Building Change Detection in VHR SAR Images via Unsupervised Deep Transcoding

Sudipan Saha, *Graduate Student Member, IEEE*, Francesca Bovolo, *Senior Member, IEEE*, and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*— Building change detection (CD), important for its application in urban monitoring, can be performed in near real time by comparing prechange and postchange very-high-spatial-resolution (VHR) synthetic-aperture-radar (SAR) images. However, multitemporal VHR SAR images are complex as they show high spatial correlation, prone to shadows, and show an inhomogeneous signature. Spatial context needs to be taken into account to effectively detect a change in such images. Recently, convolutional-neural-network (CNN)-based transfer learning techniques have shown strong performance for CD in VHR multispectral images. However, its direct use for SAR CD is impeded by the absence of labeled SAR data and, thus, pretrained networks. To overcome this, we exploit the availability of paired unlabeled SAR and optical images to train for the suboptimal task of transcoding SAR images into optical images using a cycle-consistent generative adversarial network (CycleGAN). The CycleGAN consists of two generator networks: one for transcoding SAR images into the optical image domain and the other for projecting optical images into the SAR image domain. After unsupervised training, the generator transcoding SAR images into optical ones is used as a bitemporal deep feature extractor to extract optical-like features from bitemporal SAR images. Thus, deep change vector analysis (DCVA) and fuzzy rules can be applied to identify changed buildings (new/destroyed). We validate our method on two data sets made up of pairs of bitemporal VHR SAR images on the city of L'Aquila (Italy) and Trento (Italy).**

***Index Terms*— Change detection (CD), deep change vector analysis (DCVA), generative adversarial network (GAN), multitemporal images, remote sensing, synthetic aperture radar (SAR), very high-resolution images.**

I. Introduction

CHANGE DETECTION (CD) in very-high-spatial-resolution (VHR) images [1] is important for several applications, including urban planning, disaster management, and cadastral map updating. In the last decade, a new generation of VHR satellite sensors has been launched, which can acquire images having a spatial resolution of one meter or less. The availability of VHR data allows us to analyze single man-made structures, e.g., buildings [2]. In this context, several techniques for the analysis of urban areas have been developed exploiting both passive (optical) [3], [4] and active [synthetic aperture radar (SAR)] sensors [5]–[8]. SAR sensors are particularly useful in the applications that require a quick response (e.g., disaster management) as they effectively map the affected areas irrespective of the time of the day or the weather conditions [7], thus with a potential better temporal resolution. Currently, several satellites with SAR sensors are operating (e.g., TerraSAR-X, Tandem-X, COSMO-SkyMed constellation, and COSMO-SkyMed second-generation constellation) that can acquire VHR images.

In the CD literature, unsupervised methods [9]–[12] are preferred due to the difficulty of collecting multitemporal labeled data, which becomes more severe in the case of postdisaster CD. Difference-based unsupervised CD methods [9] and its log-based variants [13], [14] (to suppress the multiplicative speckle noise) are popular in the literature. VHR SAR images are more complex than low/medium resolution images as they show high spatial correlation [15], [16]. Semantically homogeneous objects, such as buildings, show inhomogeneous signature at high resolution due to different scattering contributions from subobjects [5]. It is required to exploit the contextual and object-level information to extract change information effectively. There are few works in the literature that can handle the complexity of multitemporal VHR SAR data [5], [17], [18]. Brett and Guida [17] proposed a method using curvilinear features to detect changes caused by earthquakes. Marin *et al.* [5] proposed a method that exploits the increment and decrement of backscattering along with a set of fuzzy rules to detect building changes. Yousif and Ban [18] proposed an object-based CD method for HR SAR images. The limitations of these methods are as follows.

1) They only extract low-level features (e.g., texture and curvilinear feature) from VHR images for CD, which are not robust for representing the semantic information of bitemporal images.
2) They rely only on the model assumptions or the degree of similarity in the considered scene. Thus, they fail in exploiting an enormous amount of unlabeled remote sensing data that are currently available and can be exploited to improve performance.

Recently, deep neural networks, especially the convolutional neural network (CNN), have demonstrated remarkable performance in image processing tasks [19]. They are suitable to extract semantically rich features that capture object-level information [20]. Motivated by this, a few works have been

Manuscript received April 10, 2020; revised May 26, 2020; accepted May 28, 2020. Date of publication June 18, 2020; date of current version February 25, 2021. *(Corresponding author: Francesca Bovolo.)*

Sudipan Saha is with Fondazione Bruno Kessler, 38123 Trento, Italy, and also with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: saha@fbk.eu).

Francesca Bovolo is with Fondazione Bruno Kessler, 38123 Trento, Italy (e-mail: bovolo@fbk.eu).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.3000296

proposed for SAR CD, which are mostly supervised [21], [22] and/or deal with low/medium resolution images [23], [24]. Gong *et al.* [21] proposed a CD method that first trains a recurrent Boltzmann machine (RBM) in an unsupervised way and tunes it in a supervised fashion. Gao *et al.* [22] proposed principal component analysis net (PCANet) that exploits PCA filters as convolutional filters, and subsequently, the output of the PCA filters is fed into a classifier to predict the CD map. A preclassification scheme is designed to obtain some labeled samples of high accuracy to train the PCANet. The accuracy of the method strongly depends on the accuracy of the preclassification scheme. Li *et al.* [25] incorporated the concept of image saliency in PCANet. Li *et al.* [26] proposed a method based on a preclassification scheme where initial pseudolabels are produced through unsupervised spatial fuzzy clustering. Similarly, [27] also used spatial fuzzy clustering for pseudolabel generation. Gong *et al.* [28] used a sparse autoencoder to transform log-ratio difference image into a feature space, and those features are clustered to generate pseudolabels. Liu *et al.* [29] proposed a symmetric convolutional coupling network for CD in heterogeneous optical and SAR images. In [23], a supervised method is proposed for sea ice CD using a convolutional-wavelet neural network (CWNN) in low-resolution images. Similar to [22], the method in [23] used a preclassification scheme to generate training samples. In [30], a two-channel CNN is used to estimate the similarity between the bitemporal patches. The limitations of these methods are as follows.

