Incorporating Metric Learning and Adversarial Network for Seasonal Invariant Change Detection

Wenzhi Zhao⬤, Lichao Mou, *Student Member, IEEE*, Jiage Chen, Yanchen Bo⬤, *Member, IEEE,* and William J. Emery⬤, *Life Fellow, IEEE*

Abstract—Change detection by comparing two bitemporal images is one of the most fundamental challenges for dynamic monitoring of the Earth surface. In this article, we propose a metric learning-based generative adversarial network (GAN) (MeGAN) to automatically explore seasonal invariant features for pseudochange suppressing and real change detection. To achieve this purpose, a seasonal invariant term is introduced to maximally suppress pseudochanges, whereas the MeGAN explores the transition patterns between adjacent images in a self-learning fashion. Different from the previous works on bitemporal imagery change detection, the proposed MeGAN have the following contributions: 1) it automatically explores change patterns from the complex bitemporal background without human intervention and 2) it aims to maximally exclude pseudochanges from the seasonal transition term and map out real changes efficiently. To our best knowledge, this is the first time we incorporate the seasonal transition term and GAN for change detection between bitemporal images. At last, to demonstrate the robustness of the proposed method, we included two data sets which are the Google Earth data and the Landsat data, for bitemporal change detection and evaluation. The experimental results indicated that the proposed method is able to perform change detection with precision can be as high as 81% and 88% for the Google Earth and Landsat data set, respectively.

Index Terms—Bitemporal images, change detection, metric learning, metric learning-based generative adversarial network (GAN) (MeGAN), pseudochanges.

I. INTRODUCTION

DYNAMICALLY monitoring the Earth surface's evolving process is fundamental to understanding human roles that impact environmental transformation. With the rapid development of remote-sensing industry, it is now possible to acquire real-time satellite images to detect environmental changes at fine observation scales, for example, the Landsat and Sentinel mission. Change detection aims to identify changes taking place on the same geographical area by analyzing images that are captured at different times [1]–[3]. It has been widely used in urban planning, disaster management, and land-use monitoring. However, change detection is a challenging task due to the heterogeneous nature of remote-sensing images (which contain random noises and pseudochanges), and it needs effective strategies to detect real changes from complex background.

Recent works have been intensely studying change detection in remote-sensing images, and thus many methods are proposed to better identify changes between image pairs. Among them, the algebra-based approach is one of the most popular change detection strategies, such as the image ratio, image differencing, and change vector analysis (CVA) [4], [5]. However, the algebra-based strategy directly feeds with reflectance values which amplify the random noises introduced by atmospheric conditions. Moreover, the threshold selection is a huge challenge to identify change areas [6], [7]. To facilitate change detection, the unsupervised strategy is proposed by integrating expectation-maximization (EM) algorithm for the threshold selection [8]. However, it is still difficult to select reasonable thresholds to include all changed areas while excluding unwanted ones. Consequently, the transformation-based change detection methods introduce transformed feature representations [such as principal component analysis (PCA) and cap transformation] to highlight changed areas, while suppressing and reducing unchanged information between the bands. Although the transformed strategy exploits invariant components that are able to separate changed pixels from unchanged ones, it is difficult to label the information about the changing area in the transformed feature space. Complementarily, the classification-based change-detection strategy directly translates image pixels into land cover classes regardless of random noises for change identification. For instance, the supervised nonparametric methods have been proposed to detect land cover

transitions in remote-sensing image pairs [9]–[12]. Meanwhile, the post- or unsupervised-classification comparison [13], [14], and artificial neural network methods [15] also perform image classification before change detection. However, classification-based method provides change information which depends on the full delineation of individual images. Moreover, the variation introduced by seasonal differences in remote-sensing images is quite complex when identifying changed areas [16]–[19]. Therefore, the change detection is still one of the most challenging tasks in the remote sensing field due to the following aspects: 1) seasonal differences introduced by land cover objects (particularly, green vegetation) include lots of unwanted pseudochanges; and 2) the change detection rules empirically setting by trial-and-error selection on change sensitive features, which lack of robustness and suffer from underdetection or overdetection.

In order to exclude unwanted pseudochanges, it is important to find robust image features that can represent land cover information accurately and invariantly. Different from hand-crafted image features, deep-learning frameworks [20]–[26] are capable of finding high-level invariant features to represent complex image targets. To formulate the nonlinear change patterns of bitemporal images, the recurrent neural network (RNN) [27] is an ideal choice to tackle such complex transitions. The RNN automatically identifies abnormal changes from sequences lists, as it constantly feeds with labeled data from the training pool. Recently, a long-short-term memory (LSTM)-based RNN is deployed to learn a spectro-temporal, a joint feature representation from a bitemporal image sequence [28]–[30]. Then, the well-trained RNN model is applied to detect image content transitions based on specific land cover classes. Similarly, an RNN model based on the LSTM is used to formulate spectro-temporal features for land cover information extraction. However, the RNN model focuses on spectral sequence information of individual pixels which has neglected the spatial patterns in bitemporal images. To remedy this, the recurrent convolutional neural network (ReCNN) [31] is proposed to capture spectral, spatial, and temporal feature representations in complex bitemporal images for change-detection purposes. But, both RNN and ReCNN need lots of samples which are usually difficult to manually label from pseudochanges, especially for large-scale change detection tasks. To handle this problem, the generative adversarial network (GAN) as an un-/semi- supervised learning framework [32], [33], it requires a relatively small number of training samples to learn bitemporal transition patterns. For instance, a conditional GAN (CGAN) has been proposed to jointly use optical and SAR images for efficient change detection [34]. It significantly decreases pseudochanges by translating the optical images to satisfy the distribution of SAR data. Then, changed areas can be identified by implementing pixelwise detection from the common observation space. Similarly, a CGAN model has been successfully adapted to feed with bitemporal images for change area prediction [35]. However, GAN models that relied on maximizing likelihood formulations are hard to be stabilized, or even suffer from vanishing gradients during the training process [36], especially when feeding seasonal variation samples (such as trees with/without leaves). Metric learning aims to find a distance between data points in the same class that are smaller than those from different classes [37], [38]. It is learnt that nonlinear distance measurements using contrastive loss functions are more robust than those of GAN models. Therefore, how to incorporate metric learning to increase the stability of GAN models is urgently needed, especially with a limited number of elusive and contradictive samples.

In this article, we propose an efficient change detection method for bitemporal imagery comparison. To achieve the purpose of robust change identification, the metric learning-based GAN (MeGAN) model is introduced to exploit invariant feature representations from the limited number of training samples. Moreover, complex transition patterns between bitemporal images can be captured by the proposed change detection method, and thus it is able to identify target areas from pseudochange background. The contributions of this work are: 1) we propose the MeGAN for bitemporal image changes detection, which is able to automatically learn transition patterns in the spectro-spatio-temporal domain without human intervention and 2) to maximally suppress the pseudochange phenomenon, the seasonal transition term is considered to identify changed areas in bitemporal images.

