compared methods.

D. Experiments on Xidian University Datasets

Finally, we test our method in an real scenario. Suppose, we want to research the changes of Xidian University while there are no satellite images. We can acquire the images from Google Map, which almost covers the surface of Earth. Both latest and historical data are provided. However, with different view angles, the tall objects show different appearance and cannot be accurately coregistered, as shown in Fig. 8. This phenomenon brings great challenges for traditional change detection methods, especially unsupervised ones.

The difference images or final results generated by different methods are shown in Fig. 11. The three datasets show

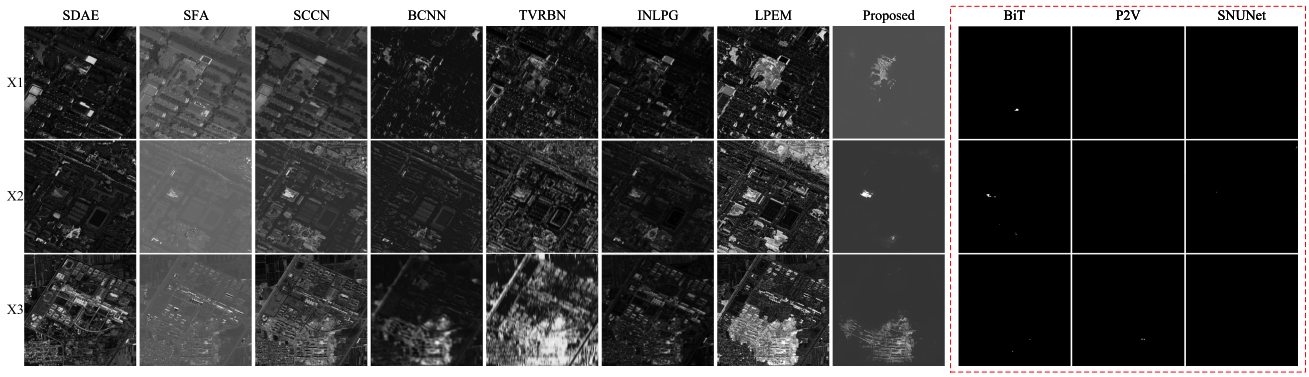


Fig. 11. Difference images or final results generated via different methods on Xidian University datasets. Final results of supervised methods are surrounded by red dashed box.

the changes of buildings. On the dataset X1, there are many details that cannot be well coregistered. The tall building is specially obvious. The compared unsupervised methods highlight unchanged objects. For example, SDAE highlights the white rooftops. The proposed method can well restrain the background objects. But changed object is not completely highlighted. On the dataset X2, the changed object is a white building, which shows distinct appearance with the background. As a consequence, SFA and DCCN can highlight it with well restrained background. However, there is another white building that influences the detection result. The proposed method is able to avoid the influence of unchanged buildings. Dataset X3 shows a changed region with many rising buildings. Many unchanged buildings are not exactly coregistered. Moreover, the different seasons also result in the variant appearance of background. As a consequence, even though some compared methods can highlight the changed region, many unchanged buildings are also detected as changed ones. In summary, as analyzed in Section II, compared methods fail to accurately distinguish intrinsic and nonintrinsic changes, which results in many false alarms in unchanged regions. From the results, the proposed method achieves clearer unchanged regions, which demonstrates its capability in restraining nonintrinsic changes. But due to the simple margin-aware constraint, which is based on the color information, many changed regions are missed. With complex scenarios with many types of ground objects, supervised methods fail to correctly detect the changed regions with limited training data.

The values of the evaluation criteria are listed in Table IV. Obviously, the proposed method significantly outperforms compared ones. Since the images are directly acquired from Google Map, there are many obstacles for unsupervised methods to deal with, such as the illumination, season, and viewing angle variance. Moreover, the scene variance between test image and training images in supervised methods also results in invalidation.