1) Most of them are supervised and do not account for the lack of labeled multitemporal data [31], [32] at large scale.
2) The methods relying on preclassification do not need labeled data. However, they are still incapable of utilizing an enormous amount of remote sensing data that are currently available and can be used to improve performance. Moreover, their accuracy depends on the accuracy of the preclassification scheme.

Recently, few deep learning-based supervised methods [33], [34] have been proposed for building CD. Chen and Yu [33] proposed a supervised method exploiting residual deep network to map earthquake-induced damaged buildings. Li *et al.* [34] used residual U-Net to detect building changes in Sentinel-1 images.

Another advancement in the field of deep learning is transfer learning that enables a model trained on a certain task to be used for another task [35]. Transfer learning has shown excellent capability in different remote sensing tasks [36], [37]. Inspired by the success of transfer learning, Saha *et al.* [4] proposed deep-change-vector-analysis (DCVA) for CD in multispectral optical satellite images by exploiting deep features extracted from a pretrained network [38]. To apply a DCVA framework, a pixelwise labeled VHR database (that is not multitemporal) is used to train a deep network that is subsequently used as a multitemporal deep feature extractor. DCVA is unsupervised as it does not require any labeled multitemporal data. However, it successfully exploits the huge amount of available remote sensing data. In spite of its success for CD in VHR optical images [4], applying DCVA for VHR SAR images is not trivial. Obtaining labeled VHR SAR data set is very challenging due to the difficulty of labeling VHR SAR images [39]. Thus, there is a need for a method that can circumnavigate the necessity of labeled SAR images.

A step forward in the paradigm of deep networks, generative adversarial networks (GANs), can learn to mimic complex data distributions from unlabeled data. It has shown the promising capability for transfer learning tasks [40] that has inspired its use in remote sensing [41], [42]. Ley *et al.* [42] showed that transcoding SAR images into optical images force GAN to learn deep features for distinguishing between different land surfaces. The mechanism does not require labeled data. Considering the availability of many VHR SAR and optical sensors traversing the Earth repeatedly [39], it is possible to obtain multitemporal pairs' SAR and optical images from the same locations on Earth, thus allowing for unsupervised transcoding of SAR images into optical ones and training of deep network circumnavigating the necessity of labeled data. The task of transcoding SAR images into optical ones is suboptimal/ill-posed as there are features in SAR images that are not present in the optical images, and vice versa [42], [43]. While SAR images emphasize the physical properties of the target surfaces, the optical images highlight structural details [43]. Even if it is not possible to completely transcode SAR data to actual optical data, Ley *et al.* [42] and Reyes *et al.* [43] observed that tasking GAN to learn this transcoding forces the GAN to learn useful semantic features [42]. Motivated by this, we propose a CD method that uses SAR-optical transcoding [43] to train a deep network that is subsequently used as bitemporal deep feature extractor in the DCVA framework. The method assumes the availability of a data set of SAR and optical images obtained from the same location or similar geographical locations, i.e., images representing similar behavior. Different variants of GAN are available in the literature [44]–[46]. Inspired by the work of Reyes *et al.* [43], the proposed method exploits cycle consistent GAN (CycleGAN) [44] framework to learn the transcoding between SAR and optical images. The CycleGAN framework does not require the presence of paired/coregistered SAR and optical pairs. The CycleGAN framework consists of two generator networks and two discriminator networks. One generator is tasked to transcode SAR images into the optical domain, while the other is tasked to transcode the optical images into the SAR domain. The network learns exploiting a set of loss functions designed for adversarial training and cycle consistency. Thus, after unsupervised training of CycleGAN, the generator network transcoding SAR images into optical is used as a deep feature extractor from multitemporal SAR images. The use of CycleGAN to learn transcoding between SAR and optical images is as follows.

1) Help to train a deep network without the requirement of any labeled training data.
2) Do not assume the presence of coregistered SAR and optical images. While SAR and optical images can be potentially collected from similar geographical locations, collecting coregistered pairs is difficult. This significantly relaxes one of the strongest constraints in CD.

3) Ingest the knowledge from a plethora of unlabeled images used in the training process. Similar to transfer learning, while subsequently using a generator of the CycleGAN as a deep feature extractor in the CD process, the framework can use the semantic features learned from the plethora of images for the CD task.

The proposed method exploits SAR-optical transcoding to learn useful semantic features for multitemporal SAR image analysis. Recovery of optical data from SAR data has its limitations [42], [43] and is beyond the scope of this article.

Coregistered prechange and postchange VHR SAR images are processed through the multilayered CNN (i.e., the generator) to obtain deep features. Deep features are compared to each other to obtain the deep change hypervectors that are processed using a DCVA framework [4] originally developed for optical images only and a fuzzy building detection model [5] to identify changed buildings (new/destroyed).

The novelty of this article is that we propose an unsupervised method to train a deep network that is used as a multitemporal optical-like deep feature extractor from SAR images to be processed in the DCVA framework. On the contrary to [22] and [26]–[28], the proposed mechanism is completely unsupervised and can ingest knowledge from a plethora of unlabeled images in the training process. Effectiveness of the multitemporal feature extractor is demonstrated by suitably coupling with DCVA framework [4] and fuzzy building detection model [5] for a practical application, i.e., building CD.

This article is organized into the following sections. Section II formulates the problem statement and presents a synopsis of the proposed solution. Section III presents in detail the proposed CD framework for detecting destroyed buildings. Experimental results are presented in Section IV. We conclude this article and discuss the scope of further investigations in Section V.

II. Problem Formulation and Synopsis of the Proposed Solution

SAR is an active imaging system, and a SAR image is formed by coherently processing the backscatter returns from successive radar pulses. Due to this acquisition mechanism, speckle noise (a salt-and-pepper granular pattern) inherently manifests itself in the SAR images [47]. Speckle noise is multiplicative in nature. Thus, to reduce its effect, we assume images to be in the dB scale.

Let X_1 and X_2 be two VHR SAR images taken over the same geographical region at times t_1 and t_2, respectively, using the same sensor and same acquisition angle. Let the set of all pixels in the bitemporal scene be represented by Ω. The proposed method aims to detect the changes corresponding to the building between X_1 and X_2 in an unsupervised manner, i.e., without using any labeled bitemporal data.