The remainder of this article is organized as follows. The general review on change detection is listed in Section I. Then, Section II provides detailed information about the MeGAN algorithm as well as the seasonal transition formulation. Section III is designed to introduce experimental data sets, network setup, and experimental results. Final conclusions are given in Section IV.

II. Methodology

The framework of metric-learning-based GAN mainly contains the following steps: 1) bitemporal images alone with seasonal transition term are fed into the generator to predict potential change areas; 2) then, the change (both reference and predicted) maps are fed into the discriminator for fake/real discrimination, and simultaneously the MeGAN is being optimized with the metric-learning penalty; and 3) lastly, the optimized MeGAN is applied to detect real changes from the bitemporal images, as shown in Fig. 1. For the first step, the seasonal transition term is calculated by the ratio of vegetation-sensitive bands between bitemporal bands, while the generator formulates the transition pattern by considering spatial and seasonal changes. Then, the discriminator is deployed to make a determination on the true or fake samples with the metric-learning term based on matching image features and, lastly, change map predictions on bitemporal images can be obtained using the well-trained MeGAN.

A. Formulation of Regular GAN

The primary objective of GAN is based on game theory that consists of two individual counterparts which are a generator $G(z)$ and a discriminator $D(x)$. The goal of the generator is to learn a generative distribution p_z as similar as possible to the real data distribution $p_d(x)$. The discriminator aims to distinguish the samples whether they are from the generator p_z
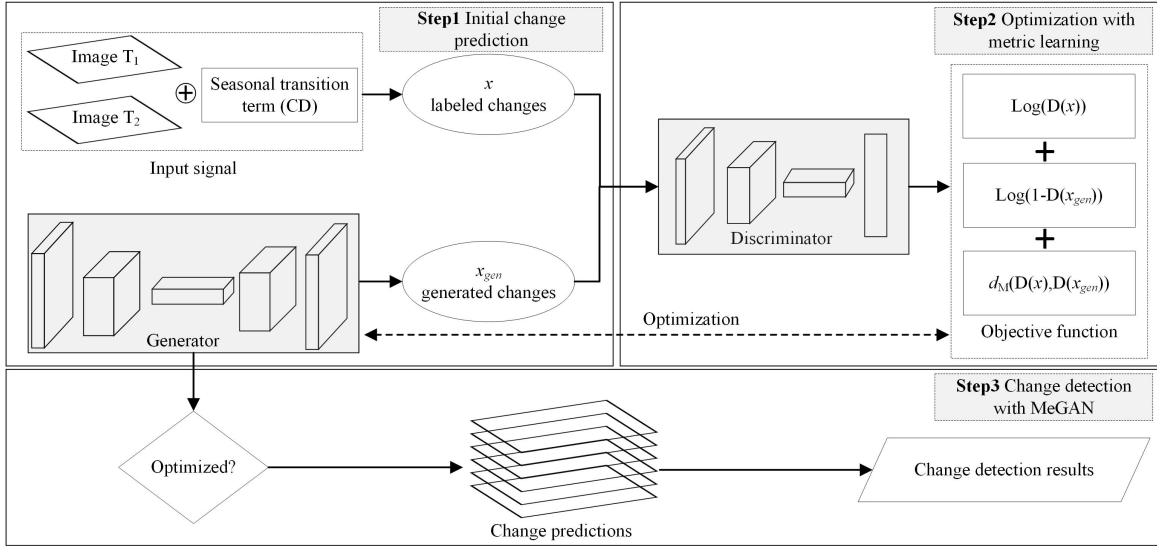
Fig. 1. Workflow of MeGAN for bitemporal image change detection. Bitemporal images are fed into the generator to output potential changes. Then, the discriminator is deployed to make a determination on the true or fake samples with the metric-learning term. Lastly, change map predictions on bitemporal images can be obtained using the well-trained MeGAN.

or from the real data distribution $p_d(x)$. During the training process, the generator $G(z)$ optimizes the parameters in the generative simulator and tries to fool the discriminator, and the discriminator $D(x)$ tries to upgrade the performance to make better discrimination on real or fake samples. To better formulate the minimax problem, the objective function of regular GAN is represented as

$$\min_G \max_D V(D, G) = E_{x \sim p_d(x)}[\log D(x)] \\ + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \quad (1)$$

To optimize this loss function, the stochastic gradient descent (SGD) approach is widely applied by sampling a minibatch of fakes from $p_z(z)$ and a minibatch of real samples from $p_d(x)$. The selected adversarial samples are used to minimize $V(D, G)$ with respect to parameters of G by assuming a constant D, and then maximize $V(D, G)$ with respect to parameters of D by assuming a constant G. The training process is to be finished when the loss function converges.

The goal of regular GAN is to train a generator to learn the real data distribution of p_d by transforming the input signal z. For the well-trained GAN, it is possible to utilize the generator to transform the input data to the target distribution. For this purpose, we have

$$p(z) = G(z) \approx p(d). \quad (2)$$

Although the concept of regular GAN is easy to deploy, it is difficult to optimize the loss function $V(D, G)$ during the training process, especially when feeding contradictory training samples (seasonal differences). Therefore, it almost becomes impossible for the generator to capture the distribution pattern of real data.

B. Metric Learning-Based GAN

To increase the stability of regular GAN during the training process, the metric learning is incorporated with

GAN (MeGAN) to enhance the performance of the GAN. Metric-learning aims to find a distance between data points in the same class that are smaller than those of different classes. Different from the regular GAN, the discriminator inside of MeGAN outputs a vector rather than a single scalar. Therefore, we denote the discriminator of MeGAN as $D(x)$. Let $X = [x_1, x_2, \ldots, x_N] \in R$ be the metric of N sample pairs, where $X_i = [x_i, x_i'] \in R (i = 1, 2, \ldots, N)$ consists of a pair of training samples. Here, in the MeGAN, x_i are from the real samples, whereas x_i' are from the fake samples from the generator. To measure the distances between data pairs, the Minkowski distances induced by ℓ_p norms are often used.

$$d_p(X, X') = \|X - X'\|_p = \left(\sum_{i=1}^{d} |x_i - x_i'| \right)^{\frac{1}{p}} \quad (3)$$

for $p \geq 1$. Three different widely used distances can be calculated. For $p = 1$, the Manhattan distance $d_{man}(X, X') = \sum_{i=1}^{d} |x_i - x_i'|$. For $p = 2$, the Euclidean distance $d_{euc}(X, X') = (\sum_{i=1} d|x_i - x_i^{prime}|^2)^{(1/2)}$, and for $p \to \infty$, the Chebyshev distance $d_{che}(X, X') = \max_i (x_i - x_i')$. Conventionally, the Mahalanobis distance is one of the most widely used term for metric learning. To better formulate the Mahalanobis distance, we have

$$d_M(X, X') = \sqrt{(X - X')M(X - X')} \quad (4)$$

where $M \in R$ is a symmetric positive semidefinite matrix. Equivalently, the Mahalanobis distance can be reformed as

$$d_M(X, X') = \sqrt{(X - X')L^T L(X - X')} \\ = \sqrt{(LX - LX')^T (LX - LX')} \quad (5)$$

where L represents the first few components of matrix M, where M has rank k, then $L \in R^{k \times d}$. Therefore, the Mahalanobis distance learns a global linear transformation which could be used to measure the discrepancy between data pairs.