TABLE IV
VALUES OF EVALUATION CRITERIA ON THE XIDIAN UNIVERSITY DATASETS.

datasets	methods	AP	AUR	Acc.(%)	Kappa
X1	SDAE	0.0580	0.6841	82.30	0.0824
	SFA	0.0947	0.8281	84.52	0.1509
	DCCN	0.1398	0.9003	89.67	0.2246
	BCNN	0.1139	0.7699	92.11	0.1694
	TVRBN	0.1277	0.8470	92.18	0.2048
	INLPG	0.1088	0.8515	87.11	0.1634
	LPEM	0.1985	0.9323	90.25	0.3040
	BiT	-	-	96.76	0.0009
	P2V	-	-	96.80	0.0003
	SNUNet	-	-	96.80	0.0
	proposed	**0.6540**	**0.9664**	**98.13**	**0.6722**
X2	SDAE	0.0287	0.8131	92.66	0.0734
	SFA	0.0944	0.8719	98.41	0.2531
	DCCN	0.2594	0.8911	99.22	0.4470
	BCNN	0.0081	0.5234	30.79	0.0017
	TVRBN	0.0618	0.8430	98.07	0.1432
	INLPG	0.0585	0.8461	96.88	0.1620
	LPEM	0.0536	0.8684	98.30	0.1062
	BiT	-	-	99.16	0.1158
	P2V	-	-	99.11	0.0
	SNUNet	-	-	99.10	0.0049
	proposed	**0.5887**	**0.9768**	**99.45**	**0.5770**
X3	SDAE	0.1486	0.4648	34.07	0.0059
	SFA	0.3346	0.7553	76.57	0.3109
	DCCN	0.4432	0.8046	81.86	0.4183
	BCNN	0.8282	0.9446	91.12	0.6934
	TVRBN	0.6494	0.8979	87.20	0.5540
	INLPG	0.3239	0.7655	72.35	0.2911
	LPEM	0.6605	0.9335	88.65	0.6430
	BiT	-	-	82.61	0.0006
	P2V	-	-	82.62	0.0014
	SNUNet	-	-	82.61	0.0
	proposed	**0.8210**	**0.9464**	**92.20**	**0.7322**

The bold entities denote the best values among the compared methods.

E. Experiment on λ

In this method, the hyperparameter λ is important to control the importance of the two terms in the energy function. Therefore, in this experiment, we set the λ as $\{0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100, 200\}$ and implement the proposed method to generate the difference images, respectively. The AP and AUR values of those difference images on the eight datasets are shown in Fig. 12. Due to the consecutive pooling layers, the output feature is robust to the local positional deviation. Slight positional deviation of regions will not influence the difference of features. Then, the alignment

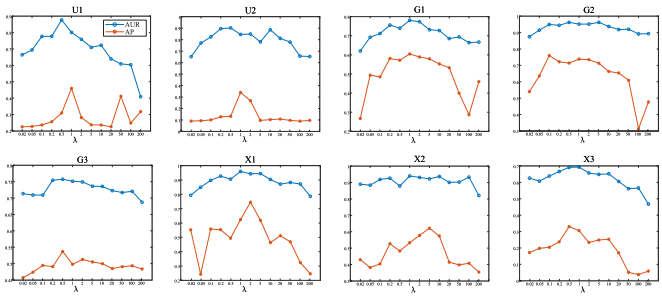


Fig. 12. AUR and AP values with different λ on the eight datasets.

TABLE V
TIME COST (S) OF EACH DATASET

Dataset	U1	U2	G1	G2	G3	X1	X2	X3
INLPG	95.0	96.9	99.9	97.9	96.1	3022.5	2899.9	2869.5
LPEM	16.7	14.2	14.4	14.7	14.8	62.4	66.0	64.4
proposed	70.6	89.6	98.8	40.9	49.0	190.5	188.9	131.5

loss $L(I_1, I_2, \theta, \boldsymbol{u})$ and constraint $C(I_1, \theta) + C(I_2, \theta)$ is not conflicted with each other. As a consequence, from the line plots, the proposed method is not very sensitive to the value of λ on some datasets. While for the U1 and U2 datasets, much better results are achieved by the λ values around 1. Different from remote sensing scenes where the misalignment is local (only buildings), the misalignment of UAV scenes is more severe because the camera captures the images around the 3-D objects instead of far from them. Therefore, almost all the objects are not well aligned. As a consequence, the choice of λ influences the result a lot on the UAV datasets. For Gaofen Challenge datasets, the positional deviation is simple to deal with and the view angle seems almost the same. Therefore, the method is not sensitive to λ on those datasets.

