Detecting Changes by Learning No Changes: Data-Enclosing-Ball Minimizing Autoencoders for One-Class Change Detection in Multispectral Imagery

Lichao Mou, Yuansheng Hua, Sudipan Saha, *Member, IEEE*, Francesca Bovolo, *Senior Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*— Change detection is a long-standing and challenging problem in remote sensing. Very often, features about changes are difficult to model beforehand, thus making the collection of changed samples a challenging task. In comparison, it is much easier to collect numerous no-change samples. It is possible to define a change detection approach using only easily available annotated no-change samples, which we henceforth call one-class change detection. Autoencoder networks being trained on no-change data are natural candidates for addressing this task due to their superior performance when compared with other one-class classification models. However, the autoencoders usually suffer from the problem of overgeneralization, i.e., they tend to generalize too well, thus risking properly reconstructing changed samples. In this article, we propose a novel data-enclosing-ball minimizing autoencoder (DebM-AE) that is trained with dual objectives—a reconstruction error criterion and a minimum volume criterion. The network learns a compact latent space, where encodings of no-change samples have low intraclass variance, which as counterpart has the identification of changed instances. We conducted extensive experiments on three real-world datasets. Results demonstrate advantages of the proposed method over other competitors. We make our data and code publicly available (https://gitlab.lrz.de/ai4eo/reasoning/DebM-AE; https://github.com/lcmou/DebM-AE).**

***Index Terms*— Autoencoder network, change detection, one-class classification, remote sensing.**

Manuscript received 24 March 2021; revised 8 August 2021, 28 December 2021, 4 April 2022, and 23 June 2022; accepted 19 July 2022. Date of publication 30 August 2022; date of current version 22 September 2022. This work was supported in part by the European Research Council (ERC) through the European Union's Horizon 2020 Research and Innovation Program under Grant ERC-2016-StG-714087, (Acronym: *So2Sat*), in part by the Helmholtz Association through the Framework of Helmholtz AI under Grant ZT-I-PF-5-01-Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)," in part by the Helmholtz Excellent Professorship "Data Science in Earth Observation-Big Data Fusion for Urban Research" under Grant W2-W3-100, in part by the German Federal Ministry of Education and Research (BMBF) in the framework of the International Future AI Laboratory "AI4EO–Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" under Grant 01DD20001, and in part by the German Federal Ministry of Economics and Technology in the framework of the "National Center of Excellence ML4Earth" under Grant 50EE2201C. *(Corresponding authors: Lichao Mou; Xiao Xiang Zhu.)*

Lichao Mou, Yuansheng Hua, and Xiao Xiang Zhu are with the Data Science in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany, and also with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: lichao.mou@tum.de; yuansheng.hua@tum.de; xiaoxiang.zhu@tum.de).

Sudipan Saha is with the Chair of Data Science in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: sudipan.saha@tum.de).

Francesca Bovolo is with the Fondazione Bruno Kessler, 38100 Trento, Italy (e-mail: bovolo@fbk.eu).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38050 Trento, Italy (e-mail: lorenzo.bruzzone@unitn.it).

Digital Object Identifier 10.1109/TGRS.2022.3200985

I. Introduction

WE LIVE in a dynamic world where things change all the time. The recent advances in remote sensing platforms open up new possibilities for observing dynamic changes of our planet from a bird's eye view. New images are populated daily, e.g., the Landsat 8 satellite acquires more than 700 scenes a day, and in the same time span Sentinel-2 produces over 4 Terabytes of fresh images. Hence, multitemporal data analysis is becoming increasingly important. In the field of multitemporal image analysis [1], [2], change detection is a long-standing research problem. It deals with changes both in natural resources and in man-made structures. For instance, in the former case, change detection enables deforestation monitoring, disaster assessment, and crop stress detection [3], [4]. In the latter one, it can aid in city monitoring and planning [5], [6], [7].

There are numerous methods for detecting changes in remote sensing imagery. For the biggest part, they can be divided into three classes, supervised, semisupervised, and unsupervised ones, differing in the use of labeled data. Both the supervised and semisupervised methods require well-labeled samples from changed and no-change areas, and the difference between them is that the latter also exploit unlabeled samples to assist the labeled ones in the training phase. Albeit successful, these approaches suffer from the fact that in most situations, ground-truth data, particularly for changed regions, are difficult to acquire. This is because features about changes are often unknown or difficult to model beforehand because of the variety of possible kinds of change. For this reason, the unsupervised change detection models are conceptually of high interest and have been widely studied over the past decades [8], [9], [10], [11]. The underlying

assumption of unsupervised models is that nothing is known about the data and task, and samples are grouped into two or more categories (change/multitype changes versus no change) by clustering or thresholding algorithms.

To take a deep dive into this task, we observe how humans identify the changes in bitemporal images. A human annotator usually switches between the pre- and post-images repeatedly and tries to spot significant differences (change) from the background (no change). The annotator may not be aware of the expected changes, but has a clear understanding and knowledge of no change. On the other hand, information about changes is often hard to obtain in many applications, e.g., disaster management, but it is much easier to have information on the nature of no-change regions, and this can be even beforehand. This has motivated researchers to formulate binary change detection as a one-class classification task where samples of only *one class*, i.e., no-change, are used to train a model that is used to distinguish changed instances from no-change ones. Hereafter, we call this paradigm as one-class change detection.

We find that autoencoders are natural candidates for this task: these models can learn no-change profiles given only no-change samples and identify changes as samples not conforming to the no-change profiles. The autoencoders are trained by minimizing the reconstruction error on no-change instances and use the reconstruction error as an indicator of changes, i.e., it is assumed that changed samples cannot be reconstructed well. However, we find that this assumption does not always hold, and the autoencoders often overgeneralize, that is, sometimes they generalize so well to closely reconstruct changed pixel pairs, thus in the end failing to recognize them as such. To address it, we propose a data-enclosing-ball minimizing autoencoder (DebM-AE). Specifically, in this article, we consider one-class change detection in multispectral images and a pixel-based model. In our model, an autoencoder is trained with dual objectives—a traditional reconstruction error criterion and a minimum volume criterion. The latter minimizes the volume of the space enclosing training sample pairs in the latent space of the autoencoder. The result of training is that the encoder learns to transform no-change instances to an extremely compact feature space where the samples have very low intraclass variance. Thus, changed pixel pairs can be better separated. The main contributions of this work are as follows. First, we explore the use of the autoencoder framework for supervised one-class change detection tasks in which only easily available labeled no-change samples are needed for model training. This paradigm shows superior performance compared with the conventional one-class classification algorithms and outlier detection models. Furthermore, we analyze the overgeneralization problem of autoencoders in this task. To address it, we propose a novel and effective autoencoder network, namely, DebM-AE, which is trained based on dual objectives and can aid in separating and thus recognizing changed pixel pairs more accurately. Second, we build polygonal training/test masks for four real-world datasets and use them to evaluate the effectiveness of the proposed network. Moreover, we conduct experiments to verify that our method has good transferability.

The remainder of this article is organized as follows. Section II gives a brief overview of the current methods for change detection in bitemporal remote sensing images with a focus on the supervised, semisupervised, and unsupervised models, as well as the existing approaches for one-class change detection. Section III presents our proposed DebM-AE architecture. In Section IV, the datasets, evaluation metrics, and competitors are introduced. Furthermore, we present the numerical results and a discussion of the observed model performance in Section V. Finally, Section VI concludes this article.

II. Related Work

In this section, we explore the state-of-the-art change detection methods which are divided into three categories, supervised, semisupervised, and unsupervised models, differing in the use of labeled data. The one-class models are particularly reviewed in Section II-A.

A. Supervised Change Detection

The development of supervised learning algorithms in the field of machine learning provides insights into the supervised change detection methods. Early efforts have gone into seeking out conventional classifiers, to name a few, random forest [12], [13], [14] and support vector machine (SVM) [15], [16]. In recent years, with the rising popularity of deep learning, neural networks have been applied to change detection tasks. For instance, [17] uses a long short-term memory (LSTM) network—a kind of recurrent neural network (RNN)—to detect the changes in bitemporal multispectral images, and this model shows good generalization results. Later, [18] proposes a recurrent convolutional neural network (CNN) that is able to significantly remove salt and pepper noise from change detection maps and thus obtain better results. In [19], a 2-D CNN is introduced to learn high-level features, and in [20], the authors use a recurrent 3-D CNN to extract spectral–spatial–temporal features for change detection. In [21], the authors introduce a Siamese CNN to extract features of bitemporal images and use a weighted contrastive loss to alleviate the influence of imbalanced data. For high spatial resolution remote sensing images, this task can be deemed as a dense (pixel-wise) prediction task and solved by the semantic segmentation network architectures [22]. In [23], several Siamese U-Nets are designed for semantic change detection with very high-resolution images. The authors of [24] use an improved UNet++ for this task. To tackle pseudo changes caused by seasonal transitions in change detection tasks, [25] proposes a metric-learning-based generative adversarial network (GAN), termed MeGAN, to learn seasonal invariant feature representations. Furthermore, some modern network architectures and techniques, e.g., CapsNet [26], [27], self-supervised learning [28], [29], attention mechanism [30], and Transformer [31], also shed light on this problem. For example, in [26], the authors use the convolutional capsule networks to extract vector-based features of bitemporal images and then produce binary change maps by computing their cosine similarities and differences. [31] tokenizes feature maps

learned by a CNN backbone and feed them into Transformer layers for encoding contexts in space-time. Afterward, tokens are rearranged as feature maps, of which differences are calculated by subtraction for predicting final change maps. Although successful, these models require massive training samples to provide sufficient supervision information, which restricts their application scenarios.

It is noteworthy that in change detection, supervision can be in the form of either labeled training pixels scattered across the target scene [18] or labeled training scenes that are spatially disjoint with the target one [23]. Our work belongs to the first category but uses polygon annotations, which is relevant in scenarios where a human annotator can quickly label some data, but not the entire target scene. For polygon-level annotations, pixels within each polygon are assigned the same label (change or no change). In addition, we also show the performance of the proposed method using the second category of labeling (see Section V-D).

One-class change detection is a special case of supervised learning algorithms, as only no-change samples are used. In [32], the authors make use of a one-class SVM (OC-SVM) for change detection. The authors of [33] propose a semisupervised OC-SVM and apply it to one-class change detection. [34] exploits a cost-sensitive SVM for the same task and uses an entire solution path algorithm to facilitate and accelerate the parameter selection of the cost-sensitive SVM. From bitemporal images to long image time series, in [35], the authors propose a one-class method for change detection in image time series, where no-change image time series is used to train an LSTM network that is subsequently applied to detect the changes from target image time series. Moreover, when the number of changed instances is smaller than that of no-change samples in a scene, the outlier detection methods? [36], [37], [38], [39], [40] can be applied into the one-class change detection tasks. We also evaluate several outlier detection approaches in our case (see Sections IV and V).