To minimize the sum of distances that should be in the same class while maximizing the sum of distances from a different class. Therefore, the original formulation is to maximize the term of $\max \sum_{(x_i, x_j) \in D} d_M(x_i, x_j)$. For convenience, the objective function for the metric learning can be rewritten as

$$\min_{M, \xi \geq 0} \|M\|_F^2 + C \sum_{i,j,k} \xi_{ijk}$$
$$\text{s.t. } d_M^2(x_i, x_k) - d_M^2(x_i, x_j) \geq 1 - \xi_{ijk} M \succeq 0 \quad (6)$$

where $C \sum_{i,j,k} \xi_{ijk}$ is the regularization term to avoid the over-fitting. Similarly, the k-NN based strategy also applied to optimize the objective function of metric learning

$$\min_{M, \xi \geq 0} (1 - \mu) \sum_{(x_i, x_j) \in R} d_M^2(x_i, x_j) + \mu \sum_{i,j,k} \xi_{ijk}$$
$$\text{s.t. } d_M^2(x_i, x_k) - d_M^2(x_i, x_j) \geq 1 - \xi_{ijk} \quad (7)$$

where $\mu \in [0, 1]$ is a tradeoff parameter for objective function optimization. To consider the nonlinear metric learning, the deep neural networks are considered. In this article, the objective function of the metric learning term can be represented as

$$d_M(D(x_i), G(z_i)) = \sum_{x_i \in p_d(x), G(z_i) \in p_z} \|D(x_i) - D(G(z_i))\|_2^2 \quad (8)$$

where $d_M(D(x_i), G(z_i))$ represents the sum of the nonlinear distance between the extracted features from both real samples and generated ones. To be specific, the deep features refer to the activation outputs from mid-layer of the discriminator, except for the last output layer. Thus, this term can be optimized by Adam optimizer during the training progress of MeGAN. In general, the objective function for MeGAN is

$$\min_G \max_D V(D, G) = E_{x \sim p_d(x)}[\log D(x)]$$
$$+ E_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$
$$+ d_M(D(x_i), G(z_i)). \quad (9)$$

As stated above, the objective of MeGAN is to generate realist samples while considering the distance loss between the generator and the discriminator. Compared to the regular GAN, the MeGAN is more stable and easier to be trained.

C. Change Detection With MeGAN

To utilize the MeGAN for change detection, the seasonal transition term determination is one of the key components before calculation. Because of the pseudochanges with bitemporal images (especially for seasonal variations), it is difficult to exclude the seasonal pseudochanges from real changes. To handle this, the change vector analysis is introduced to formulate the possible pseudochanges by quantitatively measuring color variations. The change magnitude term can be formulated by

$$CM_{pixel} = \sum_{k=1}^{n} \|V_{ijk(t2)} - V_{ijk(t1)}\|^2 \quad (10)$$

here CM_{pixel} represents the change magnitude for each pixel between bitemporal images, and V_{ijk} represents the values

of pixel (i, k) for band k. However, the change magnitude only reveals the overall changes for each pixel which fails to identify seasonal transitions. Thus, the seasonal transition term is calculated to identify pseudochanges that are induced by seasonal changes, and it can be represented by

$$CD = \frac{V_{ijk(t2)} - V_{ijk(t1)}}{\sum_{k=1}^{n} \|V_{ijk(t2)} - V_{ijk(t1)}\|^2} \quad (11)$$

here CD is the change direction term that represents the seasonal changes in the color domain. To achieve the purpose of change detection, the seasonal transition term alone with bitemporal images are integrated as the input data for the MeGAN. Suppose, the bitemporal images are T_1 and T_2, and the ground truth change map is gt, the generator has the formulation

$$x_{gen} = G((T_1, T_2, CD)|gt). \quad (12)$$

The generator is fed with bitemporal images T_1, T_2 and the change-direction term CD. To reduce the computational complexity, the bitemporal images are converted to the panchromatic mode (i.e., single band for each image), whereas the color components are stored in the change direction map. Here, x_{gen} represents the output of the generator that aims to mimic the distribution of gt. Consequently, the objective function of the MeGAN for change detection has the following formulation:

$$\min_G \max_D T(D, G) = E_{x \sim p_d(x)}[\log D(x)]$$
$$+ E_{z \sim p_z(z)}[\log(1 - D(x_{gen}))]$$
$$+ d_M(D(x_i), D(x_{gen})) \quad (13)$$

where $p_d(x)$ is the subset of gt, and $p(z)$ is the data distribution of bitemporal images. The first two losses can be categorized as the optimization term for the discriminator, and the last loss is the penalty term for the generator. To optimize the objective function of the MeGAN, both loss terms for discriminator and generator are calculated and minimized during the training process with the help of Adam optimizer. Finally, the well-trained generator inside of the MeGAN is able to generate realistic samples by considering the bitemporal images T_1 and T_2.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Description

To illustrate the change detection ability of the proposed MeGAN, the experiments are conducted on two data sets for algorithm evaluation. Specifically, the selected two data sets are season-varying bitemporal images, and they have different spatial resolutions which are roughly 0.5 and 30 m, respectively.

The first data set was provided by Lebedev et al. [35], and it contains two images acquired at different times, covering a complex residential region. This data set was obtained by Google Earth (Digital Globe) with the sizes of 2700×4275 pixels for each image. The spatial resolution is approximately 0.5 m, and the first image was captured in summer and the second image was captured in autumn. With

Fig. 2.   Google Earth data set. (a) T1 image acquired in summer time. (b) T2 image acquired in autumn time. (c) Manually labeled real changes.
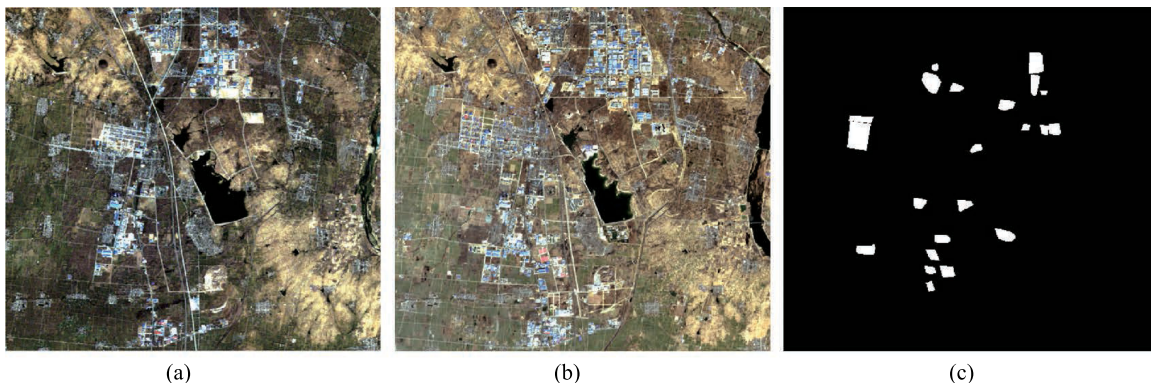