F. Running Time

As we analyzed in Section III, even though the training and change detection processes seem overstaffed, both of them focus on one pair of images. With the assistance of GPU parallel computation, the computational time can be controlled within a reasonable range. Therefore, here we list the running time of the proposed method on each dataset in Table V. The computing device is equipped with Intel i7-8700 K CPU (3.7 GHz) and NVIDA RTX3090 GPU. From the table, the computational cost for all the datasets can be kept within 5 min, which is acceptable for offline change detection. For small datasets (512×512 pixels), the whole change detection process can be kept within 2 min. For large datasets, the computational time increases significantly. But the computational time does not increase exponentially. Since real time is not required in change detection, the proposed method is feasible in practical applications with parallel computation devices.

V. CONCLUSION

We propose a new unsupervised change detection paradigm, which compares images via globally encoded features. Thus, it can mitigate the influence of many global changes such as the illumination, season, and viewing angles. To generate changed regions from global comparison, we use a deep neural network to extract the features and propose an adversarial learning method to train the network. First, to align the global features, we define a masked loss, which measures the bias between features of two images. Meanwhile, to maintain the details for pixelwise difference, we define a margin-aware constraint that restrains the differential of code with respect to input image according to the margin of regions. Second, to learn useful features with the two input images and meanwhile align the global features, we introduce a generation network and train the two networks adversatively. After training, the change detection is achieved by mapping the global difference to pixelwise difference via gradient. From the experiments, the proposed methods significantly reduces the nonintrinsic changes and achieves clear unchanged regions in the difference image, which is difficult for compared methods.

However, due to that the margin-aware constraint is simply based on gray value and spatial information of input images, the changed region cannot be completely detected. Moreover, the second derivative also encumbers the learning efficiency. Therefore, in the future work, we try to design an object-aware network that encodes the images according to complete objects, such as a car or a building instead of simple homogeneous regions. Moreover, we will consider fusing multiscale feature difference between feature maps of different layers to maintain more changed regions and directly constraint a mask to avoid the second derivative for efficiency.

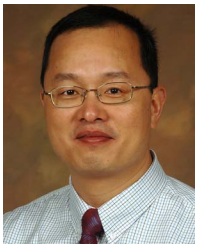
REFERENCES

[1] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.
[2] W. Wang, C. Liu, G. Liu, and X. Wang, "CF-GCN: Graph convolutional network for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5607013.
[3] B. Cui et al., "Enhanced edge information and prototype constrained clustering for SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5206116.
[4] J. Ding, X. Li, S. Xiang, and S. Chen, "Multilevel features fused and change information enhanced neural network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5502413.
[5] L. Ding, J. Zhang, H. Guo, K. Zhang, B. Liu, and L. Bruzzone, "Joint spatio-temporal modeling for semantic change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5610814.
[6] L. Ding, K. Zhu, D. Peng, H. Tang, K. Yang, and L. Bruzzone, "Adapting segment anything model for change detection in VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5611711.
[7] L. Fang, Y. Jiang, H. Yu, Y. Zhang, and J. Yue, "Point label meets remote sensing change detection: A consistency-aligned regional growth network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5603911.
[8] K. Jiang, J. Liu, W. Zhang, F. Liu, and L. Xiao, "MANet: An efficient multidimensional attention-aggregated network for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4706118.
[9] K. Jiang, W. Zhang, J. Liu, F. Liu, and L. Xiao, "Joint variation learning of fusion and difference features for change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4709918.
[10] J. Liu, W. Zhang, F. Liu, and L. Xiao, "A probabilistic model based on bipartite convolutional neural network for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4701514.