B. Semisupervised Change Detection

Semisupervised learning allows making use of a huge amount of unlabeled data and few labeled samples, thus reducing the workload of data annotation to some extent. There have been several works in this direction. For example, in [41], a Gaussian process algorithm is introduced for semisupervised change detection. [42] presents a metric learning method with Laplacian regularization by which labels can be propagated from labeled samples to unlabeled instances. Most recently, the authors of [43] propose a semisupervised change detection approach, where bitemporal images are first encoded as a graph via a multiscale parcel segmentation algorithm, and a graph convolutional network (GCN) is used for semisupervised node classification on this graph. In [44], the authors propose to extract multiscale features and learn discriminant metrics by highlighting the contribution of simple training samples. As a popular architecture in semisupervised change detection, GAN is used to generate more training samples [45] and enhance the feature distribution consistency of final predictions [46].

C. Unsupervised Change Detection

Most change detection models are unsupervised, as collecting labeled data is difficult. In the literature, a widely used methodology is based on image algebra, and the classical models are change vector analysis (CVA) [8], [9] and its variations [10], [11]. Furthermore, some image-transformation-based unsupervised change detection approaches have been proposed to learn new feature representations from the original data, to highlight changes while suppressing no changes in the new feature spaces. For instance, [47] and [48] apply principal component analysis (PCA)—a well-known subspace learning algorithm—on stacked images and difference images, respectively, for unsupervised change detection tasks. In [49], the authors introduce a multivariate statistical transformation, termed multivariate alteration detection (MAD), to identify land cover changes in satellite images, and this method is invariant to the linear scaling of the input data. The authors of [50] propose to make use of slow feature analysis (SFA), which is capable of learning slowly varying features from a time series, in change detection. By doing so, in the learned feature space, differences among no-change pixel pairs are suppressed so that changed instances stand out more clearly. Later, deep SFA, a combination of networks and SFA, is proposed in [51]. In [52], self-supervised multitemporal segmentation is used for unsupervised change detection. In addition, some networks, e.g., autoencoders [53], [54], [55], [56], [57], deep belief networks (DBNs) [58], and CNNs [19], [20], [59], [60], [61], [62], are also used to learn better feature representations in an unsupervised fashion for unsupervised change detection. The authors of [63] incorporate deep neural networks and low-rank decomposition for predicting saliency maps where high values indicate large change probabilities. Another important branch in unsupervised change detection is heterogeneous change detection, also known as multimodal change detection, which aims to detect the changes between heterogeneous images. In [64], the authors propose to build up an observation field from paired pixels in bitemporal heterogeneous images for modeling features that are invariant to diverse imaging modalities. In [65], the authors train a stacked sparse autoencoder for heterogeneous change detection. [66] measures the similarity of affinity matrices computed from bitemporal heterogeneous images to identify pixels being likely to be unchanged, which are then used as pseudo-training data to learn transformations between different modalities. Following this work, [67] introduces vertex degrees and cycle consistency in learning such transformations. Aiming to project features of multimodal images into a shared space, [68] proposes an iterative coupled dictionary learning model to establish coupled dictionaries for multimodal images, and [69] takes variational autoencoder (VAE) as the backbone and builds coupled GANs for heterogenous change detection.

III. METHODOLOGY

A. Autoencoders in One-Class Change Detection and Their Limitations

A conventional autoencoder usually includes two major components: an encoder and a decoder. The former attempts

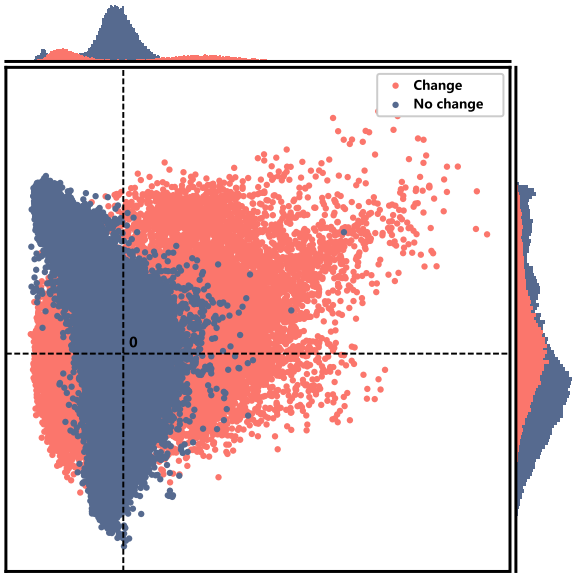


Fig. 1. Visualization of the latent space of an autoencoder for one-class change detection, obtained by projecting to the first two principal components. Blue and red points indicate no-change and changed samples in the test set, respectively. As can be seen, a considerable number of changed samples share the same feature space with no-change ones, demonstrating the overgeneralization problem of the autoencoder.

to represent the input in a latent space, and the latter learns to reconstruct the input from the encoded representation. Let us assume that we have pre- and post-change images acquired over the same geographical area. We stack them together and use them as the input to an autoencoder. Note that only pairs of no-change pixels are used in the training phase. Formally, let $\mathcal{X}$ be the domain of labeled no-change samples and $\mathcal{Z}$ represent the domain of encodings. Thus, the encoder and the decoder can be denoted as $f(\cdot) : \mathcal{X} \to \mathcal{Z}$ and $g(\cdot) : \mathcal{Z} \to \mathcal{X}$, respectively. Given a no-change pixel pair $\boldsymbol{x}_i \in \mathcal{X}$, we have

$$\boldsymbol{z}_i = f(\boldsymbol{x}_i; \boldsymbol{\Theta}_e) \tag{1}$$

$$\hat{\boldsymbol{x}}_i = g(\boldsymbol{z}_i; \boldsymbol{\Theta}_d) \tag{2}$$

where $\boldsymbol{z} \in \mathcal{Z}$, $\boldsymbol{\Theta}_e$ and $\boldsymbol{\Theta}_d$ are the parameters of the encoder and the decoder, respectively, and $\hat{\boldsymbol{x}}_i$ denotes the reconstruction. Training the autoencoder means optimizing the parameters $\{\boldsymbol{\Theta}_e, \boldsymbol{\Theta}_d\}$ to reduce the average loss on the training samples. In the context of one-class change detection, an autoencoder makes use of the reconstruction error as a change score. This method assumes that in the inference phase, no-change samples can be reconstructed more accurately than changed ones.

In theory, the autoencoder architecture is an ideal model for the one-class change detection tasks. However, in real-world scenarios, we observe that it often suffers from the problem of overgeneralization, i.e., sometimes an autoencoder trained on no-change examples generalizes so well that it is able to reconstruct some changed instances perfectly in the inference phase. Fig. 1 shows an example visualization of the learned latent space of an autoencoder for one-class change detection. This is obtained by projecting to the first two principal components, and red and blue indicate change and no change, respectively. As can be seen, a considerable number of changed samples share the same feature space with no-change instances, demonstrating the overgeneralization problem of the autoencoder. There are two possible reasons: 1) some changed pixel pairs share common components (e.g., spectral bands) with no-change ones and 2) the encodings of changed examples have high intraclass variance, which leads to difficulty in distinguishing changes from no changes.

B. Data-Enclosing-Ball Minimizing Autoencoder

Our insight is that reducing the intraclass distance of training data (i.e., labeled no-change samples) in the latent space can help reduce the overgeneralization problem of the autoencoder in the one-class change detection tasks. In this work, a new autoencoder architecture is proposed with dual objectives—a traditional reconstruction error criterion and a minimum volume criterion that minimizes the volume of the latent space enclosing encoded representations of training samples. Fig. 2 shows the overview architecture of the proposed DebM-AE.

1) Reconstruction Error Criterion: In this article, we make use of the ℓ_2-norm-based mean square error (MSE), that is,

$$\mathcal{L}_{\text{rec}}(\boldsymbol{\Theta}_e, \boldsymbol{\Theta}_d) = \sum_{i=1}^{N} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|_2^2 \tag{3}$$

as the reconstruction loss, where N is the number of training samples. The proposed model updates the encoder and the decoder to minimize the reconstruction errors of the inputs.

2) Minimum Volume Criterion: Our aim is to jointly learn the encoder parameters $\boldsymbol{\Theta}_e$ together with minimizing the volume of the enclosing space of training samples in the latent space $\mathcal{Z}$. Here, we make use of a hypersphere as the enclosing space, as it is simple, effective, and easy to implement. Let $\boldsymbol{c}$ and R be the center and radius of the hypersphere, respectively. Thus, the minimum volume criterion can be defined as follows:

$$\mathcal{L}_{\text{mv}}(R, \boldsymbol{c}, \boldsymbol{\Theta}_e) = R^2 + \frac{1}{\mu N} \sum_{i=1}^{N} \max\{0, \|\boldsymbol{z}_i - \boldsymbol{c}\|^2 - R^2\}. \tag{4}$$

Minimizing the first term of this objective, R^2, minimizes the volume of the hypersphere. The minimization of the second term is a penalty that encodings outside the sphere ($\|\boldsymbol{z}_i - \boldsymbol{c}\|^2 > R^2$) get penalized. This makes the hypersphere include as much training data as possible. The hyperparameter $\mu \in (0, 1]$ controls the trade-off between the volume and boundary violations. By optimizing (4), the encoder learns to closely map no-change pixel pairs to the hypersphere center $\boldsymbol{c}$. Through this, we can learn a compact latent space for no change. Note that the more compact the latent space, the better the robustness and transferability of our model. After sufficient training, changed samples can be better separated, as they are supposed to be further away from $\boldsymbol{c}$ or outside the sphere.

Furthermore, we consider a special case where the hypersphere becomes a point ($R = 0$). In this case, (4) can be

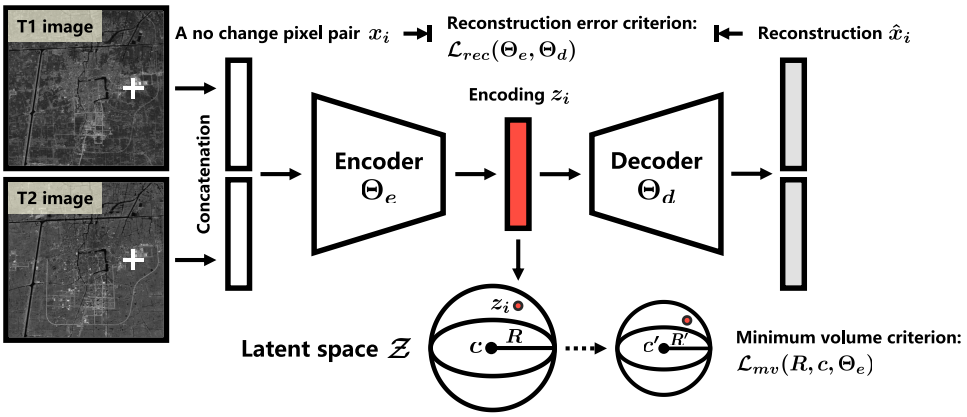


Fig. 2. Architecture of a DebM-AE. The top row shows a standard autoencoder that reconstructs a no-change pixel pair $\boldsymbol{x}_i$ from a latent encoding $\boldsymbol{z}_i$ based on the reconstruction error criterion. The bottom row shows our minimum volume criterion that minimizes the volume of the space enclosing training examples in the latent space. In the inference phase, a sample whose reconstruction loss is greater than a threshold is predicted as a changed instance.

simplified and rewritten as follows:

$$\mathcal{L}_{\text{mv}} = \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{z}_i - \boldsymbol{c}\|^2. \tag{5}$$

i.e., the minimum volume criterion is converted into a quadratic loss which penalizes the distance of each encoded representation to the hypersphere center.