Let us assume that generic data sets of unlabeled VHR SAR patches $\mathbf{X} = \{\mathbf{x_i} \forall \mathbf{i} = \mathbf{1}, \ldots, \mathbf{I}\}$ and VHR optical patches $\mathbf{Z} = \{\mathbf{z_i} \forall \mathbf{i} = \mathbf{1}, \ldots, \mathbf{I}\}$ are available. Considering the difficulty of the SAR-optical transcoding task, we consider optical patches are panchromatic. $\mathbf{I}$ is the number of patches in the data sets. $\mathbf{X}$ and $\mathbf{Z}$ do not need to be paired/coregistered. However, they must be acquired from similar geographical locations, thus implying similar information content and image distribution. This condition significantly relaxes the typical hypothesis of CD methods where images pairs have to be coregistered. Furthermore, $\mathbf{X}$ and $\mathbf{Z}$ do not necessarily include images of the geographical region of X_1 and X_2. The proposed method exploits $\mathbf{X}$ and $\mathbf{Z}$ to train a CycleGAN framework that consists of two generators: one for transcoding SAR images from $\mathbf{X}$ into the optical domain of $\mathbf{Z}$ and the other for transcoding images from $\mathbf{Z}$ into the domain of $\mathbf{X}$. After training, the generator that transcodes images from $\mathbf{X}$ into $\mathbf{Z}$ is used to extract optical-like bi-temporal deep features from SAR images X_1 and X_2. DCVA framework defined for VHR optical images [4] is applied to divide the set of all pixels Ω in an unsupervised manner into two subsets Ω_c and ω_{nc} corresponding to changed and unchanged pixels, respectively. The pixels in Ω_c are further analyzed to cluster into two different types of change, corresponding to increment (ξ^+) and decrement (ξ^-) in deep feature space. Following this, a fuzzy building detection model [5] is employed for building CD. The block scheme of the proposed method is shown in Fig. 1.

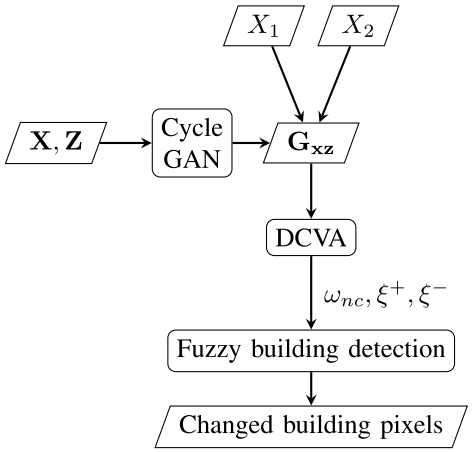


Fig. 1. Proposed CD framework.

III. Proposed Method

The proposed method is accomplished in the following steps: 1) learning transcoding between SAR and optical images using CycleGAN; 2) exploiting the CycleGAN for bitemporal optical-like deep feature extraction from SAR images; and 3) changed building detection using the DCVA framework and the fuzzy building detection model.

A. Learning Transcoding Between SAR and Optical

CycleGAN [44] is chosen to learn the transcoding between VHR SAR and VHR optical domains due to their capability to work with spatially uncoupled images. The CycleGAN training process is achieved with data sets of unlabeled VHR SAR $\mathbf{X} = \{\mathbf{x_i} \forall \mathbf{i} = \mathbf{1}, \ldots, \mathbf{I}\}$ and optical $\mathbf{Z} = \{\mathbf{z_i} \forall \mathbf{i} = \mathbf{1}, \ldots, \mathbf{I}\}$ patches. Assuming that the SAR patches in $\mathbf{X}$ are drawn from a distribution $p^*(\mathbf{x})$ and the optical patches in $\mathbf{Z}$ are drawn from the distribution $q^*(\mathbf{z})$, the unpaired patch-to-patch translation learns correspondence between distributions

of SAR ($p^*(\mathbf{x})$) and optical ($q^*(\mathbf{z})$). Learning such a transocding/correspondence is nontrivial and suboptimal since the network must learn semantic entities to synthesize the corresponding optical textures from SAR images. However, we train a SAR-optical transcoder only as a proxy task that facilitates learning semantic attributes from SAR images. To accomplish the training process, CycleGAN [44] uses two generators $\mathbf{G_{xz}}$ and $\mathbf{G_{zx}}$ and two discriminators $\mathbf{D_z}$ and $\mathbf{D_x}$. These components interact following two criteria.

1) Adversarial criterion optimizes $\mathbf{G_{xz}}$ and $\mathbf{D_z}$ together. The generator $\mathbf{G_{xz}}$ has no access to the real optical patches $\mathbf{Z}$. Given $\mathbf{X}$, it learns to project it to the distribution $q^*(\mathbf{z})$ only through its interaction with the discriminator $\mathbf{D_z}$. $\mathbf{D_z}$ has access to both patches in $\mathbf{Z}$ and patches generated by $\mathbf{G_{xz}}$. The adversarial mechanism works based on the assumption that if $\mathbf{G_{xz}}$ successfully learns to transform images in $\mathbf{X}$ to those in $\mathbf{Z}$, $\mathbf{D_z}$ will fail to distinguish between real patches in $\mathbf{Z}$ and those generated by $\mathbf{G_{xz}}$. Thus, based on the feedback produced by the discriminator $\mathbf{D_z}$, $\mathbf{G_{xz}}$ improves its approximation of $q^*(\mathbf{z})$ in the iterative fashion. The following is in more detail.
   a) The generator $\mathbf{G_{xz}}$ generates $\tilde{\mathbf{Z}}$ that mimics the distribution $q^*(\mathbf{z})$ given $\mathbf{X}$ that is drawn from the distribution $p^*(\mathbf{x})$. The generator has an encoder-transformer-decoder architecture that consists of a series of convolutional layers, ResNet blocks, and deconvolutional layers.
   b) The discriminator $\mathbf{D_z}$ tries to distinguish patches $\tilde{\mathbf{Z}}$ generated by $\mathbf{G_{xz}}$ (commonly called fake patches [44]) from real patches drawn from $\mathbf{Z}$. The generator $\mathbf{G_{xz}}$ and discriminator $\mathbf{D_z}$ interact in a minimax fashion where $\mathbf{G_{xz}}$ tries to minimize, while $\mathbf{D_z}$ tries to maximize the same objective function

$$\min_{\mathbf{G_{xz}}} \max_{\mathbf{D_z}} \mathbb{E}[\log \mathbf{D_z}(\mathbf{z})] + \mathbb{E}[\log(1 - \mathbf{D_z}(\tilde{\mathbf{z}}))]. \quad (1)$$