[11] W. Zhang, L. Jiao, F. Liu, S. Yang, W. Song, and J. Liu, "Sparse feature clustering network for unsupervised SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5226713.
[12] W. Zhang, L. Jiao, F. Liu, S. Yang, and J. Liu, "Adaptive contourlet fusion clustering for SAR image change detection," *IEEE Trans. Image Process.*, vol. 31, pp. 2295–2308, 2022.
[13] S. Bianco, G. Ciocca, and R. Schettini, "Combination of video change detection algorithms by genetic programming," *IEEE Trans. Evol. Comput.*, vol. 21, no. 6, pp. 914–928, Dec. 2017.
[14] K. Korotkov, J. Quintana, S. Puig, J. Malvehy, and R. Garcia, "A new total body scanning system for automatic change detection in multiple pigmented skin lesions," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 317–338, Jan. 2015.
[15] R. Saxena et al., "Towards a polyalgorithm for land use change detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 144, pp. 217–234, 2018.
[16] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and bayesian soft fusion," *Remote Sens. Environ.*, vol. 199, pp. 241–255, 2017.
[17] S. Jin, L. Yang, Z. Zhu, and C. Homer, "A land cover change detection and classification protocol for updating Alaska NLCD 2001 to 2011," *Remote Sens. Environ.*, vol. 195, pp. 44–55, 2017.
[18] X. Zhang, C. Gilliam, and T. Blu, "All-pass parametric image registration," *IEEE Trans. Image Process.*, vol. 29, pp. 5625–5640, 2020.
[19] X. Chen, L. Vierling, and D. Deering, "A simple and effective radiometric correction method to improve landscape change detection across sensors and across time," *Remote Sens. Environ.*, vol. 98, no. 1, pp. 63–79, 2005.
[20] X. Li et al., "Adaptive complex wavelet informed transformer operator," *IEEE Trans. Multimedia*, vol. 27, pp. 3513–3526, 2025.
[21] X. Yi et al., "Contour-aware dynamic low-high frequency integration for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5402213.
[22] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.
[23] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1662.
[24] M. Liu, Q. Shi, A. Marinoni, D. He, X. Liu, and L. Zhang, "Super-resolution-based change detection network with stacked attention module for images with different resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4403718.
[25] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604816.
[26] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5607514.
[27] M. Lebedev, Y. Vizilter, O. Vygolov, V. Knyaz, and A. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *ISPRS - Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. XLII-2, pp. 565–571, 2018.
[28] J. Liu, H. Luo, W. Zhang, F. Liu, and L. Xiao, "Multiscale self-supervised constraints and change-masks-guided network for weakly supervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 4701415.
[29] F. Liu, P. Zhang, J. Liu, J. Yang, X. Tang, and L. Xiao, "Background-driven and foreground-refined network for weakly supervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 4704116.
[30] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
[31] D. Marinelli, F. Bovolo, and L. Bruzzone, "A novel change detection method for multitemporal hyperspectral images based on binary hyperspectral change vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4913–4928, Jul. 2019.
[32] 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.
[33] 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.
[34] L. Hu, J. Liu, and L. Xiao, "A total variation regularized bipartite network for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5239518.
[35] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016.
[36] J. Shi, T. Wu, A. K. Qin, Y. Lei, and G. Jeon, "Self-guided autoencoders for unsupervised change detection in heterogeneous remote sensing images," *IEEE Trans. Artif. Intell.*, vol. 5, no. 6, pp. 2458–2471, Jun. 2024.
[37] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
[38] K. Xiao, Y. Sun, G. Kuang, and L. Lei, "Change alignment-based graph structure learning for unsupervised heterogeneous change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 2504405.
[39] 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.
[40] Y. Sun, L. Lei, D. Guan, G. Kuang, Z. Li, and L. Liu, "Locality preservation for unsupervised multimodal change detection in remote sensing imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 4, pp. 6955–6969, Apr. 2025.
[41] Q. Xu, Y. Shi, J. Guo, C. Ouyang, and X. X. Zhu, "UCDFormer: Unsupervised change detection using a transformer-driven image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5619917.
[42] L. Hu, Q. Liu, J. Liu, and L. Xiao, "PRBCD-Net: Predict-refining-involved bidirectional contrastive difference network for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5620717.
[43] Y. Zhou, X. Li, K. Chen, and S.-Y. Kung, "Progressive learning for unsupervised change detection on aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5601413.
[44] Y. Xing, Q. Zhang, L. Ran, X. Zhang, H. Yin, and Y. Zhang, "Progressive modality-alignment for unsupervised heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5614612.
[45] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.
[46] M. Gong, S. Zhao, L. Jiao, D. Tian, and W. Shuang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, Jul. 2014.
[47] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 24–41, 2016.
[48] M. Lin, G. Yang, and H. Zhang, "Transition is a process: Pair-to-video change detection networks for very high resolution remote sensing images," *IEEE Trans. Image Process.*, vol. 32, pp. 57–71, 2023.
[49] 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.
[50] K. Boyd, K. H. Eng, and C. D. Page, "Area under the precision-recall curve: Point estimates and confidence intervals," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2013, pp. 451–466.
[51] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2005.
[52] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.
[53] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier networks," in *Proc. NIPS 2010 Workshop Deep Learn. Unsupervised Feature Learn.*, 2010, pp. 315–323.