Fig. 3.   Landsat data set. (a) T1 image acquired in January 2010. (b) T2 image acquired in October 2015. (c) Manually labeled real changes.
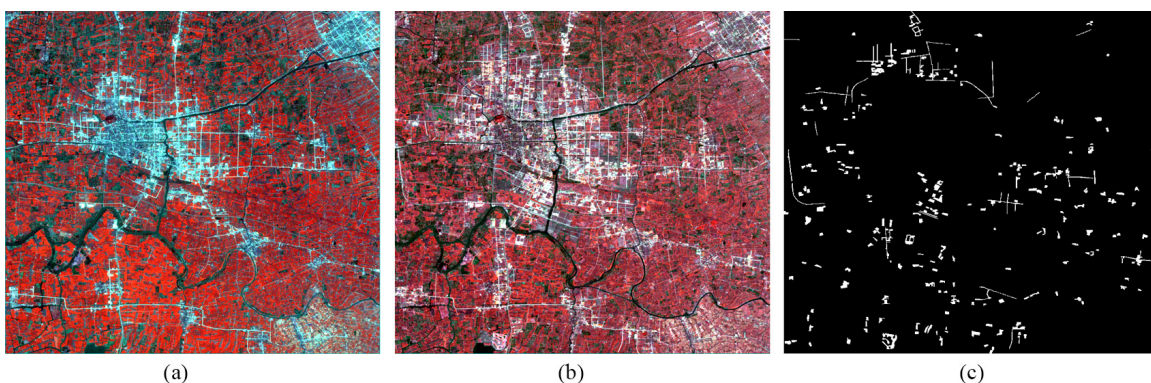

Fig. 4.   Kunshan data set. (a) T1 image acquired in March 2000. (b) T2 image acquired in February 2003. (c) Manually labeled real changes.

the submeter resolution, it allows us to detect small changes (such as cars, buildings, and roads) between the bitemporal images. Meanwhile, it is noticeable that there are lots of pseudochanges that are induced by seasonal changes, for instance, the trees, illumination conditions, as shown in Fig. 2. Therefore, the accurate change detection over such a complex region is quite challenging.

The second data set is the bitemporal subset data of Landsat satellite. This study area belongs to the North China Plain located in central Shandong Province, China. Data were acquired by the Landsat Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI) image for Path 122, Row 35. Data set consists of two images covering the city of Tai'an in January 2010 and October 2015. Both images in this data set are 506×469 (with near-infrared, red, and green bands), and the major land cover types of the area include

cropland, water, urban built-up, forest, and barren. Over the past few decades, rapid industrialization and urbanization have greatly changed the agricultural land pattern in this study area. The bitemporal images and its reference map are shown in Fig. 3.

To further demonstrate the robustness of the proposed method, we also introduced another Kunshan data set to test the performance of the proposed method. This data set consists of two images covering the city of Kunshan, China, on March 2000 and February 2003, acquired by the Landsat Enhanced Thematic Mapper Plus (ETM+) [39]. Each image contains 800×800 pixels where bitemporal changes are mainly involved with urban expansion. The spatial resolution is 30 m and we refer to the near-infrared, red, and green bands for change detection. The change reference map is acquired by manual labeling, as shown in Fig. 4.

TABLE I
DETAILED INFORMATION ABOUT THE CONFIGURATION OF THE MEGAN

Name	Layer	Kernel	Stride	Features	Activation
	1	128×128	2	2*ndf	ReLu
	2	64×64	2	2*ndf	ReLu
	3	32×32	2	4*ndf	ReLu
	4	16×16	2	8*ndf	ReLu
G	5	2×2	8	8*ndf	ReLu
	6	16×16	8	8*ndf	ReLu
	7	32×32	2	4*ndf	ReLu
	8	64×64	2	2*ndf	ReLu
	9	128×128	2	2*ndf	ReLu
	1	128×128	2	2*ndf	ReLu
	2	64×64	2	4*ndf	ReLu
D	3	32×32	2	8*ndf	tanh
	4	31×31	1	16*ndf	tanh
	5	30×30	1	1	tanh

B. MeGAN Configuration and Analysis

For the purpose of change detection, the seasonal transition term is calculated based on the near-infrared bands of these two data sets, respectively. However, due to the spectral information limitation of the Google Earth data set, we chose the green band for change direction estimation. In the meantime, the multiband images are transformed into the panchromatic mode for computational efficiency. During the experiment, the original bitemporal data sets were randomly cropped into small patches with the sizes of 256×256. Also, to increase the number of training samples the adjacent images have a distance of 64 pixels between central points for the Google Earth data set. Finally, Google Earth data set has 2457 patches, where 30% of the samples were taken as training samples and the rest are test samples. In addition, to avoid the possible correlation between training samples and test samples, we separate the whole data set into two individual areas. Specifically, the training area is 1:810 in rows and 1:4275 in columns, and the rest area is for evaluation. Similarly, for the Landsat data set, we separate it into two areas where the distance between adjacent patches is 32%, and 30% samples (16 samples versus 56 samples in total) are also included for training and the rest for the test. Finally, for the Kunshan data set, the training area is 1:256 in rows and the rest for the test.

1) Configurations of MeGAN: In order to achieve the purpose of end-to-end change detection between bitemporal images, we utilized the pix2pix-similar framework to directly translate the bitemporal images into change maps, as shown in Table I. Different from the previous implementation, the change direction feature was fed into the generator for change detection. The generator is based on the U-net framework which contains encoder–decoder convolution operations and skips links for image transformation.

The discriminator is modified from a deep CNN, where the main contribution of the modified framework is the metric-learning term. In this implementation, the discriminator feeds with bitemporal images alone with the change maps either from the real or fake data set. Then the

discriminator tries to discriminate the change maps that come from ground-truth data set rather than the generator. At the same time, the metric-learning term is calculated by measuring the discrepancy between real samples and generated ones. During the training process, the learning rate is set to 0.0002, while the number of generator filters in the first convolution layer is 20 and the number of discriminator filters in the first layer is also 20.

2) Impacts of Metric Learning Term: The GAN framework that aims to capture the distribution pattern of the real data set often encountered the problem of gradients vanishing which leads to failure training. To stabilize the adversarial network, the metric learning term is introduced to increase the stability of the GAN framework. To be specific, the metric distances between the generated samples and the real ones are measured by the discriminator through hierarchical convolution and activation. In this article, we measured the distance by matching the first four deep features generated by the discriminator. Then, the loss function is incorporated into adversarial losses to be optimized by Adam optimizer. By incorporating metric learning, it is possible to train the adversarial network stably and efficiently.