   Similarly, the generator $\mathbf{G_{zx}}$ and the discriminator $\mathbf{D_x}$ jointly learn to generate $\tilde{\mathbf{X}}$ that mimics the distribution $p^*(\mathbf{x})$ given $\mathbf{Z}$ that is drawn from the distribution $q^*(\mathbf{z})$. $\mathbf{G_{zx}}$ and $\mathbf{D_x}$ are trained in adversarial fashion to optimize the objective function

$$\min_{\mathbf{G_{zx}}} \max_{\mathbf{D_x}} \mathbb{E}[\log \mathbf{D_x}(\mathbf{x})] + \mathbb{E}[\log(1 - \mathbf{D_x}(\tilde{\mathbf{x}}))]. \quad (2)$$

   The adversarial criterion is not sufficient to learn appropriate transcoding between images in $\mathbf{X}$ and $\mathbf{Z}$ [44]. This is because it only tasks $\mathbf{G_{xz}}$ to translate patches in $\mathbf{X}$ to look like patches in $\mathbf{Z}$. However, it does not ensure that input and output correspond to the same object. For example, a patch showing a building in $\mathbf{X}$ can be converted to a patch showing a realistic road in $\mathbf{Z}$. Adversarial criterion only enforces that the generated output is of the appropriate domain and does not ensure that output is semantically related to the input.
2) Cycle-consistency criterion works on the limitation of the adversarial criterion and is inspired from the circular strategy in the domain adaptation literature [48]. It ensures that if patches sampled from SAR images ($\mathbf{X}$) are transformed twice consecutively through $\mathbf{G_{xz}}$ and $\mathbf{G_{zx}}$, we get back the original patches in $\mathbf{X}$

$$\mathbf{G_{zx}}(\mathbf{G_{xz}})(\mathbf{x}) \approx \mathbf{x}. \quad (3)$$

   Similarly, if patches sampled from optical images ($\mathbf{Z}$) are transformed twice consecutively through $\mathbf{G_{zx}}$ and $\mathbf{G_{xz}}$, we get back the original optical patches in $\mathbf{Z}$

$$\mathbf{G_{xz}}(\mathbf{G_{zx}})(\mathbf{z}) \approx \mathbf{z}. \quad (4)$$

   Using this criterion, it is possible to ensure that output generated by $\mathbf{G_{xz}}$, and $\mathbf{G_{zx}}$ is semantically related to their corresponding input. This is because if the generators learn to transcode to other domains without learning object to object correspondence, it is highly improbable that after transcoding twice, the same object will be obtained, e.g., if $\mathbf{G_{xz}}$ projects a building to a realistic road, with all probability that $\mathbf{G_{zx}}$ will fail to project back the road to a realistic building. Thus, the cycle-consistency ensures that semantic consistency is maintained in the transcoding process. Moreover, constraint of cycle consistency helps the CycleGAN network to learn transcoding between SAR and optical domains from unpaired patches in $\mathbf{X}$ and $\mathbf{Z}$, i.e., given a training patch in $\mathbf{X}$, the corresponding exactly geolocated patch in $\mathbf{Z}$ is not required for the training process. Thus, the cycle consistency loss added to the adversarial loss makes the training process more robust [43].

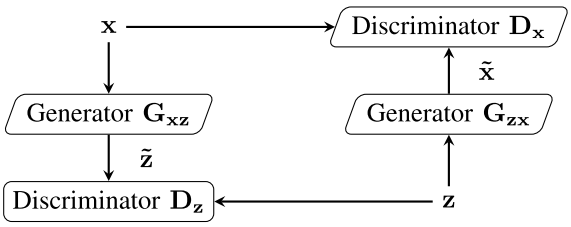


Fig. 2. CycleGAN training process.

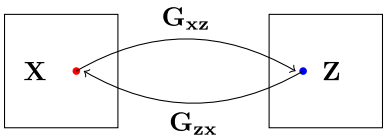


Fig. 3. Cycle-consistency constraint: the left and right rectangles represent the distribution space spanned by $\mathbf{X}$ ($p^*(\mathbf{x})$) and $\mathbf{Z}$ ($q^*(\mathbf{z})$), respectively. Drawing from the distribution $p^*(\mathbf{x})$ and processing twice through $\mathbf{G_{xz}}$ and $\mathbf{G_{zx}}$ yields the origin in $p^*(\mathbf{x})$.

Combining the cycle-consistency criterion with the adversarial losses yields full objective for learning transcoding between $p^*(\mathbf{x})$ and $q^*(\mathbf{z})$. This learning phase is completely unsupervised, i.e., no labeled training data are required for it. Fig. 2 shows the adversarial learning mechanism, and Fig. 3 shows the cycle-consistency criterion.

B. CycleGAN-Based Bitemporal Deep Features Extraction

After training CycleGAN, the weights of the generator $\mathbf{G_{xz}}$ are frozen and used as a bitemporal deep feature extractor for CD. The CNN $\mathbf{G_{xz}}$ consists of multiple convolutional layers, but it does not have any fully connected layer. Hence, this

CNN behaves as a fully convolutional CNN, and the input of any spatial size can be fed to it. Prechange and postchange SAR images X_1 and X_2 are separately processed through $\mathbf{G_{xz}}$ to obtain optical-like multitemporal deep features for each pixel of the analyzed scene. This allows for the use of CD approaches designed for optical images. Here, we use an approach inspired by DCVA [4] that was originally proposed for VHR optical images.

Obtaining features from multiple layers of CNN [49], [50] allows reasoning at multiple levels of abstraction and scales. Based on this, we further chose suitable layers L to extract features for CD. The first convolutional layer of the $\mathbf{G_{xz}}$ captures primitive features, such as edges, and may add noise to the CD process. Previous works on transfer learning [4], [42], [49], [51], [52] demonstrated that intermediate layers are more suitable for transfer learning tasks. The deeper convolutional layers are more oriented toward the task for which the network is trained and less suitable for transfer learning. Thus, the method chooses the layers in L from the intermediate layers. Features from layer l in L are upsampled using bilinear interpolation [4] to the spatial size of the input SAR images to obtain f_l^1 and f_l^2.