C. Learning and Inference

We jointly optimize the parameters of the proposed network $\{\boldsymbol{\Theta}_e, \boldsymbol{\Theta}_d, R, \boldsymbol{c}\}$ in an end-to-end fashion by making use of the following loss function:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda \mathcal{L}_{\text{mv}} \tag{6}$$

where λ is a hyperparameter to introduce a weight on the minimum volume criterion and can model the relative importance of the two criteria.

In the inference phase, we use the trained network to detect changes by determining how well the network can reconstruct the input pixel pairs. More specifically, we

1) calculate MSE loss on training samples and obtain the mean (mean) and standard deviation (std) of loss values;
2) take mean $+ 2 *$ std as the threshold [70], [71], [72];
3) predict a sample whose reconstruction loss is greater than the threshold as a changed instance.

IV. Dataset and Experimental Setup

A. Data Description

We evaluate the proposed network on three datasets. All of them are acquired by the Landsat Enhanced Thematic Mapper Plus (ETM+) sensor onboard the Landsat satellite. Each image in the datasets has a spatial resolution of 30 m/pixel and six spectral bands. They are preprocessed (e.g., coregistration and calibration).

1) Kunshan Data: The first dataset consists of two images captured in March 2000 and February 2003, respectively, and covers the city of Kunshan, China. It is with a WGS-84 projection and a coordinate range of 32°26′09N–32°32′28N, 119°50′32E–119°58′24E. The spatial size of the two images is 800 × 800 pixels. The main types of changes in this dataset involve city expansion and farmland changes. Fig. 3 shows the two images as well as training and test set maps.

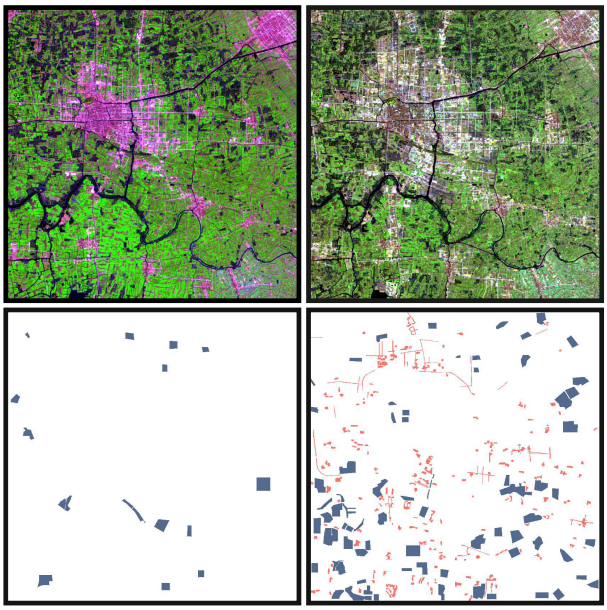


Fig. 3. (Top Row) Bitemporal images in the Kunshan dataset. (Bottom Row) Polygon-level training (left) and test (right) set maps where **blue** and **red** pixels indicate no-change and changed pixel pairs, respectively.

2) Lake Eppalock Data: The second dataset covers the Lake Eppalock in North Central Victoria, Australia. The bitemporal images of the dataset are captured in February 1991 and March 2009 with a coordinate range of 36°49′10S–37°00′52S, 144°27′52E–144°37′35E. The spatial size of both the images is 602 × 631 pixels. Several types of changes in natural resources, e.g., water, are presented in this dataset. Fig. 4 shows the two images, a training set map and a test set map for evaluating the change detection methods.

3) Taizhou Data: The third dataset (cf. Fig. 5) includes two preprocessed images covering the city of Taizhou, China. The two images are acquired in March 2000 and February 2003 with a WGS-84 projection and a coordinate range of 31°14′56N–31°27′39N, 120°02′24E–121°07′45E, and each of

TABLE I
NUMBERS OF TRAINING AND TEST SAMPLES IN THE KUNSHAN, LAKE EPPALOCK, TAIZHOU, AND BEIRUT DATASETS

data sets.

Data Set	Training		Test				Total			
	No change		Change		No change		Change		No change	
	# Polygons	# Pixels	# Polygons	# Pixels	# Polygons	# Pixels	# Polygons	# Pixels	# Polygons	# Pixels
Kunshan	15	8224	214	17327	69	40777	214	17327	84	49001
Lake Eppalock	15	1423	32	3880	30	3592	32	3880	45	5303
Taizhou	10	3385	61	4555	50	14076	61	4555	60	17461
Beirut*	18	76442	-	33918	-	1152240	-	33918	-	1228682

* indicates that all pixels outside training polygons are used to test the model performance.

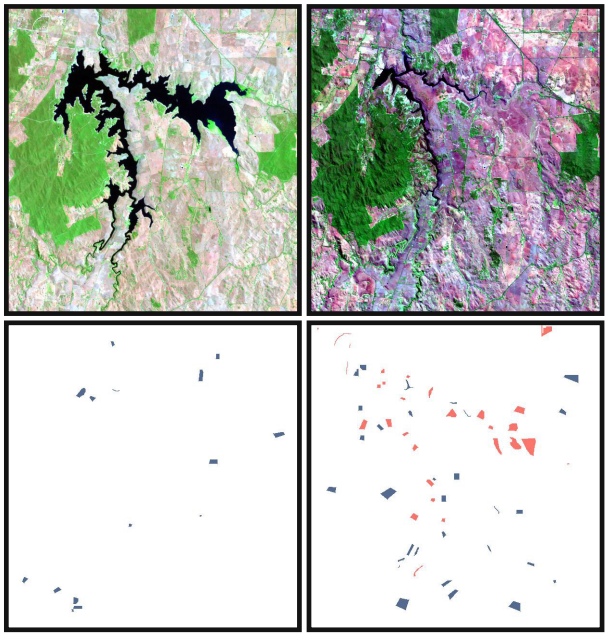


Fig. 4. (Top Row) Bitemporal images in the Lake Eppalock dataset. (Bottom Row) Polygon-level training (left) and test (right) set maps where **blue** and **red** pixels indicate no-change and changed pixel pairs, respectively.

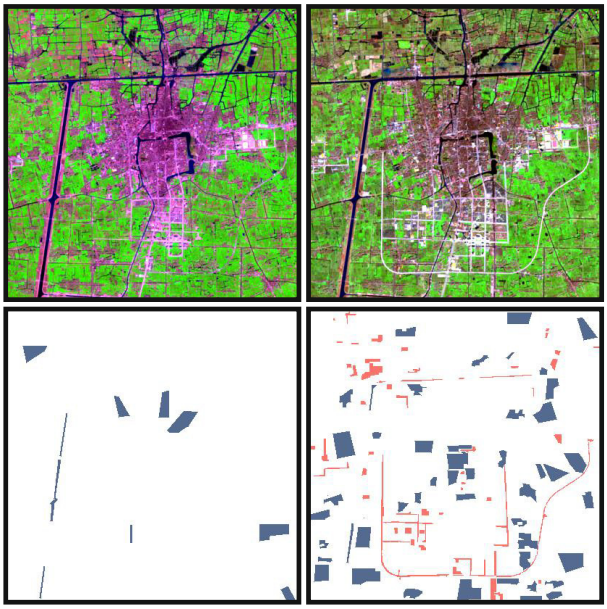


Fig. 5. (Top Row) Bitemporal images in the Taizhou dataset. (Bottom Row) Polygon-level training (left) and test (right) set maps where **blue** and **red** pixels indicate no-change and changed pixel pairs, respectively.

them has 400 × 400 pixels. The main changes shown in this dataset are city expansion, changes in water bodies, and soil changes.

4) Beirut Data: The forth dataset (see Fig. 6) is reproduced from the ONERA Satellite Change Detection (OSCD) dataset [73]. To accord with other datasets, we select bitemporal images of one city, Beirut, which is very challenging due to its large scale and various land covers. Specifically, the images were collected from Sentinel-2 satellites in August 2015 and October 2017, and all 13 bands are available, leading to variant spatial resolutions (10–60 m/pixel). Images of lower resolutions are upsampled to ensure that image sizes with respect to diverse bands are identical, i.e., 1070 × 1180 pixels. In the Beirut dataset, all the pixels are automatically classified into change or no change by resorting to OpenStreetMap data, and thus, changes mainly occur in buildings and roads.

Table I shows the numbers of training and test instances in the three datasets.

B. Implementation

As to the implementation, the encoder consists of two fully connected layers with 128 and 512 units, respectively, and yields a 512-D latent representation z_i. The decoder reconstructs $\hat{\boldsymbol{x}}_i$ via two fully connected layers, and the numbers of their units are 128 and C. Here, C denotes the number of channels of the input pixel pairs. The outputs of the first three layers are activated with rectified linear units (ReLUs), while the last layer is linearly activated. Table II shows the architecture of the used autoencoder. $\boldsymbol{c}$ is defined as a 512-D learnable vector and initialized with values sampled from a normal distribution in the training phase, and the initial R is set to 0. All the weights of the fully connected layers are initialized with a Glorot uniform initializer. Besides, we set μ as 0.01 empirically, and λ is set as 1. The learning rate is fixed as $1e-4$, and the size of minibatches is 32. The networks are implemented with TensorFlow and trained on one NVIDIA Tesla P100 16-GB GPU. All the network weights are trainable and optimized by Nestrov Adam [74] using $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e-8$ as recommended.