In order to quantitatively measure the impacts of metric learning, we tested the MeGAN with/without distance regularization. Specifically, the Mahalanobis distance term is added to measure the similarity between real and fake samples. Based on the previous settings, the training processes of the MeGAN with/without distance regularization are displayed in Fig. 5. From this figure, we can conclude that both discriminator and generator losses are changing rapidly without distance regular regularization term. After 80k iterations, the discriminator is minimized to around 0.2 where the generator keeps rising until it reaches 7. Therefore, the generator is generally losing its power in generating realistic samples that are distinguished by the discriminator. Compared to nonmetric learning MeGAN, the metric learning-based MeGAN is more stable during the training process. To be specific, the loss values of the discriminator are gradually decreased to 0 around 20k iterations. After that, the generator gains the power of generating realistic samples and the loss values keep rising from 0 up to around 0.5 in the end. Meanwhile, the generator keeps optimizing its loss function and increasing the power of realistic sample generation. In general, the MeGAN with distance regularization is more suitable for change detection for seasonal varying bitemporal images.

3) Seasonal Pseudochange Suppression: Owing to the contradictory samples induced by seasonal differences, it is difficult to learn invariant features that represent real changes. For instance, the vegetation including trees and grassland are quite green in the growing season, whereas the leaves decay during winter, as shown in Fig. 6. Therefore, the huge differences in terms of color and spatial contents are the main obstacle in identifying the real changes between bitemporal images. Specifically, the ground information (such as cars and roofs) may be occluded by trees and other vegetation canopies, which increased the difficulty in change detection. In some extreme cases, the image patches may be fully covered by canopies that blocked the land cover objects. How to
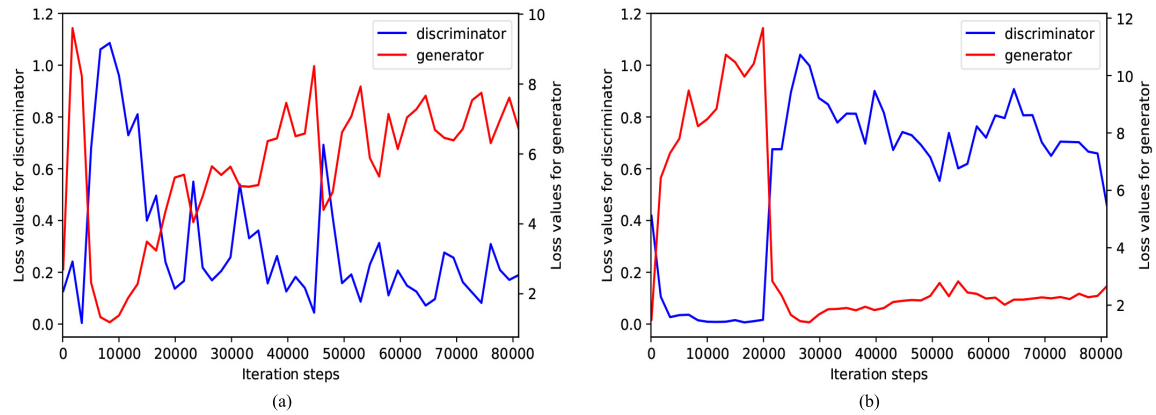
Fig. 5.   Loss function values. (a) MeGAN without distance metric regularization. (b) MeGAN with the metric learning term.
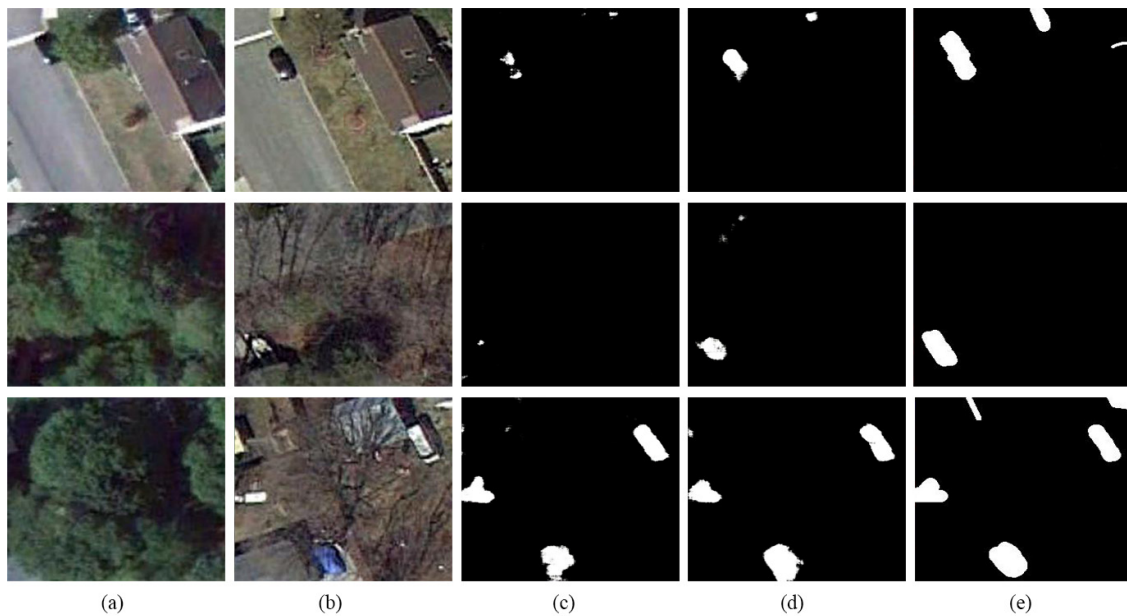

Fig. 6.   Illustration of seasonal varying between bitemporal images and the performances of the MeGAN for change detection. (a) Original image of T1. (b) Original image of T2. (c) Change detection results by using MeGAN without using seasonal transition term. (d) Change detection results by using MeGAN with seasonal transition term. (e) Reference map.

detect the changed areas while ignoring pseudochanges caused by vegetation canopies is one of the biggest challenges for bitemporal imagery change detection.

In the framework of MeGAN, we introduced the seasonal transition term to represent the pseudochanges induced by seasonal varying. To achieve this purpose, we calculated the change direction [as shown in (11)] of the bitemporal images by using the vegetation sensitive bands (green or near-infrared). Then, the vegetation transition term is fed into the MeGAN alone with the bitemporal images for change detection. To illustrate the seasonal variation between bitemporal images, some parts of the change detection results are shown in Fig. 6. Without seasonal transition regularization, the change detection results may suffer from under-identifying. In the meantime, the change direction term [i.e., (11)] is able to capture the pseudochanges caused by vegetation, while the MeGAN can automatically reject any pseudochanges that it represents. Therefore, the change-detection results have

significantly improved when the seasonal transition term is introduced. In general, the results of change detection are quite promising in terms of identifying contradictory image pairs to find real change areas.