Layerwise deep features' difference (by subtracting f_l^1 from f_l^2) is taken to obtain a change vector G_l, corresponding to layer l. Some features carry relevant information for CD, while others do not. We assume that features capturing relevant change information have higher variance/standard deviation than those features less responsive to change information [4], [53]. After computing the difference image, features in G_l not affected by change show values that all tend to zero (no change means that the pixel has similar values over time). Features in G_l affected by change show both values that tend to zero for the portion being not affected by the change and values far from zero for the portion being affected (change means that the pixel assumes dissimilar values over time). Accordingly, the variance of the features in the latter case tends to be greater than the former one. Based on this, we employ a variance-based automatic feature selection strategy [4] to layerwise select discriminative features. The resulting deep change hypervector G_l' ($G_l' \in G_l$) effectively emphasizes change information. Layerwise selected features G_l' are concatenated for all layers l in L to obtain a D-dimensional deep change hypervector G that captures multiscale change information from chosen layers

$$G = (G_1', \ldots, G_l', \ldots, G_L'). \tag{5}$$

C. Changed Building Detection

Components of deep change hypervector G are represented as g^d ($d = 1, \ldots, D$). Assuming that unchanged pixels yield similar deep features, while changed ones do not, it can be postulated that components of G (i.e., g^d) have smaller absolute values for unchanged pixels (ω_{nc}) compared with changed pixels (Ω_c) [4]. Using this property, for each feature component g^d, we segregate pixels into two sets $\Omega_c^d(\forall |g^d| \geq \mathcal{T}^d)$ and $\omega_{nc}^d(\forall |g^d| < \mathcal{T}^d)$ using a component-specific threshold $\mathcal{T}^d$. Any automatic and unsupervised thresholding scheme [31] can be used to determine $\mathcal{T}^d$. Thus, D different CD maps are obtained, one for each feature g^d in G. A suitability score τ is assigned to each pixel that denotes the fraction of the D features that agree that a pixel is changed. Taking inspiration from multiscale ensemble decision level fusion in [54], pixels are segregated into changed and unchanged using majority voting, i.e., pixels are segregated into Ω_c (if $\tau \geq 0.5$) and unchanged ω_{nc} (if $\tau < 0.5$).

Ω_c includes two complimentary classes: changed buildings and all other changes. Changed buildings (new/destroyed) generate the specific signature in terms of the combination of the increment (ξ^+) and decrement (ξ^-) in deep feature space that allows them to be identified and separated from other kinds of changes [5]. Thus, Ω_c is further analyzed by clustering the deep change hypervector G into two classes using the deep direction analysis, as described in [4]. The presence/absence of new/destroyed buildings is analyzed by employing the fuzzy building CD system, as proposed by Marin *et al.* [5]. A building generates a signature that is characterized by the presence of: 1) backscattering contributions coming from the ground, the vertical wall, and the roof of the building (a layover area); 2) multiple scattering between the ground and the vertical wall (a double bounce line); and 3) the occlusion of the sensor due to the building (a shadow area). The appearance/disappearance of a building in the scene causes the appearance/disappearance of such primitives in the VHR SAR image [5]. The two dominant changes in the deep feature space are identified as $\delta^+ \in \xi^+$ and $\delta^- \in \xi^-$, and they are evaluated using the fuzzy rules to identify the changed buildings.

IV. Data Set and Experimental Results

The data set for training CycleGAN-based deep transcoder is detailed in Section IV-A. Experiments were performed on two data sets described in Section IV-B. Choice of the method for comparison is stated in Section IV-C. Choice of layers for bitemporal deep feature extraction is detailed in Section IV-D. Following that, Section IV-E discusses the deep feature visualization, and Section IV-F presents our results in detail.

A. CycleGAN Training

The SARptical data set [39] that is proposed in context of urban analysis is used in this article for CycleGAN training (both $\mathbf{X}$ and $\mathbf{Z}$). The data set provides more than 10 000 pairs of paired SAR and optical patches extracted from TerraSAR-X spotlight images and aerial UltraCAM optical images [39]. Even though the SARptical data set provides paired patches, the proposed method does not need them to be paired. For our training process, 8000 patches from both SAR and optical images are used. Table I[1] shows key structure of the generators $\mathbf{G_{xz}}$ and $\mathbf{G_{zx}}$. Table II shows key structure of the discriminators $\mathbf{D_z}$ and $\mathbf{D_x}$.

For CycleGAN training, we use the Adam optimizer [55] with a batch size of 1. The training is performed for 750 epochs with a learning rate of 0.0001 and momentum

[1]Detailed CycleGAN structure: https://github.com/sudipansaha/sarCdUsingDeepTranscoding