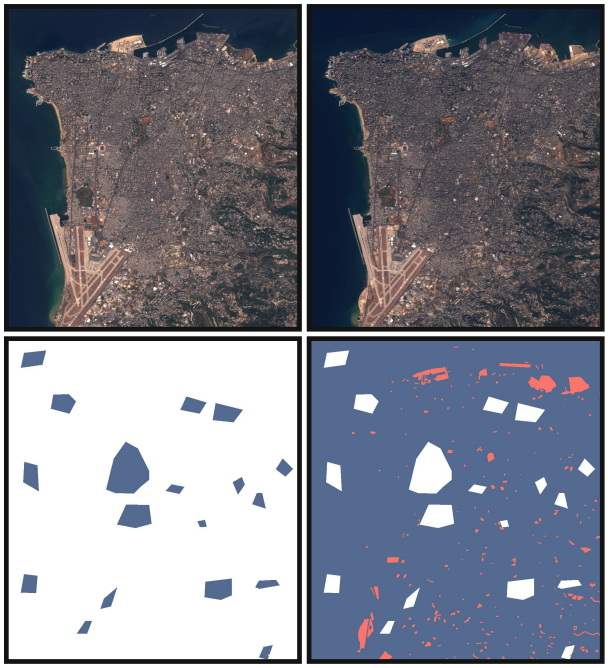


Fig. 6. (Top Row) Bitemporal images in the Beirut dataset. (Bottom Row) Polygon-level training (left) and test (right) set maps where **blue** and **red** pixels indicate no-change and changed pixel pairs, respectively.

TABLE II
ARCHITECTURE OF THE AUTOENCODER

Stage	Layer	Activation	Output size
Input	-	-	$N \times C$
Encoder	Fully-connected, 128 units	ReLU	$N \times 128$
	Fully-connected, 512 units	ReLU	$N \times 512$
Decoder	Fully-connected, 128 units	ReLU	$N \times 128$
	Fully-connected, C units	-	$N \times C$

N denotes the size of mini-batch.
C represents the channel dimension of each pixel.

C. Evaluation Metrics and Competitors

To quantitatively evaluate the performance of the proposed one-class change detection model and other competitors, we make use of the following metrics.

1) *Kappa Coefficient:* Kappa is a statistical measurement to measure the agreement between a predicted change detection map and the ground truth. In general, it is thought to be a more robust indicator than a simple percentage agreement calculation, as the possibility of a random agreement is taken into account. Note that several works in remote sensing point out issues with using Kappa coefficient as a performance metric [75], [76], [77], [78]. Since it is widely used in the change detection tasks, in this article, we still make use of it as one of the metrics.
2) *Overall accuracy (OA):* This metric indicates the number of samples that are correctly identified, divided by the number of test instances.
3) *Precision:* This index calculates the ability of a model to not label a true negative observation as positive.
4) *Recall:* Recall, a.k.a., sensitivity, calculates the capability of a model to find positive observations in the test set.
5) *F1 score:* We use F1 score to compare any two models. It is difficult to compare two models with low precision and high recall or vice versa. F1 score can help measure recall and precision at the same time. It exploits harmonic mean in place of arithmetic mean by punishing extreme values more.
6) *False alarm rate (FAR):* FAR, also known as false positive ratio, indicates the probability of wrongly detected negatives, e.g., no-change pixels in our case.
7) *Missed detection rate (MDR):* MDR measures the probability that a model fails to infer true positive observations.

Note that we compute F1 score, precision, and recall for both change and no change classes.

Considering that the proposed method is for one-class change detection, we take the one-class classification algorithms as the main competing methods, to evaluate the effectiveness of our approach. To verify the benefit of using labeled no-change samples, we also include unsupervised change detection models in comparisons. Besides, given that the outlier detection approaches can be applied to the one-class change detection tasks, we compare our network with some outlier detection algorithms. All the methods included in experiments are summarized as follows.

1) CVA [9], a classical image-algebra-based unsupervised model for change detection in multi and hyperspectral images.
2) PCA [48], [79], a transformation-based unsupervised approach having a fast computing speed.
3) MAD [49], a classical image-transformation-based unsupervised change detection method.[1]
4) IRMAD [80], iteratively reweighted MAD, which is an extension of MAD by introducing an iterative scheme.[2]
5) SFA [50], an SFA-based unsupervised change detection method.[3]
6) DSFA [51], a network based on SFA for unsupervised change detection.[4]
7) KPCAMNet [81], a Siamese deep network built on weight-shared kernel PCA convolutions. The input of each branch is an image patch, and the outputs are clustered into change and no change in an unsupervised manner. Note that compared with other competing methods, this model uses spatial information (patches).[5]
8) DSMSCN [82], a Siamese multiscale convolutional network for change detection. Following the setting in [82], we use CAV to generate pseudo labels.[6]

[1] http://www.imm.dtu.dk/~alan/software.html
[2] http://www.imm.dtu.dk/~alan/software.html
[3] https://github.com/I-Hope-Peace/ChangeDetectionRepository/tree/master/Methodology/Traditional
[4] https://github.com/rulixiang/DSFANet
[5] https://github.com/I-Hope-Peace/KPCAMNet
[6] https://github.com/I-Hope-Peace/DSMSCN

9) cGAN [83], a conditional adversarial network for unsupervised change detection.[7]
10) iForest [36], isolation forest, is an algorithm for outlier detection. In our case, it builds an ensemble of isolation trees for a given dataset, and then changes are the samples that have shorter average path lengths on the trees.[8]
11) EE [37], elliptic envelope, is an outlier detection model that fits a robust covariance estimate to the data and thus fits an ellipse to central data samples.[9]
12) LOF [38], local outlier factor, a density-based algorithm to identify outliers/novelties in a dataset. It can be used in a dataset, which has a mixture of data distributions, and is applicable to our one-class change detection tasks. LOF can be used as both an unsupervised outlier detection model and a one-class classifier, and in our experiments, we choose the latter.[10]
13) SVDD [84], support vector data description, a one-class classification model trying to find the smallest hypersphere that contains all no-change samples from the training set.[11]
14) OC-SVM [15], a one-class SVM that separates no-change instances from changed ones by finding a hyperplane of maximal distance from the origin.[12]
15) Deep SVDD [85], a network using the SVDD loss. We make the structure of the used Deep SVDD model consistent with that of the encoder of our DebM-AE for a fair comparison.[13]
16) VAE [86], VAE that imposes a probabilistic distribution, e.g., Gaussian, on the latent space and penalizes the difference between the prior and posterior using Kullback–Leibler divergence. The configuration of the used VAE is the same as that of our DebM-AE described in Section IV-B. Note that z_{mean} and z_{std} are separate fully connected layers.
17) AE, an autoencoder whose configuration is the same as above.
18) DebM-VAE, the proposed network implemented on VAE.
19) DebM-AE, the proposed network implemented on AE.

V. Results

In this section, the hyperparameters are evaluated in Section V-A, and the effectiveness of the minimum volume criterion is analyzed in Section V-B. Afterward, we compare our model with the aforementioned competitors in Section V-C, and Section V-D validates the transferability of the proposed model. Section V-E explores the feasibility of integrating spatial information into our model, and Section V-F shows the effect of the number of polygon annotations.

[7]https://github.com/llu025/Heterogeneous_CD/tree/master/Code-Aligned_Autoencoders

[8]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html#sklearn.ensemble.IsolationForest

[9]https://scikit-learn.org/stable/modules/generated/sklearn.covariance.EllipticEnvelope.html#sklearn.covariance.EllipticEnvelope

[10]https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html#sklearn.neighbors.LocalOutlierFactor

[11]https://ww2.mathworks.cn/matlabcentral/fileexchange/69296-support-vector-data-description-svdd

[12]https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[13]https://github.com/lukasruff/Deep-SVDD-PyTorch

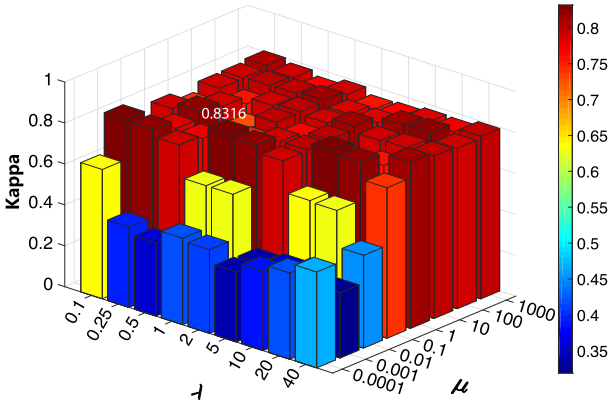


Fig. 7. Evaluation of hyperparameters μ and λ. The highest Kappa (0.8316) is achieved when μ and λ are set to 0.01 and 1, respectively.

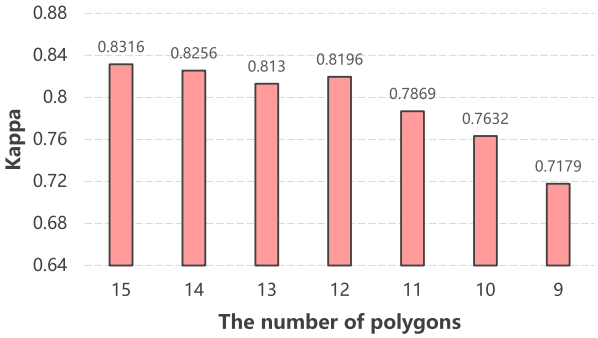


Fig. 8. Effect of the number of labeled polygons in the training set.

A. Evaluation of Hyperparameters

We evaluate the following hyperparameters: μ in (4) and λ in (6). The former is a hyperparameter that controls the relative importance of the volume of the data-enclosing-ball and boundary violations. It actually allows us to control the ratio of outliers in our model during the training phase. The latter, λ, is a trade-off hyperparameter between the two objectives. Fig. 7 shows that the Kappa coefficient increases initially as μ becomes larger, but slightly decreases after $\mu = 0.01$, and λ of large values can lead to a decrease in the network performance. In our experiments, we use $\mu = 0.01$ and $\lambda = 1$.

B. Ablation Study on the Minimum Volume Criterion

We report the numerical results in Tables III–VI. The observed performance increase from VAE/AE to DebM-VAE/DebM-AE is a central point of our article. Take AE and DebM-AE as an example. The increases in Kappa are 0.0515, 0.2981, 0.0892, and 0.0727 on the Kunshan, Lake Eppalock, Taizhou, and Beirut datasets, respectively, which represents a significant gain. The fact that DebM-VAE/DebM-AE outperforms VAE/AE can be attributed to learning a more compact data-enclosing latent space for no-change pixel pairs.