C. Change Detection Results and Comparison

To demonstrate the robustness of the proposed method, we included the traditional change-detection methods such as CVA [40], postclassification change detection (PCC) [41], and pix2pix framework [42]. For the CVA, we calculated the distances of each pixel between the bitemporal images. Based on the distance values, the change magnitude map can be obtained. At last, the change areas are identified by implementing the auto-threshold algorithm. For the PCC method, the ISO data-based unsupervised segmentation strategy is applied to generate classification maps for each image. Then, the change map can be derived from these two frames of classification
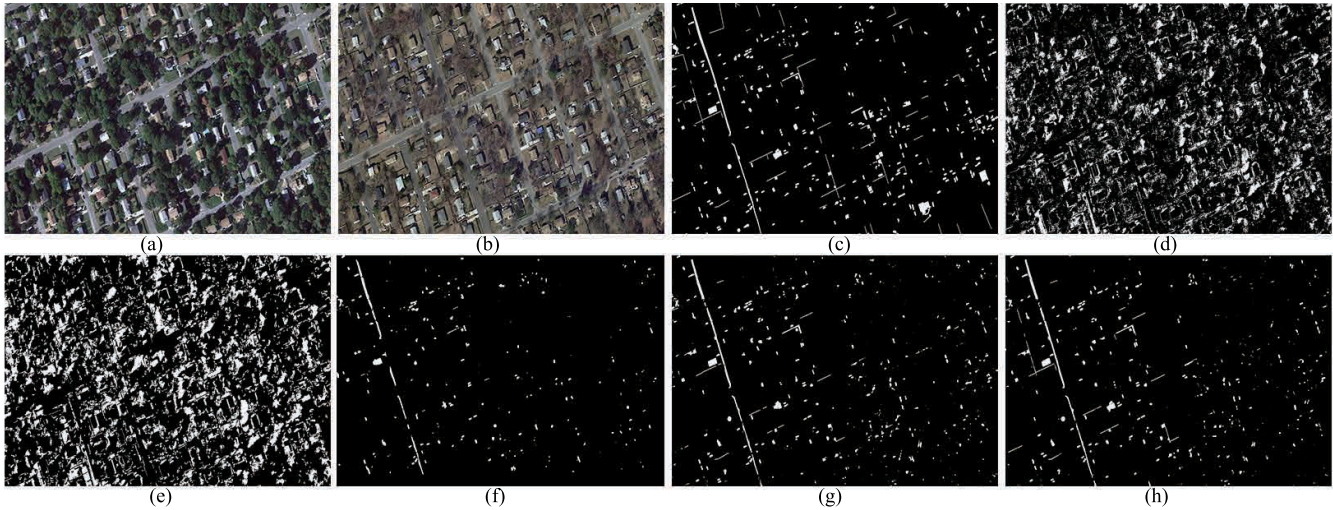
Fig. 7. Change detection results on Google Earth data set. (a) Original image of T1. (b) Original image of T2. (c) Reference change map. (d) CVA change detection result. (e) PCC change detection result. (f) pix2pix GAN change detection result. (g) MeGAN change detection result without metric learning. (h) MeGAN change detection result.

map. In addition, for the pix2pix GAN, we implemented it as mentioned in [35], and two bitemporal images are directly fed into a framework for change map prediction.

Meanwhile, to quantitatively measure the change detection accuracy, we calculated the bounding boxes and their intersection over union (IoU) of polygons to represent the performance of change maps. In particular, the IoU is calculated by implementing IoU = (True/Union). For convenience, we mark a correct detection as true positive (TP), a wrong detection as false positive (FP), and a ground truth not detected as false negative (FN). Precision is the percentage of correct positive predictions and is given by Precision = (TP/TP + FP) = (TP/all detections) and the recall represents percentage of TP detected among all relevant ground truths and is given by Recall = (TP/TP + FN) = (TP/all ground truths) [43].

D. Google Earth Data Set

For the Google Earth data set due to the high spatial resolution and significant seasonal changes, it is difficult to separate real changes from noisy background information. To identify change areas between bitemporal images, we implemented the CVA, PCC, and pix2pix framework as the standard benchmark. In addition, the proposed MeGAN is also been included for comparisons. The classification map is illustrated in Fig. 7. As shown in this figure, we can conclude that traditional change detection methods have failed in identifying real changes given such a complex background. To be specific, the CVA calculated the change magnitude as well as the change direction vectors for change detection. However, due to the significant difference in illumination and phenology conditions, the CVA falsely identified roofs, roads, and canopies as potential change areas. Similarly, the PCC first classified land cover objects by using unsupervised clustering algorithms. Then, change areas can be derived by using postclassification comparisons. But, the classification of bitemporal images is difficult when the spatial resolution is high. Therefore, how
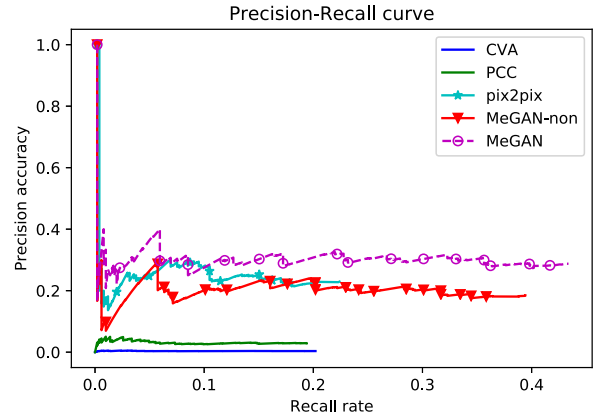

Fig. 8. Precision-recall curve of change detection results on Google Earth data set. (a) CVA change detection. (b) PCC change detection. (c) pix2pix GAN change detection. (d) MeGAN change detection without metric learning. (e) MeGAN change detection.

to extract robust change features from the self-contradictory image samples is one of the biggest changes for accuracy change detection. To handle it, the pix2pix framework aims to automatically explore high-level robust features to represent real changes between bitemporal images. The pix2pix learns the translation pattern from the T1 image to the T2 image, and the change map can be obtained by predicting the entire data set. The change detection map of pix2pix is illustrated in Fig. 7(f), where lots of pseudochanges are excluded but with significant omission errors. Complementary, the MeGAN is able to predict the real changes by simultaneously considering seasonal transition and metric learning term.

To quantitatively measure the change detection results, we borrowed the idea for measuring the object detection results and introduced the bounding box fitness (IoU) as an indicator to calculate precision and recall rate. The precision and recall rate is plotted in Fig. 8, where the IoU threshold
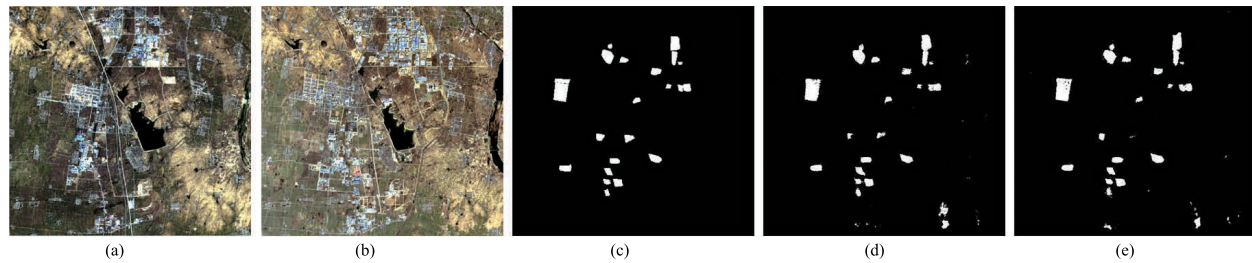
Fig. 9.   Change detection results on Landsat data set. (a) Original image of T1. (b) Original image of T2. (c) Reference change map. (d) MeGAN without metric learning. (e) MeGAN change detection.