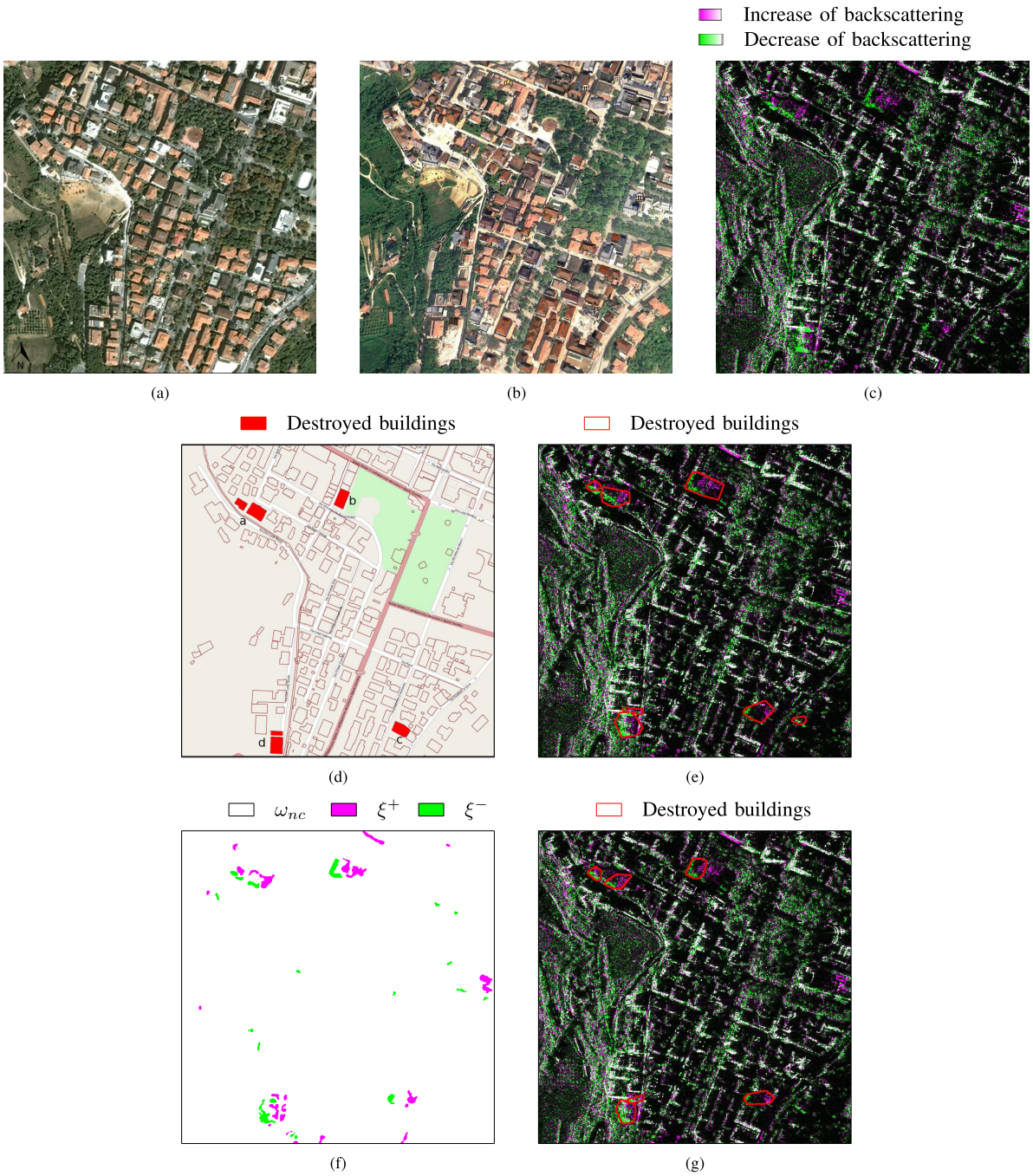


Fig. 4. L'Aquila (Italy) data set. (a) Optical image (September 4, 2006 [5]) [58]. (b) Optical image (May 8, 2009) [58]. (c) RGB multitemporal composition of COSMO-SKyMed images (R: September 12, 2009; G: April 5, 2009; and B: September 12, 2009). (Agenzia Spaziale Italiana, 2009. All Rights Reserved.) (d) Cadastral map of the area. (e) Destroyed buildings detected by Marin *et al.* [5]. (f) Increase and decrease of deep feature space detected by the proposed method. (g) Destroyed buildings detected by the proposed method.

(β_1) [55] of 0.5. The learning rate is kept fixed for the first 250 epochs and linearly decayed to zero over the next 500 epochs.

B. CD Data Set

Experiments were conducted on two data sets. One is related to the 2009 L'Aquila earthquake that occurred in the region of Abruzzo in central Italy. The other one captures the urban evolution of the city of Trento, Italy, between 2011 and 2013.

L'Aquila Earthquake data set [5] consists of two spotlight-mode X-band COSMO-SkyMed one-look amplitude images acquired in HH polarization on April 5, 2009, and September 12, 2009, over the city of L'Aquila, Italy (42°21′ N, 13°24′ E). L'Aquila was impacted by an earthquake of 6.3-moment magnitude on April 6, 2009. The images show an area of 1024 × 1024 pixels. Thus, the prechange image is acquired before the earthquake, and the postchange image is acquired after the immediate relief operation is finished. Fig. 4(a) and (b) shows the prechange and postchange optical images corresponding to the area of