TABLE III

NUMERICAL RESULTS FOR THE EVALUATED MODELS ON THE KUNSHAN DATASET. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**

Model	Kappa	OA	Mean F1	Change					No change		
				F1 score	Precision	Recall	FAR	MDR	F1 score	Precision	Recall
CVA	0.4764	78.23	73.82	63.07	63.83	62.32	15.00	37.68	84.57	84.15	85.00
PCA	0.5646	84.49	77.48	64.91	99.71	48.12	0.06	51.88	90.04	81.93	99.94
MAD	0.5513	83.25	77.19	65.42	85.12	53.12	3.95	46.88	88.95	82.82	96.05
IRMAD	0.5675	83.98	77.95	66.42	88.59	53.13	2.91	46.86	89.48	82.98	97.09
SFA	0.6058	83.72	80.28	72.06	73.79	70.40	10.62	29.60	88.51	87.66	89.38
DSFA	0.5660	78.51	78.06	81.19	74.41	89.33	33.18	10.67	74.93	85.29	66.82
KPCAMNet	0.8265	**92.76**	91.32	87.80	88.20	87.40	4.97	12.60	**94.85**	94.67	95.03
DSMSCN	0.6391	86.56	81.64	72.13	94.50	58.32	1.44	41.68	91.14	84.77	98.56
cGAN	0.6619	88.55	82.85	72.95	95.27	59.10	1.04	40.90	92.74	87.25	98.96
iForest	0.4763	77.80	73.81	63.59	62.23	65.01	16.77	34.99	84.03	84.85	83.23
EE	0.5025	82.45	74.15	59.51	95.34	43.25	0.90	56.75	88.79	80.43	99.10
LOF	0.6314	82.24	81.09	76.42	63.25	96.53	23.83	3.47	85.75	98.10	76.17
SVDD	0.7624	89.46	88.07	84.00	76.70	92.84	11.98	7.16	92.14	96.66	88.02
OC-SVM	0.2097	56.74	56.43	52.76	39.12	81.00	53.57	19.00	60.10	85.19	46.43
Deep SVDD	0.7607	90.11	88.04	83.05	84.96	81.22	6.11	18.78	93.02	92.17	93.89
VAE	0.6507	86.19	82.46	74.37	83.31	67.16	5.72	32.84	90.55	87.11	94.28
AE	0.7801	90.75	89.01	84.62	83.96	85.29	6.93	14.71	93.39	93.71	93.07
DebM-VAE	0.8197	90.94	90.92	90.50	99.29	83.14	6.40	16.85	91.34	84.51	99.36
DebM-AE	**0.8316**	92.74	**91.57**	**88.44**	84.16	93.18	7.46	6.82	94.70	96.97	92.54

Besides quantitative results, we also visualize the learned latent spaces of VAE, AE, DebM-VAE, and DebM-AE. The right two images of Fig. 9 show latent spaces of the test data resulting from the proposed methodology (DebM-VAE and DebM-AE) trained on the Taizhou dataset. The learned manifolds exhibit very compact distributions for no-change samples, which in practice means better discrimination between change and no change categories. In contrast, the left column of Fig. 9 shows the learned feature spaces of VAE and AE with the same network configurations used in DebM-VAE and DebM-AE in experiments. We can see that distributions of changed and no-change instances severely overlap. Furthermore, Fig. 10 shows the evolution of the visualization of training instances on the Kunshan dataset in the latent space over training epochs.

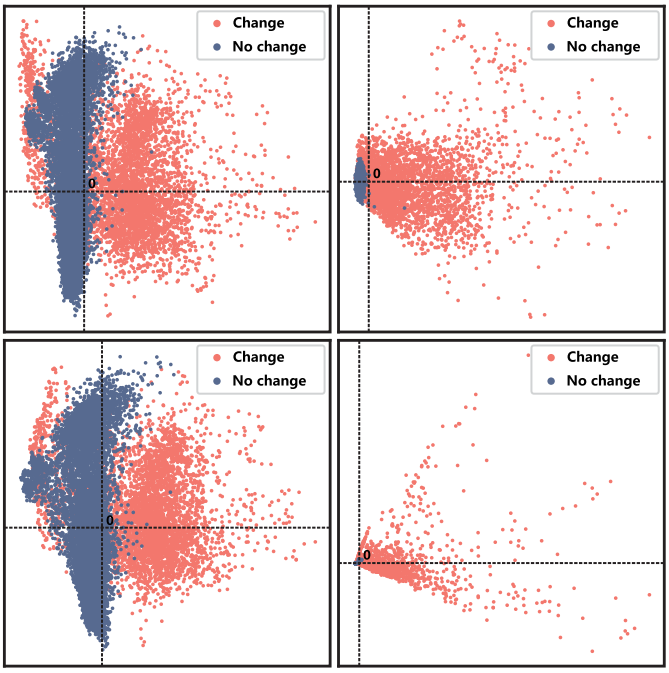


Fig. 9. Comparison of the learned latent spaces of test samples resulting from VAE (top left), AE (bottom left), DebM-VAE (top right), and DebM-AE (bottom right) trained on the Taizhou dataset. Nte that there are some overlaps between blue and red dots due to overplotting.

C. Comparison With Other Competitors

It can be seen from Tables III–VI that the unsupervised change detection methods based on deep networks outperform other conventional models on the four datasets, due to their superior ability of feature learning. On the other hand, among one-class classification methods (LOF, SVDD, OC-SVM, and Deep SVDD), deep SVDD performs relatively well considering the four datasets together. Deep SVDD shows comparative performance on the Kunshan and Lake Eppalock datasets and achieves higher Kappa and mean F1 on the Taizhou dataset compared with SVDD owing to its exploitation of deep neural networks. Furthermore, more interestingly, the one-class classifiers that use partial labeled data (i.e., labeled no-change pixel pairs for training) perform better than most unsupervised change detection approaches on the Kunshan

TABLE IV

NUMERICAL RESULTS FOR THE EVALUATED MODELS ON THE LAKE EPPALOCK DATASET. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**

Model	Kappa	OA	Mean F1	Change					No change		
				F1 score	Precision	Recall	FAR	MDR	F1 score	Precision	Recall
CVA	0.6099	80.77	79.95	84.00	73.95	97.22	37.00	2.78	75.90	**95.44**	63.00
PCA	0.5353	76.98	76.53	79.78	73.34	87.47	34.35	12.53	73.28	82.91	65.65
MAD	0.7850	89.19	89.15	88.52	98.67	80.26	1.17	19.74	89.78	82.25	98.83
IRMAD	0.7890	89.39	89.35	88.70	99.23	80.18	0.67	19.82	90.00	82.27	99.33
SFA	0.8096	90.50	90.48	**90.87**	90.68	91.06	10.11	8.94	90.09	90.30	89.89
DSFA	0.7086	88.00	85.43	79.30	81.67	77.06	7.35	22.94	**91.55**	90.48	92.65
KPCAMNet	0.7889	89.41	89.41	89.18	94.99	84.05	4.79	15.95	89.63	84.67	95.2
DSMSCN	0.6271	81.62	80.80	84.78	74.37	98.58	36.69	1.42	76.81	97.64	63.31
cGAN	0.5708	78.54	78.45	77.09	72.14	82.76	24.73	17.24	79.82	84.95	75.27
iForest	0.5274	76.66	75.90	80.19	71.69	90.98	38.81	9.02	71.60	86.26	61.19
EE	0.0380	50.12	36.72	7.59	100	3.94	0	96.06	65.84	49.08	100
LOF	0.2008	61.24	53.47	72.49	57.40	98.32	78.81	1.68	34.45	92.13	21.19
SVDD	0.4641	73.43	72.97	76.52	70.72	83.35	37.28	16.65	69.42	77.72	62.72
OC-SVM	0.5548	77.93	77.55	80.48	74.41	87.63	32.54	12.37	74.61	83.47	67.46
Deep SVDD	0.4751	74.20	72.65	79.17	68.16	94.41	47.63	5.59	66.12	89.66	52.37
VAE	0.4604	72.50	70.92	64.14	99.30	47.37	0.36	52.63	77.69	63.67	99.64
AE	0.5245	75.84	74.95	70.20	97.66	54.79	1.42	45.21	79.69	66.87	98.58
DebM-VAE	0.8213	91.02	89.09	90.60	99.29	83.30	0.64	16.70	91.41	84.63	99.40
DebM-AE	**0.8226**	**91.07**	**91.13**	90.69	99.11	83.58	0.81	16.42	91.45	84.83	99.19

TABLE V

NUMERICAL RESULTS FOR THE EVALUATED MODELS ON THE TAIZHOU DATASET. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**

Model	Kappa	OA	Mean F1	Change					No change		
				F1 score	Precision	Recall	FAR	MDR	F1 score	Precision	Recall
CVA	0.0793	65.71	53.97	30.71	30.35	31.09	23.09	68.91	77.22	77.52	76.91
PCA	0.8103	93.59	90.46	85.00	99.24	74.34	0.19	25.62	95.92	92.32	99.82
MAD	0.8179	93.67	90.88	85.82	94.92	78.31	1.36	21.71	95.93	93.36	98.64
IRMAD	0.8266	94.05	91.30	86.40	97.83	77.37	0.55	22.63	96.19	93.14	99.45
SFA	0.8245	93.77	91.21	86.47	92.26	81.36	2.21	18.64	95.96	94.19	97.79
DSFA	0.8369	94.12	91.84	87.53	90.86	84.43	2.75	15.57	96.15	95.08	97.25
KPCAMNet	0.7595	87.91	87.89	87.37	95.51	80.52	4.09	19.48	88.41	82.00	95.91
DSMSCN	0.6383	87.88	81.79	71.26	84.79	61.45	3.57	38.55	92.32	88.55	96.43
cGAN	0.6052	87.37	80.25	68.39	70.93	66.04	7.06	33.96	92.11	91.30	92.94
iForest	0.7501	90.05	87.46	81.75	74.13	91.11	10.29	8.89	93.16	96.89	89.71
EE	0.4116	83.29	69.06	48.08	99.93	31.66	0.01	68.34	90.04	81.89	99.99
LOF	0.5914	81.61	78.99	71.57	57.52	94.69	22.63	5.31	86.41	97.83	77.37
SVDD	0.6810	86.21	83.77	77.47	64.50	96.95	17.27	3.05	90.07	98.82	82.74
OC-SVM	0.2576	58.01	57.32	51.89	36.04	92.62	53.19	7.38	62.75	95.15	46.81
Deep SVDD	0.7640	90.55	88.15	82.80	74.61	93.02	10.24	6.98	93.49	97.54	89.76
VAE	0.7637	91.47	88.19	81.95	84.91	79.19	4.55	20.81	94.42	93.41	95.45
AE	0.8027	92.51	90.13	85.28	82.05	88.78	6.29	11.22	94.97	96.27	93.71
DebM-VAE	0.8055	92.42	90.26	85.66	79.69	92.60	7.64	7.40	94.85	97.47	92.36
DebM-AE	**0.8919**	**95.92**	**94.59**	**91.91**	89.07	94.95	3.77	5.05	**97.27**	98.33	96.23

TABLE VI
NUMERICAL RESULTS FOR THE EVALUATED MODELS ON THE BEIRUT DATASET. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**