TABLE II
CHANGE DETECTION ACCURACY ON GOOGLE EARTH
DATA SET (IN PERCENTAGE)

Method	precision	recall	f1-score	kappa
CVA	0.09	0.37	0.14	0.08
PCC	0.07	0.39	0.11	0.05
Pix2Pix	0.80	0.33	0.44	0.43
MeGAN-non	0.76	0.51	0.59	0.58
MeGAN	0.82	0.51	0.63	0.60

TABLE III
CHANGE DETECTION ACCURACY ON LANDSAT
DATA SET (IN PERCENTAGE)

Method	precision	recall	f1-score	kappa
MeGAN-non	0.84	0.81	0.83	0.83
MeGAN	0.88	0.83	0.86	0.85

was set to 0.5. In general, the accuracy of traditional change detection methods is significantly lower than that of deep learning-based methods. In this figure, a good change detector means that its precision stays high as recall increases. From this point of view, both CVA and PCC are quite low in precision and recall rates, which are even lower than 0.1. The pix2pix framework is significantly better than the traditional change-detection strategies. The precision of pix2pix can be as high as 0.3 when the recall rate is around 0.1. However, due to the huge differences in terms of vegetation land covers, the pix2pix method suffered from significant omission errors. Complementarily, the MeGAN that introduced the seasonal transition term is able to suppress pseudochanges. Thus, the recall rates can almost reach 0.4 for the MeGAN without metric learning (MeGAN-non) and 0.45 for the MeGAN. Meanwhile, the precision accuracy at a recall rate of 0.4 can be as high as 0.38 for the MeGAN. Finally, the average precision (AP) can be calculated based on areas between the precision-recall curve and the x-axis. Therefore, the average precision for the CVA is the lowest one which is only 0.08%, and the PCC has an accuracy of 0.67%. When using the deep-learning method, the average accuracy is increased to 5.81%. Finally, with the introduction of seasonal transition term and metric-learning strategy, the average accuracies can be as high as 8.80% and 14.24%, respectively. Meanwhile, we also introduce the f1-score and kappa index to measure the change detection results, as shown in Table II.

E. Landsat Data Set

Similarly, the proposed MeGAN is applied to the Landsat data set for change detection. Different from the Google Earth data set, it has a coarser resolution, and the land covers are more complex without any fixed spatial patterns. For the Landsat scene, we focus on the changes taken place on the

cropland which has been dramatically occupied due to fast urbanization in China. To increase the difficulty of change detection, we chose two images with significant seasonal differences. Most pseudochanges that occurred on croplands were usually green in summer and became bare soil in winter. Different from the Google data set, we utilized the standard pseudocomposite (near-infrared, red, and green) to demonstrate the robustness of the seasonal transition term and the MeGAN model. Therefore, the non-MeGAN and MeGAN were employed in this section to detect real changes that occurred between the bitemporal images. Change detection results are shown in Fig. 9.

From the change detection maps, we can conclude that the proposed MeGAN has achieved a good performance in terms of identifying real changes even with the near-infrared band. This demonstrates that both natural color combination and standard pseudocomposite are suitable for seasonal transition pattern recognition. The detailed information about change detection results of the Landsat scene is illustrated in Table III. The change detection results indicate that the MeGAN is able to learn stable transition patterns from bitemporal images. Specifically, the vegetation cover in the T1 image is denser than that in the T2 image. In addition, due to the vari- ation of illumination conditions between seasonal varying images, the reflectance values for the same land cover targets may present with different values. To handle this problem, the framework of the GAN can effectively shed light into the invariant features that indicate pseudochanges as well as the real changes. In this data set, the nonmetric learning strategy and the metric learning strategy have been applied to predict the real changes between seasonal varying images. As a result, the recall rate for the non-MeGAN is 0.82 and 0.84 for the metric learning-based MeGAN.

F. Kunshan Data Set

Different from the previous Landsat data set, the Kunshan data set has a larger geographical coverage with a
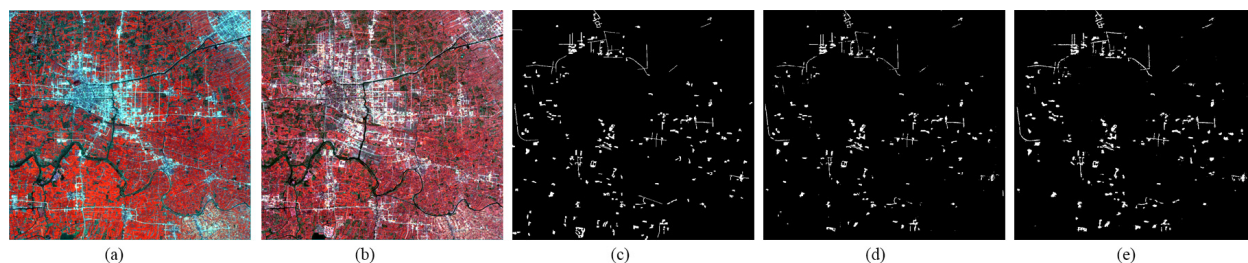
Fig. 10. Change detection results on Kunshan data set. (a) Original image of T1. (b) Original image of T2. (c) Reference change map. (d) MeGAN without metric learning. (e) MeGAN change detection.

TABLE IV
CHANGE DETECTION ACCURACY ON KUNSHAN
DATA SET (IN PERCENTAGE)

Method	precision	recall	f1-score	kappa
MeGAN-non	0.71	0.64	0.71	0.69
MeGAN	0.82	0.70	0.72	0.71

complex urban background. Because of the seasonal differences, the vegetation presented change-detection results in Fig. 10(a) and (b) with significant variations. To reduce the impact of seasonal differences, the MeGAN with seasonal transition term was deployed to detect real changes from a complex background. The changes occurred in these bitemporal images are mainly due to urban expansion. In order to detect real changes, the seasonal transition term was calculated to maximally suppress pseudochanges. Meanwhile, the MeGAN model aims to capture the transition pattern between the two images. Change detection results are presented in Fig. 10 and their accuracies are reported in Table IV. We can conclude that the MeGAN is able to capture the bitemporal transition patterns while excluding noises introduced by seasonal difference.

IV. CONCLUSION

In this article, we proposed a MeGAN to detect real changes from bitemporal images. To be specific, a season-invariant transition term is proposed to suppress the pseudochanges in bitemporal images. Then, the metric learning strategy is introduced to improve the performance of the traditional GAN, in terms of change detection. The main contribution of this article is that we introduced a metric learning-based strategy that incorporating a seasonal transition term to learn invariant bitemporal transitional features. Although the proposed MeGAN has the ability to learn robust transitional features, still, the task of change detection in long-term time-series images is quite challenging, especially at larger scales. How to find periodic transition patterns in time-series images will be one of the major challenges in our future work.

ACKNOWLEDGMENT

REFERENCES

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

[2] 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.

[3] 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.

[4] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, Jan. 2015.

[5] V. Ferraris, N. Dobigeon, Q. Wei, and M. Chabert, "Detecting changes between optical images of different spatial and spectral resolutions: A fusion-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1566–1578, Mar. 2018.