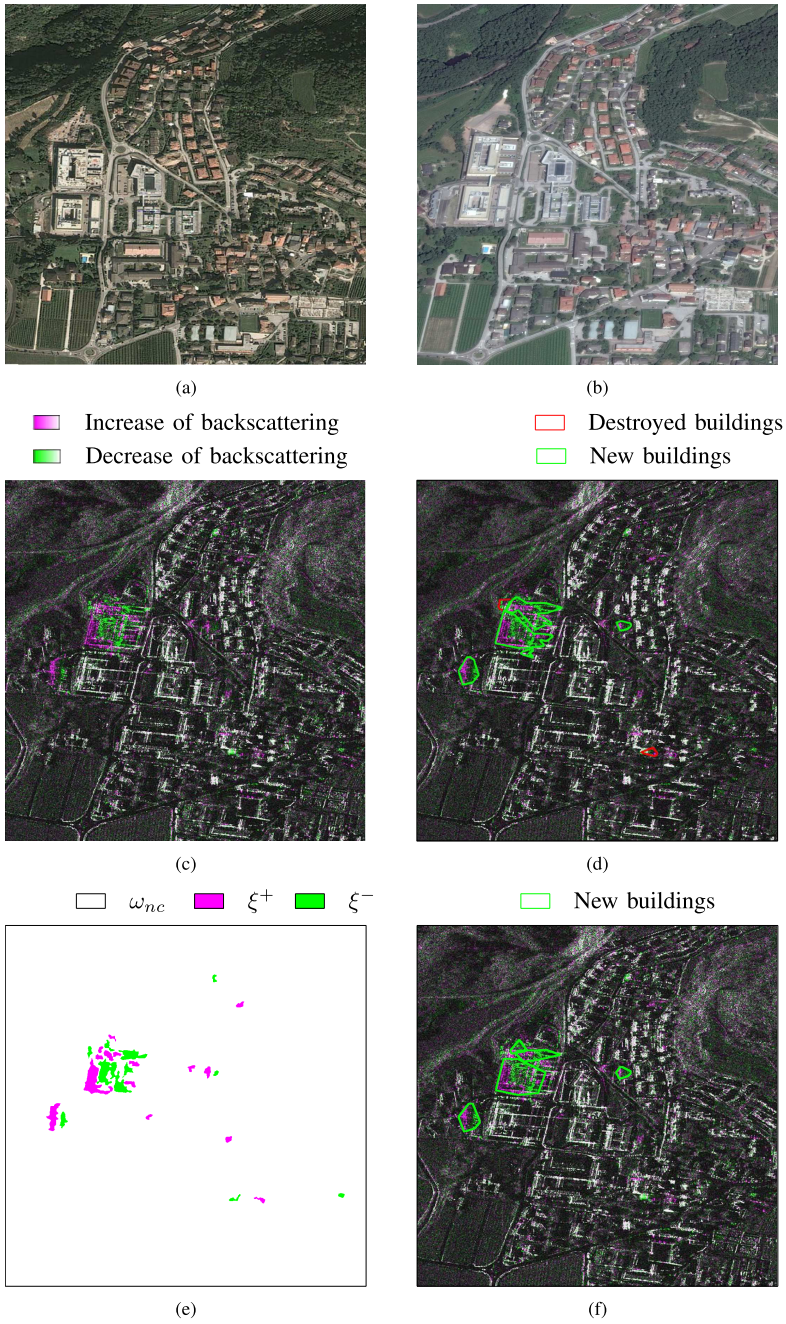


Fig. 5. Trento (Italy) data set. (a) Optical image (2011 [58]). (b) Optical image (2014 [58]). (c) RGB multitemporal composition of spotlight TerraSAR-X and TanDEM-X images (R: April 3, 2013; G: January 21, 2011; and B: April 3, 2013). (d) Changed buildings detected by Marin *et al.* [5]. (e) Increase and decrease of deep feature space detected by the proposed method. (f) Changed buildings detected by the proposed method.

interest. Multitemporal false-color composition of the data set (red channel: September 12, 2009; green channel: April 5, 2009; and blue channel: September 12, 2009) is shown in Fig. 4(c). Unchanged pixels appear in grayscale, while pixels with an increase in the value of backscattering appear in magenta tone, and pixels with a decrease in the value of backscattering appear in a green tone. The cadastral map of the area is shown in Fig. 4(d). Six buildings were identified as totally destroyed after the earthquake. Some destroyed buildings are in close proximity to each other. The six destroyed buildings are found in four regions that are marked as a–d in Fig. 4(d). The number of other small changes exists in the analyzed scene that does not correspond to buildings.

The Trento data set [5] consists of two spotlight-mode high-resolution X-band Tandem-X and TerraSAR-X images acquired in HH polarization on January 21, 2011, and April 3, 2013, over the city of Trento, Italy (46°04′ N, 11°07′ E). The selected test site is a section (1024 × 1024 pixels), which covers the area around the Department of Engineering and Computer Science, University of Trento. Fig. 5(a) and (b) shows the prechange and postchange optical images corresponding to the area of interest. Multitemporal false-color composition of the data set (red channel:

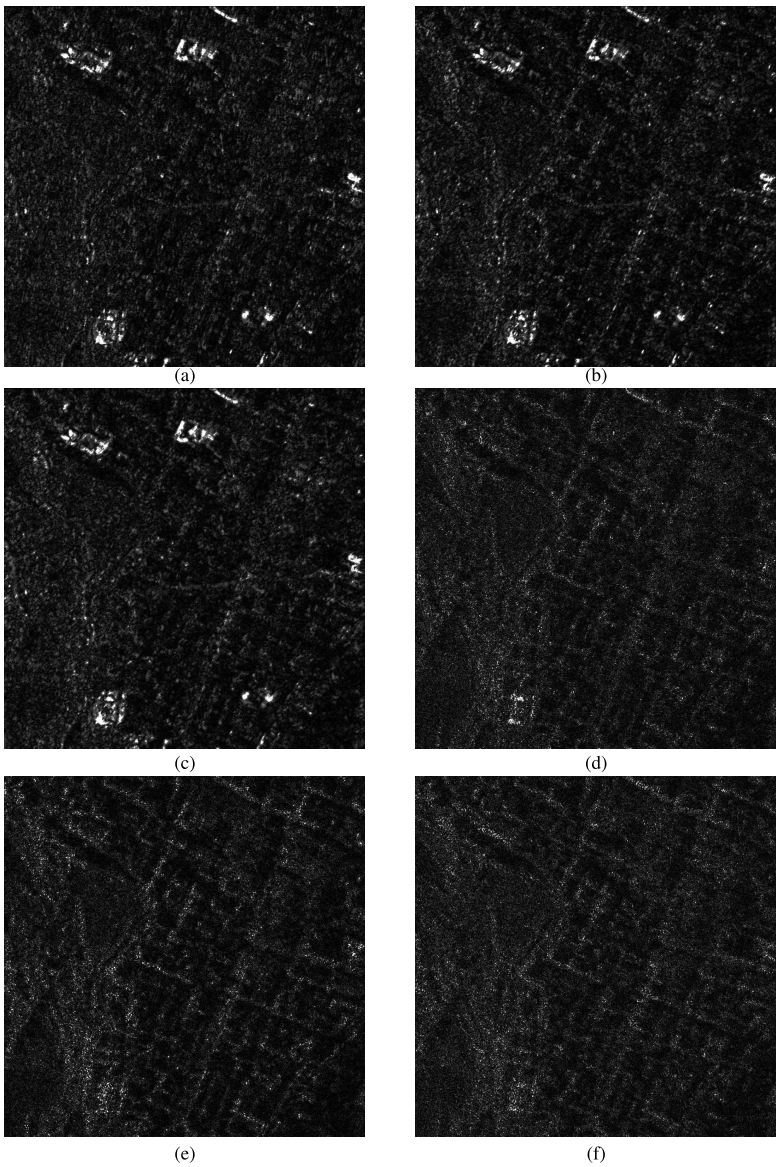


Fig. 6. Visualization of features from second convolutional layer on L'Aquila (Italy) data set. (a)–(c) Top three features. (d)–(f) Bottom three features.