Model	Kappa	OA	Mean F1	Change					No change		
				F1 score	Precision	Recall	FAR	MDR	F1 score	Precision	Recall
CVA	0.1468	82.36	54.40	18.70	10.67	75.54	17.46	24.46	90.10	99.19	82.54
PCA	0.0955	75.07	49.69	13.96	7.69	75.27	24.93	24.73	85.42	99.10	75.07
MAD	0.1086	79.10	51.61	15.12	8.49	69.29	20.62	30.71	88.09	98.94	79.38
IRMAD	0.1097	79.17	51.68	15.23	8.55	69.67	20.57	30.33	88.12	98.96	79.43
SFA	0.1057	72.19	49.40	15.45	8.50	84.67	28.20	15.33	83.36	99.34	71.80
DSFA	0.1089	81.21	52.47	15.52	8.97	57.49	18.05	42.51	89.43	98.42	81.95
KPCAMNet	0.2316	89.47	60.57	26.81	16.94	64.27	9.75	35.73	94.33	98.79	90.25
DSMSCN	0.3051	95.70	65.25	32.72	30.85	34.84	2.42	65.16	97.78	97.98	97.58
cGAN	0.1729	84.45	56.25	21.13	12.23	77.54	15.36	22.46	91.37	99.27	84.64
iForest	0.0568	71.50	46.82	10.59	5.82	59.07	28.14	40.93	83.04	98.35	71.86
EE	0.0104	97.15	49.82	1.08	70.88	0.55	0.01	99.45	98.55	97.16	99.99
LOF	0.0355	48.23	36.32	8.78	4.62	87.17	52.92	12.83	63.86	99.20	47.08
Deep SVDD	0.2166	87.95	59.40	25.35	15.40	71.52	11.56	28.48	93.45	99.06	88.44
VAE	0.2003	90.04	59.14	23.60	15.12	53.82	8.90	46.18	94.67	98.53	91.10
AE	0.1518	85.95	55.77	19.23	11.51	58.52	13.25	41.48	92.30	98.61	86.75
DebM-VAE	**0.3578**	**96.22**	**67.89**	**37.72**	35.66	40.03	2.13	59.97	**98.05**	98.23	97.87
DebM-AE	0.2264	88.81	60.09	26.23	16.16	69.54	10.62	30.46	93.95	99.01	89.38

and Beirut datasets, but they do much worse on the Lake Eppalock and Taizhou scenes. This means that the one-class classification models have unstable behaviors with different datasets. In contrast, the autoencoder-based models, especially the proposed DebM-AE, demonstrate good robustness against different scenes and high accuracy. Table VI presents the experimental results on the Beirut dataset, where DebM-VAE achieves the highest Kappa and mean F1 score. It is interesting to observe that DebM-VAE outperforms DebM-AE, and VAE benefits more from the proposed minimum volume criterion, e.g., increments of 0.1575% and 8.75% in Kappa and mean F1, respectively. However, it can be seen that all the models gain high OA but relatively low Kappa and mean F1 on the dataset. The reasons could be 1) the dataset includes few changed samples and is highly imbalanced and 2) bands with lower resolutions deliver limited effective visual information. Fig. 11 shows change detection maps obtained by the proposed method and three representative competitors.

We also analyze the model complexity by measuring floating point operations (FLOPs) and computation time. Taking the Kunshan dataset as an example, compared with deep learning-based competitors, our method needs comparative FLOPs and time. Specifically, AE and VAE take $2.69e5$ and $4.04e5$ FLOPs, while DebM-AE and DebM-VAE have $2.71e5$ and $4.05e5$ FLOPs, respectively. Besides, it takes 1 and 2 s per epoch to train AE and Deep SVDD, respectively, while for DebM-AE, one epoch training requires 2 s. As to the traditional algorithms, SVDD and IRMAD need 72.4 and 6.1 s in total, respectively, which are faster than our method. But taking the model performance into consideration, the computational complexity of our model is acceptable.

D. Model Transferability

In this section, we discuss the transferability of the proposed method. Several experiments are conducted to verify the performance of a model trained on a dataset and tested on other unseen scenes. Since both Kunshan and Taizhou are relevant to urban changes, here, we use these two datasets. The experimental settings are as follows.

1) *Kunshan → Taizhou:* Training samples are from the training set of the Kunshan dataset (cf. the bottom left image in Fig. 3), and the trained model is evaluated on the test set of the Taizhou scene (see the bottom right image in Fig. 5). The model is marked with $K \rightarrow T$.
2) *Taizhou → Kunshan:* We train a model using the training set of the Taizhou dataset (see the bottom left image in Fig. 5) and test it on the test set of the Kunshan dataset (cf. the bottom right image in Fig. 3). In this case, the method is marked with $T \rightarrow K$.

Table VII reports the experimental results. Overall, the proposed method shows good transfer performance. For example, Kappa and mean F1 of DebM-AE$_{K\rightarrow T}$ are 0.8522% and 92.61%, respectively, which are second to those of DebM-AE (in Table V) but significantly better than those of other models trained on the Taizhou dataset.

Moreover, to explore further the transferability of the proposed model, we conduct experiments on the source and

TABLE VII
EVALUATING TRANSFERABILITY OF MODELS. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**

Model	Kappa	OA	Mean F1	Change					No change		
				F1 score	Precision	Recall	FAR	MDR	F1 score	Precision	Recall
Kunshan → Taizhou											
$VAE_{K\to T}$	0.7633	92.07	88.09	81.19	96.69	69.97	0.77	30.03	94.98	91.08	99.23
$AE_{K\to T}$	0.8182	93.73	90.88	85.78	96.16	77.43	1.00	22.57	95.97	93.13	99.00
DebM-$VAE_{K\to T}$	0.8214	93.81	91.04	86.06	95.78	78.13	1.12	21.87	96.02	93.32	98.88
DebM-$AE_{K\to T}$	**0.8522**	**94.56**	**92.61**	**88.81**	89.41	88.21	3.38	11.79	**96.41**	96.20	96.62
Taizhou → Kunshan											
$VAE_{T\to K}$	0.6438	86.25	82.04	73.34	86.98	63.39	4.03	36.61	90.74	86.05	95.97
$AE_{T\to K}$	0.6096	84.80	80.35	71.00	82.35	62.40	5.68	37.60	89.70	85.51	94.32
DebM-$VAE_{T\to K}$	0.6273	85.93	81.11	71.57	89.99	59.41	2.81	40.59	90.65	84.93	97.19
DebM-$AE_{T\to K}$	**0.7880**	**91.30**	**89.40**	**84.90**	87.96	82.05	4.77	17.95	**93.89**	92.58	95.23

TABLE VIII
EVALUATING PATCH- AND PIXEL-BASED AUTOENCODERS. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**

Model	Kappa	OA	Mean F1	Change					No change		
				F1 score	Precision	Recall	FAR	MDR	F1 score	Precision	Recall
DebM-AE	0.8316	92.74	91.57	88.44	84.16	93.18	7.46	6.82	94.70	96.97	92.54
DebM-CAE	**0.8479**	**93.55**	**92.39**	**89.42**	87.59	91.32	5.50	8.68	**95.37**	96.24	94.50

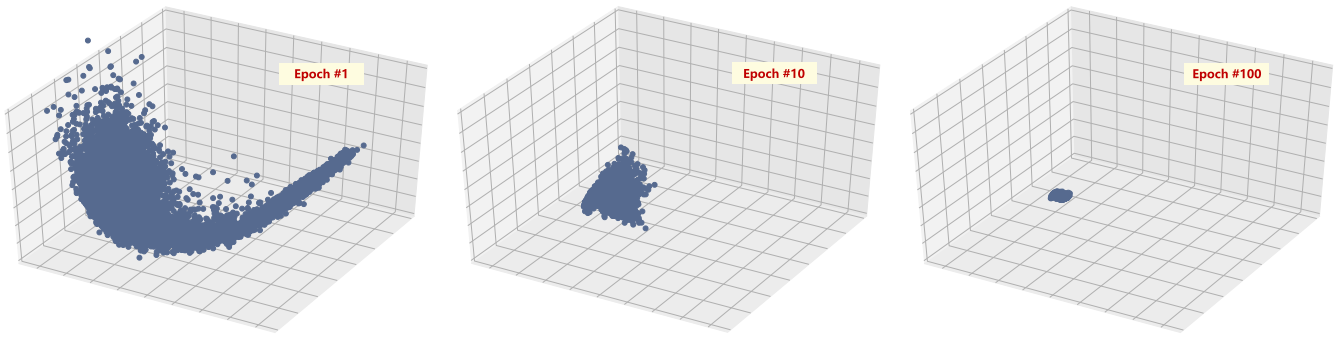


Fig. 10. Evolution of the visualization of the latent space while training a DebM-AE for one-class change detection.

target datasets having different scenes (e.g., Kunshan and Lake Eppalock). The experimental results are unsatisfactory. We speculate that the transferability of the model is limited by the scene consistency between two datasets, i.e., if the source and target datasets have the same scene type (e.g., *city* in the Kunshan and Taizhou datasets), the model's transferability is satisfactory; otherwise, the model does not work well.

E. Taking Into Account Spatial Information

We note that it is feasible to further boost the performance of our model by 1) taking image patches as input and 2) substituting convolutional layers for fully connected layers in the network architecture. Specifically, we crop patches with the original training pixels being at the center and generate patch pairs by concatenating patches from bitemporal images. The number of filters in each convolutional layer is identical to that in the corresponding fully connected layer. The experiments are conducted on the Kunshan dataset, and Table VIII shows that the use of spatial information can boost the change detection performance.