[6] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multi-scale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, Sep. 2017.

[7] 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.

[8] P. He, W. Shi, H. Zhang, and M. Hao, "A novel dynamic threshold method for unsupervised change detection from remotely sensed images," *Remote Sens. Lett.*, vol. 5, no. 4, pp. 396–403, 2014.

[9] L. Bruzzone, R. Cossu, and G. Vernazza, "Detection of land-cover transitions by combining multidate classifiers," *Pattern Recognit. Lett.*, vol. 25, no. 13, pp. 1491–1500, 2004.

[10] M. Zanetti and L. Bruzzone, "A theoretical framework for change detection based on a compound multiclass statistical model of the difference image," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1129–1143, Feb. 2018.

[11] 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, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425717303140

[12] H. Luo, C. Liu, C. Wu, and X. Guo, "Urban change detection based on Dempster–Shafer theory for multitemporal very high-resolution imagery," *Remote Sens.*, vol. 10, no. 7, p. 980, 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/7/980

[13] T. Leichtle, C. Geiß, M. Wurm, T. Lakes, and H. Taubenböck, "Unsupervised change detection in VHR remote sensing imagery—An object-based clustering approach in a dynamic urban environment," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 54, pp. 15–27, Feb. 2017.

[14] G. Cao, X. Li, and L. Zhou, "Unsupervised change detection in high spatial resolution remote sensing images based on a conditional random field model," *Eur. J. Remote Sens.*, vol. 49, no. 1, pp. 225–237, Jun. 2016.

[15] F. S. Erbek, C. Özkan, and M. Taberner, "Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1733–1748, 2004.

[16] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[17] J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 3, pp. 236–248, 2007.

[18] V. Walter, "Object-based classification of remote sensing data for change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, nos. 3–4, pp. 225–238, 2004.

[19] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogram. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.

[20] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[21] W. Zhao, S. Du, and W. J. Emery, "Object-based convolutional neural network for high-resolution imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3386–3396, Jul. 2017.

[22] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[23] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.

[24] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.

[25] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 5071–5074.

[26] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2016, pp. 1–9.

[27] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[28] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[29] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, p. 506, 2016.

[30] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.

[31] 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.

[32] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[33] Q. Shi, X. Liu, and X. Li, "Road detection from remote sensing images by generative adversarial networks," *IEEE Access*, vol. 6, pp. 25486–25494, 2018.

[34] M. Gong, X. Niu, P. Zhang, and Z. Li, "Generative adversarial networks for change detection in multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2310–2314, Nov. 2017.

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

[36] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[37] H.-J. Ye, D.-C. Zhan, X.-M. Si, Y. Jiang, and Z.-H. Zhou, "What makes objects similar: A unified multi-metric learning approach," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1235–1243.

[38] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019, doi: 10.1016/j.ins.2019.02.008.

[39] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.

[40] J. Chen, P. Gong, C. He, R. Pu, and P. Shi, "Land-use/land-cover change detection using improved change-vector analysis," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 4, pp. 369–379, 2003.

[41] P. Serra, X. Pons, and D. Saurí, "Post-classification change detection with data from different sensors: Some accuracy considerations," *Int. J. Remote Sens.*, vol. 24, no. 16, pp. 3311–3340, 2003.

[42] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv:1611.07004*. [Online]. Available: https://arxiv.org/abs/1611.07004

[43] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, 2015, Art. no. e0118432.
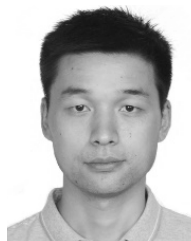
Wenzhi Zhao received the Ph.D. degree from Peking University, Beijing, China, in 2018.

In 2017, he was a Visiting Scholar with the University of Colorado, Boulder, CO, USA. He currently holds a Post-Doctoral/Lecturer position at Beijing Normal University, Beijing. His research interests include remote sensing big data, spatial-temporal data mining, and machine learning, especially deep networks and their applications in remote sensing.
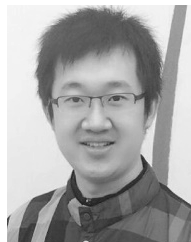
Lichao Mou (S'16) received the bachelor's degree in automation from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2012, and the master's degree in signal and information processing from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2015. He is currently pursuing the Ph.D. degree with the German Aerospace Center (DLR), Wessling, Germany, and the Technical University of Munich (TUM), Munich, Germany.

In 2015, he spent six months with the Computer Vision Group, University of Freiburg, Breisgau, Germany. In 2019, he was a Visiting Researcher with the University of Cambridge, Cambridge, U.K. His research interests include remote sensing, computer vision, and machine learning, especially deep networks and their applications in remote sensing.

Mr. Mou received first place in the 2016 IEEE GRSS Data Fusion Contest and was a finalist for the Best Student Paper Award by the 2017 Joint Urban Remote Sensing Event and for the 2019 Joint Urban Remote Sensing Event.

Jiage Chen received the Ph.D. degree from the Faculty of Geographical Science, Beijing Normal University, Beijing, China.

She is working as an Engineer with the National Geomatics Center of China, Beijing. Her research interests are in the areas of remote sensing image processing, change detection, and time series analysis in remote sensing images.

Yanchen Bo (M'08) received the B.S. degree in geography from Lanzhou University, Lanzhou, China, in 1996, and the M.S. and Ph.D. degrees in GIS and remote sensing from the Chinese Academy of Sciences, Beijing, China, in 1999 and 2002, respectively.

He is currently a Professor of remote sensing with the Faculty of Geographical Science, Beijing Normal University, Beijing, and the State Key Laboratory of Remote Sensing Science, jointly sponsored by Beijing Normal University, Beijing, and the Chinese Academy of Sciences, Beijing. His research interests focus on multi-source remote sensing data fusion and spatio-temporal scaling of remote sensing information, as well as the time series remote sensing data analysis and change detection.

William J. Emery (M'90–SM'01–F'03–LF'19) received the Ph.D. degree in physical oceanography from the University of Hawaii, Honolulu, HI, USA, in 1975.

He was with Texas A&M University, College Station, TX, USA. In 1978, he moved to The University of British Columbia, Vancouver, BC, Canada, where he created a Satellite Oceanography Facility/Education/Research Program. In 1987, he was appointed as a Professor of aerospace engineering sciences with the University of Colorado, Boulder, CO, USA. He is also an Adjunct Professor of informatics with the University of Rome Tor Vergata, Rome, Italy. He has authored more than 182 refereed publications, 2 textbooks, and more than 150 conference papers.

Dr. Emery was an Elected Fellow of the American Meteorological Society in 2010, the American Astronautical Society in 2011, and the American Geophysical Union in 2012. He was a recipient of the 2004 IEEE Geoscience and Remote Sensing Society (GRSS) Educational Award and the 2009 GRSS Outstanding Service Award. He was the Founding Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2004 and its Editor-in-Chief for six years. He serves as the Editor-in-Chief for the *Journal of Atmospheric and Oceanic Technology*. He is also the Vice President of Publications of the IEEE GRSS.