Wenhua Zhang (Member, IEEE) received the B.S. degree in communication engineering from the North University of China, Taiyuan, China, in 2015, and the Ph.D. degree in electronic circuit and system from Xidian University, Xi'an, China, in 2021.

She is currently an Associate Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. Her research interests include object tracking and image processing.



Qihao Weng (Fellow, IEEE) received the A.S. degree in geography from Minjiang University, Fuzhou, China, in 1984, the M.S. degree in geography from South China Normal University, Guangzhou, China, in 1990, the M.A. degree in geography from the University of Arizona, Tucson, AZ, USA, in 1996, and the Ph.D. degree in remote sensing and geographic information system (GIS) from the University of Georgia, Athens, GA, USA, in 1999.

From 2008 to 2009, he visited the NASA Marshall Space Flight Center, Huntsville, AL, USA, as a Senior Research Fellow. He is currently a Professor of Earth and environmental systems and the Director with the Center for Urban and Environmental Change, Indiana State University, Terre Haute, IN, USA. He has authored 241 articles and 14 books, with more than 19900 citations and an H-index of 64. His research interests include remote sensing analysis of urban ecological and environmental systems, land-use and land-cover changes, urbanization impacts, and environmental sustainability.

Dr. Weng was the recipient of distinguished career awards, including the NASA Senior Fellowship, the AAG Distinguished Scholarship Honors Award, the Taylor & Francis Lifetime Achievements Award, and the Japan Society for the Promotion of Science (Short-term S[E]) Fellowship. He is also an Elected Fellow of American Association for the Advancement of Science (AAAS), American Society for Photogrammetry and Remote Sensing (ASPRS), and a Member of International Society for Photogrammetry and Remote Sensing (ISPRS), American Geophysical Union (AGU), and American Association of Geographers (AAG). He has been the Organizer and Program Committee Chair of the biennial IEEE sponsored International Workshop on Earth Observation and Remote Sensing Applications (EORSA) conference series since 2008. He was the National Director of ASPRS from 2007 to 2010. He has been invited to give more than 110 talks by organizations and conferences worldwide. He also serves as the Editor-in-Chief of the *ISPRS Journal of Photogrammetry and Remote Sensing*. He is also a Series Editor of *Remote Sensing Applications Series* (Taylor & Francis) and *The Imaging Science Journal* series (Taylor & Francis).

Jia Liu (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 2013 and 2018, respectively.

He is currently an Associate Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. His current research interests include computational intelligence and image understanding.

Qu Zeng received the Ph.D. degree in instrument science and technology from the North University of China, Taiyuan, China, in 2021.

He is currently an Associate Professor with the School of Electrical and Control Engineering, North University of China. His research interests include electromagnetic metamaterials, filters, electromagnetic stealth technology, and micro/nano device fabrication and system integration.

Yong Li received the PhD degree from the Institute of Computing Technology (ICT), Chinese Academy of Sciences, in 2020. He is currently an Associate Professor with the School of Computer Science and Engineering, Southeast University, Nanjing, China. His research results have been expounded in more than 40 publications at prestigious journals and prominent conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON MULTIMEDIA, *Neural Information Processing Systems*, *Computer Vision and Pattern Recognition*, and *International Conference on Computer Vision*. His research interests include deep learning and human-centered affective computing.

Jinjin Chai received the B.S. degree in electronic science and technology from Tianjin Normal University of China, Tianjin, China, in 2015, and the Ph.D. degree in cryptography from Xidian University, Xi'an, China, in 2021.

She is currently a Lecturer with the School of Air Defence and Antimissile, Air Force Engineering University, Xi'an. Her research interests include object tracking and situation awareness.

Liang Xiao (Senior Member, IEEE) received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively.

From 2006 to 2008, he was a Postdoctoral Research Fellow with the Pattern Recognition Laboratory, NJUST. From 2009 to 2010, he was a Postdoctoral Fellow with the Rensselaer Polytechnic Institute, Troy, NY, USA. Since 2013, he has been the Deputy Director with the Jiangsu Key Laboratory of Spectral Imaging Intelligent Perception, Nanjing. Since 2014, he has been the Vice-Director with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, NJUST, where he is currently a Professor with the School of Computer Science and Engineering. His research interests include remote sensing image processing, image modeling, computer vision, machine learning, and pattern recognition.