F. Effect of the Number of Polygon Annotations

In this section, we study the effect of the number of polygon annotations by training networks with variant subsets. Taking the Kunshan dataset (including 15 polygon annotations) as an example, we train our model on six subsets, which are produced by randomly remaining 9, 10, 11, 12, 13, and 14 polygon-level annotations, respectively. As shown in Fig. 8, the performance of DebM-AE gradually decreases with the decrement of the number of polygon annotations. Therefore, as a compromise between the classification accuracy and

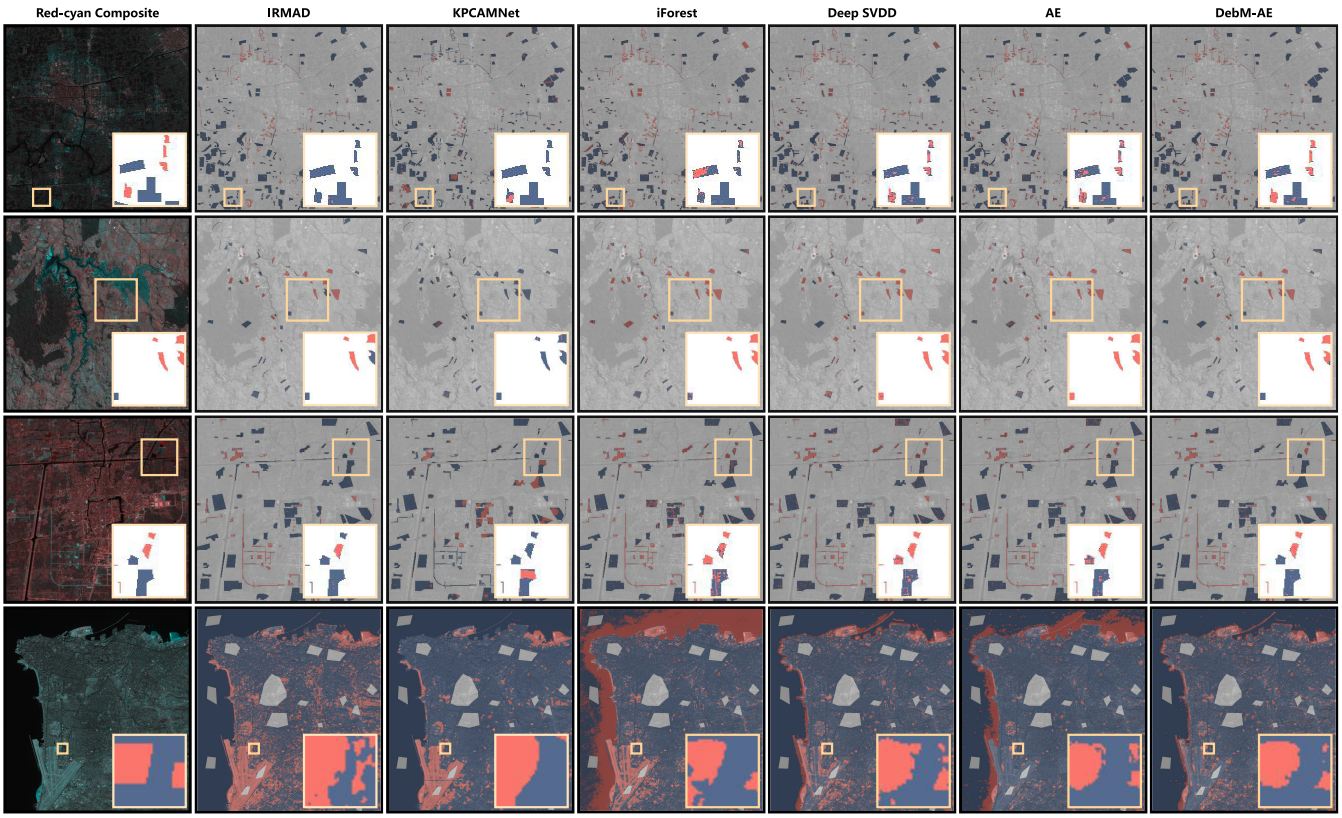


Fig. 11. Change detection maps obtained by our DebM-AE and five representative competitors—IRMAD, KPCAMNet, iForest, Deep SVDD, and AE. The first column shows red–cyan color composites of bitemporal images of the three datasets, where red is for T_1 image and cyan for is T_2 image. Best viewed zoomed in color.

human labor, the number of polygons is supposed to be no less than 12.

VI. Conclusion

In this article, we introduce a network for the one-class change detection tasks. The proposed DebM-AE learns to simultaneously reconstruct the input pairs of no-change pixels and minimize the volume of the latent space enclosing encoded representations of the training samples. By doing so, we encourage the model to learn a compact feature space for no-change data, which is conducive to recognizing changed examples in the test phase. Compared with the unsupervised change detection approaches, one-class classification models, and regular autoencoder networks, the proposed method gains significant improvements. Furthermore, our DebM-AE shows very good robustness and transferability. In future, we will extend the proposed method for change/anomaly detection in image time series. Incorporating spatial information into the proposed approach will also be studied, and we believe that the resulting model would lead to better results. In addition, applying the idea of one-class change detection and the proposed methodology to change detection in high spatial resolution images would be another interesting topic to investigate.

Acknowledgment

The authors would like to thank Prof. C. Wu and Prof. B. Du for providing the Kunshan and Taizhou datasets, and Dr. H. Lyu for providing the Lake Eppalock dataset.

References

[1] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.

[2] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens.*, vol. 12, no. 10, p. 1688, May 2020.

[3] A. Shapiro, C. Trettin, H. Küchly, S. Alavinapanah, and S. Bandeira, "The mangroves of the Zambezi Delta: Increase in extent observed via satellite from 1994 to 2013," *Remote Sens.*, vol. 7, no. 12, pp. 16504–16518, Dec. 2015.

[4] A. Koltunov and S. L. Ustin, "Early fire detection using non-linear multitemporal prediction of thermal imagery," *Remote Sens. Environ.*, vol. 110, no. 1, pp. 18–28, 2007.

[5] Y. Yuan *et al.*, "Continuous change detection and classification using hidden Markov model: A case study for monitoring urban encroachment onto farmland in Beijing," *Remote Sens.*, vol. 7, no. 11, pp. 15318–15339, Nov. 2015.

[6] N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention UNet for building segmentation in remote sensing images," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–12, Apr. 2020.

[7] J. Xie, N. He, L. Fang, and A. Plaza, "Scale-free convolutional neural network for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6916–6928, Sep. 2019.

[8] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with Landsat," in *Proc. LARS Symposia*, 1980. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.1459&rep=rep1&type=pdf

[9] 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. 2006.

[10] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, May 2012.

[11] 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.

[12] L. Auret and C. Aldrich, "Change point detection in time series data with random forests," *Control Eng. Pract.*, vol. 18, no. 8, pp. 990–1002, Aug. 2010.

[13] K. J. Wessels *et al.*, "Rapid land cover map updates using change detection and robust random forest classifiers," *Remote Sens.*, vol. 8, no. 11, p. 888, Oct. 2016.

[14] D. Seo, Y. Kim, Y. Eo, M. Lee, and W. Park, "Fusion of SAR and multispectral images using random forest regression for change detection," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 10, p. 401, 2018.

[15] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 1999, pp. 1–7.

[16] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observat. Geoinf.*, vol. 20, pp. 77–85, Feb. 2013.

[17] 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.

[18] 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.

[19] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[20] A. Song, J. Choi, Y. Han, and Y. Kim, "Change detection in hyperspectral images using recurrent 3D fully convolutional networks," *Remote Sens.*, vol. 10, no. 11, p. 1827, Nov. 2018.

[21] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep Siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.

[22] Z. Zhang, G. Vosselman, M. Gerke, D. Tuia, and M. Ying Yang, "Change detection between multimodal remote sensing data using Siamese CNN," 2018, *arXiv:1807.09562*.

[23] R. C. Daudt, B. L. Saux, A. Boulch, and Y. Gousseau, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Understand.*, vol. 187, Oct. 2019, Art. no. 102783.

[24] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++," *Remote Sens.*, vol. 11, no. 11, p. 1382, 2019.

[25] W. Zhao, L. Mou, J. Chen, Y. Bo, and W. J. Emery, "Incorporating metric learning and adversarial network for seasonal invariant change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2720–2731, Apr. 2020.

[26] Y. Gao, F. Gao, J. Dong, and H.-C. Li, "SAR image change detection based on multiscale capsule network," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 484–488, Mar. 2021.

[27] Q. Xu, K. Chen, G. Zhou, and X. Sun, "Change capsule network for optical remote sensing image change detection," *Remote Sens.*, vol. 13, no. 14, p. 2646, Jul. 2021.

[28] S. Saha, P. Ebel, and X. X. Zhu, "Self-supervised multisensor change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, 2022, doi: 10.1109/TGRS.2021.3109957.

[29] Y. Chen and L. Bruzzone, "Self-supervised change detection in multi-view remote sensing images," 2021, *arXiv:2103.05969*.

[30] 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, p. 1662, May 2020.

[31] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3095166.

[32] P. Li and H. Xu, "Land-cover change detection using one-class support vector machine," *Photogram. Eng. Remote Sens.*, vol. 76, no. 3, pp. 255–263, Mar. 2010.

[33] J. Mũnoz-Marí, F. Bovolo, L. Gomez-Chova, L. Bruzzone, and G. Camp-Valls, "Semisupervised one-class support vector machines for classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3188–3197, Aug. 2010.

[34] F. de Morsier, D. Tuia, M. Borgeaud, V. Gass, and J.-P. Thiran, "Semi-supervised novelty detection using SVM entire solution path," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1939–1950, Apr. 2013.

[35] S. Saha, F. Bovolo, and L. Bruzzone, "Change detection in image time-series using unsupervised LSTM," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[36] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2008, pp. 413–422.

[37] P. J. Rousseeuw and K. Van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[38] M. M. Breunig, H. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *ACM SIGMOD Int. Conf. Manag. Data*, 2000, pp. 93–104.

[39] H. R. Kerner *et al.*, "Comparison of novelty detection methods for multispectral images in rover-based planetary exploration missions," *Data Mining Knowl. Discovery*, vol. 34, no. 6, pp. 1642–1675, Nov. 2020.

[40] V. Böhm and U. Seljak, "Probabilistic auto-encoder," 2020, *arXiv:2006.05479*.

[41] K. Chen, C. Huo, Z. Zhou, H. Lu, and J. Cheng, "Semi-supervised change detection via Gaussian processes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2009, p. 996.

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

[43] S. Saha, L. Mou, X. Zhu, F. Bovolo, and L. Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 607–611, Apr. 2021.

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

[45] J. Liu *et al.*, "Semi-supervised change detection based on graphs with generative adversarial networks," in *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul./Aug. 2019, pp. 74–77.

[46] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5891–5906, Jul. 2021.

[47] X. Li and A. G. O. Yeh, "Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the pearl river delta," *Int. J. Remote Sens.*, vol. 19, no. 8, pp. 1501–1518, Jan. 1998.

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

[49] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.

[50] 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.

[51] 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.

[52] S. Saha, L. Mou, C. Qiu, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Unsupervised deep joint segmentation of multitemporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8780–8792, Dec. 2020.

[53] L. Bergamasco, S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised change-detection based on convolutional-autoencoder feature extraction," *Proc. SPIE*, vol. 11155, Oct. 2019, Art. no. 1115510.

[54] Y. Xu, S. Xiang, C. Huo, and C. Pan, "Change detection based on auto-encoder model for VHR images," *Proc. SPIE*, vol. 8919, Oct. 2013, Art. no. 891902.

[55] 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. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.

[56] J. Geng, H. Wang, J. Fan, and X. Ma, "Change detection of SAR images based on supervised contractive autoencoders and fuzzy clustering," in *Proc. Int. Workshop Remote Sens. Intell. Process. (RSIP)*, May 2017, pp. 1–3.

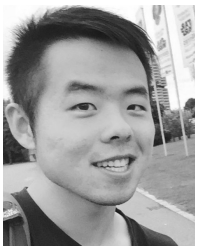
[57] E. Kalinicheva, J. Sublime, and M. Trocan, "Change detection in satellite images using reconstruction errors of joint autoencoders," in *Proc. Int. Conf. Artif. Neural Netw. (ICANN)*, 2019, pp. 637–648.

[58] Y. Chu, G. Cao, and H. Hayat, "Change detection of remote sensing image based on deep neural networks," in *Proc. 2nd Int. Conf. Artif. Intell. Ind. Eng. (AIIE)*, 2016, pp. 262–267.
[59] Y. Li, C. Peng, Y. Chen, L. Jiao, L. Zhou, and R. Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5751–5763, Aug. 2019.
[60] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.*, vol. 11, no. 3, p. 258, 2019.
[61] M. Yang, L. Jiao, F. Liu, B. Hou, and S. Yang, "Transferred deep learning-based change detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6960–6973, Sep. 2019.
[62] Z. Yuan, Q. Wang, and X. Li, "ROBUST PCANet for hyperspectral image change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 4931–4934.
[63] B. Hou, Y. Wang, and Q. Liu, "Change detection based on deep features and low rank," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2418–2422, Dec. 2017.
[64] R. Touati, M. Mignotte, and M. Dahmane, "Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based Markov random field model," *IEEE Trans. Image Process.*, vol. 29, pp. 757–767, 2020.
[65] R. Touati, M. Mignotte, and M. Dahmane, "Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 588–600, 2020.
[66] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019.
[67] L. T. Luppino *et al.*, "Deep image translation with an affinity-based change prior for unsupervised multimodal change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, 2022.
[68] M. G. Gong, P. Z. 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.
[69] M. Gong, X. Niu, T. Zhan, and M. Zhang, "A coupling translation network for change detection in heterogeneous images," *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3647–3672, May 2019.
[70] J. Miller, "Reaction time analysis with outlier exclusion: Bias varies with sample size," *Quart. J. Exp. Psychol. A*, vol. 43, no. 4, pp. 907–912, 1991.
[71] J. L. C. Hoffmann, L. P. Horstmann, and A. A. Fröhlich, "Anomaly detection in multicore embedded systems," in *Proc. IX Brazilian Symp. Comput. Syst. Eng. (SBESC)*, Nov. 2019, pp. 1–8.
[72] M. Van Selst and P. Jolicoeur, "A solution to the effect of sample size on outlier elimination," *Quart. J. Exp. Psychol. A*, vol. 47, no. 3, pp. 631–650, Aug. 1994.
[73] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral Earth observation using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 2115–2118.
[74] T. Dozat, "Incorporating Nesterov momentum into Adam," in *Proc. Int. Conf. Learn. Represent. Workshops (ICLRW)*, 2016.
[75] P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, "Good practices for estimating area and assessing accuracy of land change," *Remote Sens. Environ.*, vol. 148, pp. 42–57, May 2014.
[76] A. H. Strahler *et al.*, "Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps," *Eur. Communities, Luxembourg*, vol. 51, no. 4, pp. 1–60, 2006.
[77] R. G. Pontius, Jr., and M. Millones, "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment," *Int. J. Remote Sens.*, vol. 32, no. 15, pp. 4407–4429, Aug. 2011.
[78] G. M. Foody, "Explaining the unsuitability of the Kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification," *Remote Sens. Environ.*, vol. 239, Mar. 2020, Art. no. 111630.
[79] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
[80] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
[81] C. Wu, H. Chen, B. Du, and L. Zhang, "Unsupervised change detection in multitemporal VHR images based on deep kernel PCA convolutional mapping network," *IEEE Trans. Cybern.*, early access, Jul. 8, 2021, doi: 10.1109/TCYB.2021.3086884.
[82] H. Chen, C. Wu, B. Du, and L. Zhang, "Deep Siamese multi-scale convolutional network for change detection in multi-temporal VHR images," in *Proc. 10th Int. Workshop Anal. Multitemporal Remote Sens. Images (MultiTemp)*, Aug. 2019, pp. 1–4.
[83] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, Jan. 2019.
[84] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.
[85] L. Ruff *et al.*, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 4393–4402.
[86] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–14.

Lichao Mou received the bachelor's degree in automation from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2012, the master's degree in signal and information processing from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2015, and the Dr.-Ing. degree from the Technical University of Munich (TUM), Munich, Germany, in 2020.

He is currently a Guest Professor with the Munich AI Future Laboratory AI4EO, TUM and the Head of Visual Learning and Reasoning Team, Department "EO Data Science", Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany. Since 2019, he has been a Research Scientist at DLR-IMF and an AI Consultant for the Helmholtz Artificial Intelligence Cooperation Unit (HAICU). In 2015, he spent six months with the Computer Vision Group, University of Freiburg, Freiburg, Germany. In 2019, he was a Visiting Researcher with the Cambridge Image Analysis Group (CIA), University of Cambridge, Cambridge, U.K.

Dr. Mou was a recipient of the first place in the 2016 IEEE GRSS Data Fusion Contest and finalists for the Best Student Paper Award at the 2017 Joint Urban Remote Sensing Event and 2019 Joint Urban Remote Sensing Event.



Yuansheng Hua received the bachelor's degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, and the double master's degrees in Earth oriented space science and technology (ESPACE) and photogrammetry and remote sensing from the Technical University of Munich (TUM), Munich, Germany, and Wuhan University, in 2018 and 2019, respectively, and the Dr.-Ing. degree from TUM, in 2022.

In 2019, he was a Visiting Researcher with the Wageningen University and Research, Wageningen, the Netherlands. His research interests include remote sensing, computer vision, and deep learning, especially their applications in remote sensing.

Sudipan Saha (Member, IEEE) received the M.Tech. degree in electrical engineering from IIT Bombay, Mumbai, India, in 2014, and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, and Fondazione Bruno Kessler, Trento, in 2020.

He worked as an Engineer with TSMC Limited, Hsinchu, Taiwan, from 2015 to 2016, where he was a Guest Researcher in 2019. He is currently a Post-Doctoral Researcher with the Technical University of Munich (TUM), Munich, Germany. His research interests are related to multitemporal remote sensing image analysis, domain adaptation, uncertainty quantification, image segmentation, deep learning, and image processing. He is a reviewer for several international journals and computer vision conferences.

Dr. Saha was a recipient of the Fondazione Bruno Kessler Best Student Award 2020. He served as a Guest Editor at *Remote Sensing* (MDPI) special issue on "Advanced Artificial Intelligence for Remote Sensing: Methodology and Application."

Francesca Bovolo (Senior Member, IEEE) received the Laurea (B.S.) degree and the Laurea Specialistica (M.S.) degree (*summa cum laude*) in telecommunication engineering, and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003, and 2006, respectively.

She was a Research Fellow with the University of Trento, until 2013. She is currently the Founder and the Head of Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler, Trento, and a member of the Remote Sensing Laboratory, Trento. She is one of the coinvestigators of the Radar for Icy Moon Exploration instrument of the European Space Agency Jupiter Icy Moons Explorer and a member of the Science Study Team of the EnVision Mission to Venus. Her research interests include remote sensing image processing, multitemporal remote sensing image analysis, change detection in multispectral, hyperspectral, synthetic aperture radar images, and very high-resolution images, time series analysis, content-based time series retrieval, domain adaptation, and light detection and ranging (LiDAR) and radar sounders. She conducts research on these topics within the context of several national and international projects.

Dr. Bovolo is a member of the Program and Scientific Committee of several international conferences and workshops. She was a recipient of the First Place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, 2006). She was the Technical Chair of the Sixth International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp 2011, and 2019). She has been a Co-Chair of the SPIE International Conference on Signal and Image Processing for Remote Sensing since 2014. She is the Publication Chair of the International Geoscience and Remote Sensing Symposium in 2015. She has been an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2011 and the Guest Editor of the Special Issue on Analysis of Multitemporal Remote Sensing Data of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She is a Referee for several international journals.

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

He is currently a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications.

Dr. Bruzzone is the Founder and the Director of the Remote Sensing Laboratory (https://rslab.disi.unitn.it/), Department of Information Engineering and Computer Science, University of Trento. His research interests are in the areas of remote sensing, radar and SAR, signal processing, machine learning, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the Principal Investigator of many research projects. Among the others, he is currently the Principal Investigator of the *Radar for icy Moon exploration (RIME)* instrument in the framework of the *JUpiter ICy moons Explorer (JUICE)* mission of the European Space Agency (ESA) and the Science Lead for the *High Resolution Land Cover* project in the framework of the Climate Change Initiative of ESA. He is the author (or coauthor) of 294 scientific publications in referred international journals (221 in IEEE journals), more than 340 articles in conference proceedings, and 22 book chapters. He is the Editor/Coeditor of 18 books/conference proceedings and one scientific book. His articles are highly cited, as proven from the total number of citations (more than 37 000) and the value of the h-index (89) (source: Google Scholar). He was invited as the Keynote Speaker in more than 40 international conferences and workshops. Since 2009, he has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS), where since 2019 he has been the Vice-President for Professional Activities. He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since that he was a recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards, and the 2019 WHISPER Outstanding Paper Award. He was a Guest Coeditor of many Special Issues of international journals. He is the Co-Founder of the IEEE International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the Founder of the *IEEE Geoscience and Remote Sensing Magazine* for which he has been the Editor-in-Chief from 2013 to 2017. He is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has been the Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society from 2012 to 2016.

Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.-Ing., and "Habilitation" degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is currently the Professor for Data Science in Earth Observation (former: Signal Processing in Earth Observation), TUM, and the Head of the Department "EO Data Science," Remote Sensing Technology Institute, German Aerospace Center (DLR). Since 2019, she has been a Co-Coordinator of the Munich Data Science Research School (www.mu-ds.de). Since 2019, she also heads the Helmholtz Artificial Intelligence—Research Field "Aeronautics, Space and Transport." Since May 2020, she is the Director of the international future AI Laboratory "AI4EO–Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond," Munich, Germany. Since October 2020, she also serves as a Co-Director of the Munich Data Science Institute (MDSI), TUM. She was a Guest Scientist or the Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan, and University of California, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently a Visiting AI Professor at ESA's Phi-lab. Her main research interests are remote sensing and Earth observation, signal processing, machine learning, and data science, with their applications in tackling societal grand challenges, e.g., Global Urbanization, UN's SDGs, and Climate Change.

Dr. Zhu is a member of Young Academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She serves in the Scientific Advisory Board in several research organizations, among others the German Research Center for Geosciences (GFZ) and Potsdam Institute for Climate Impact Research (PIK). She is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and serves as the Area Editor responsible for special issues of the *IEEE Signal Processing Magazine*.