TABLE I
KEY STRUCTURE OF THE GENERATOR

Layer number	Layer type	Kernel number	Kernel size
1	convolutional	64	(7,7)
2	convolutional	128	(3,3)
3	convolutional	256	(3,3)
4	residual block	256	(3,3)
5	residual block	256	(3,3)
6	residual block	256	(3,3)
7	residual block	256	(3,3)
8	residual block	256	(3,3)
9	residual block	256	(3,3)
10	residual block	256	(3,3)
11	residual block	256	(3,3)
12	residual block	256	(3,3)
13	transposed convolutional	128	(3,3)
14	transposed convolutional	64	(3,3)
15	convolutional	1	(7,7)
16	Tanh	1	0

TABLE II
KEY STRUCTURE OF THE DISCRIMINATOR

Layer number	Layer type	Kernel number	Kernel size
1	convolutional	64	(4,4)
2	convolutional	128	(4,4)
3	convolutional	256	(4,4)
4	convolutional	512	(4,4)
5	convolutional	1	(4,4)

April 3, 2013; green channel: January 21, 2011; and blue channel: April 3, 2013) is shown in Fig. 5(c). Unchanged pixels appear in grayscale, while pixels with an increase in the value of backscattering appear in magenta tone, and pixels with a decrease in the value of backscattering appear in a green tone. Three new buildings were built up in the site during the considered period [5]. A large building that is still partially under construction during the second acquisition is in the center left of the image. A medium-size building is in the left part of the image, and a small building is in the center of the image. Thus, the size of changed buildings in the data set is not homogeneous.

C. Methods for Comparison

The proposed method is compared with the state-of-the-art unsupervised building CD method proposed by Marin *et al.* [5]. As the proposed method is unsupervised and

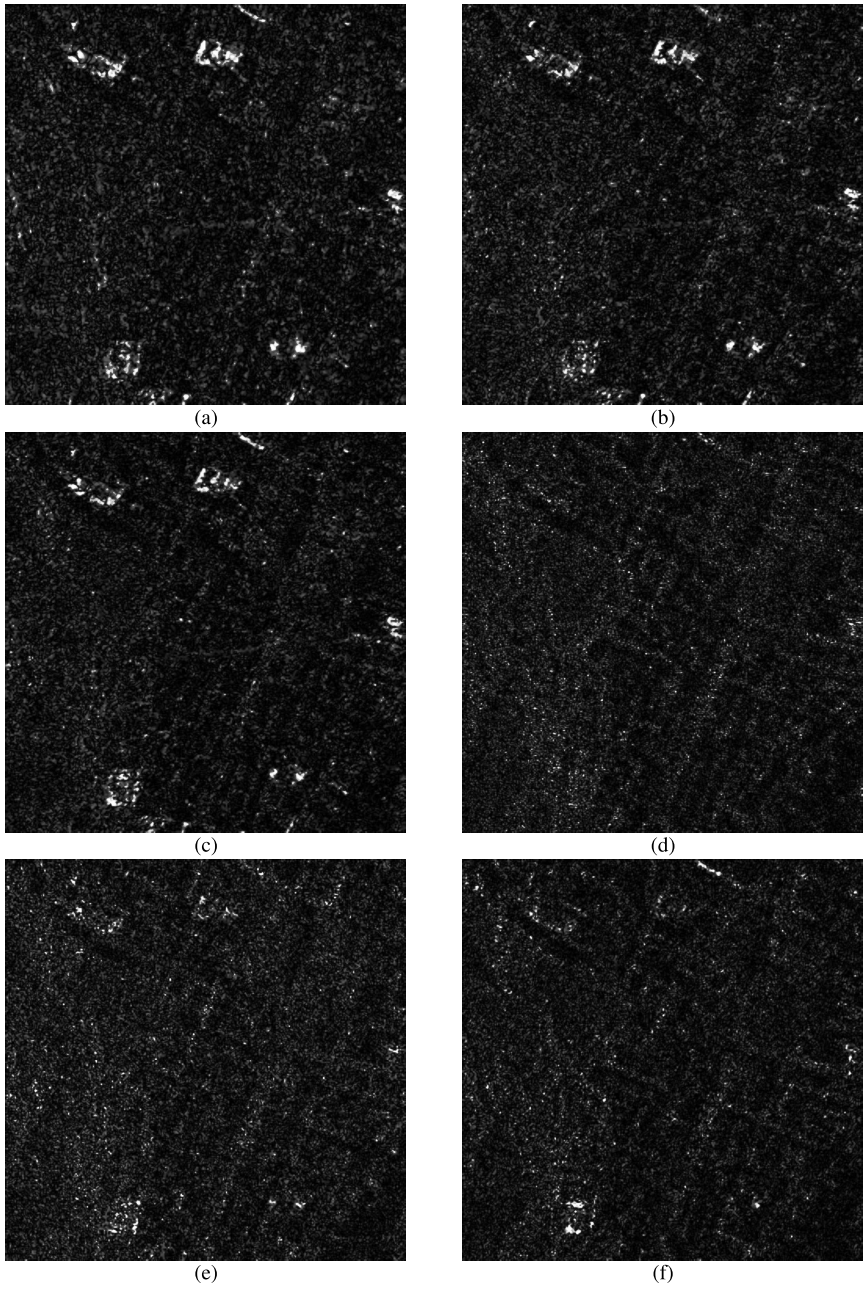


Fig. 7. Visualization of features from 3rd convolutional layer on L'Aquila (Italy) data set. (a)–(c) Top three features. (d)–(f) Bottom three features.

is not related in its objective/ design choice with respect to the state-of-the-art deep learning-based methods, they are not compared here. These are shown in more detail in the following.

1) The deep learning-based building CD methods [33], [34] are supervised. A comparison of the proposed unsupervised method with such supervised methods is unfair.
2) Most SAR CD methods in the literature are supervised [21]–[23]. Moreover, they are not designed to handle building CD.
3) The methods based on preclassification scheme [26]–[29] can work without supervision. However, their performance depends on preclassification schemes. Moreover, building change is generally sparsely distributed in the image, and hence, the generation of pseudolabeled data to further train the deep network is practically impossible.

D. Choice of Layers

Table I shows detailed structure of the CNN $\mathbf{G_{xz}}$ for bitemporal deep feature extraction. The previous works on transfer learning [4], [49], [51], [56], [57] showed that intermediate layers are more suitable for transfer learning tasks. The same is evident in the work of Ley *et al.* [42], where after SAR-to-optical transcoding-based pretraining, only the shallower layers of the generator are retained to train as a classifier. Based on this, deep features are extracted from intermediate $L = \{2, 3, 4, 5, 6\}$ layers.

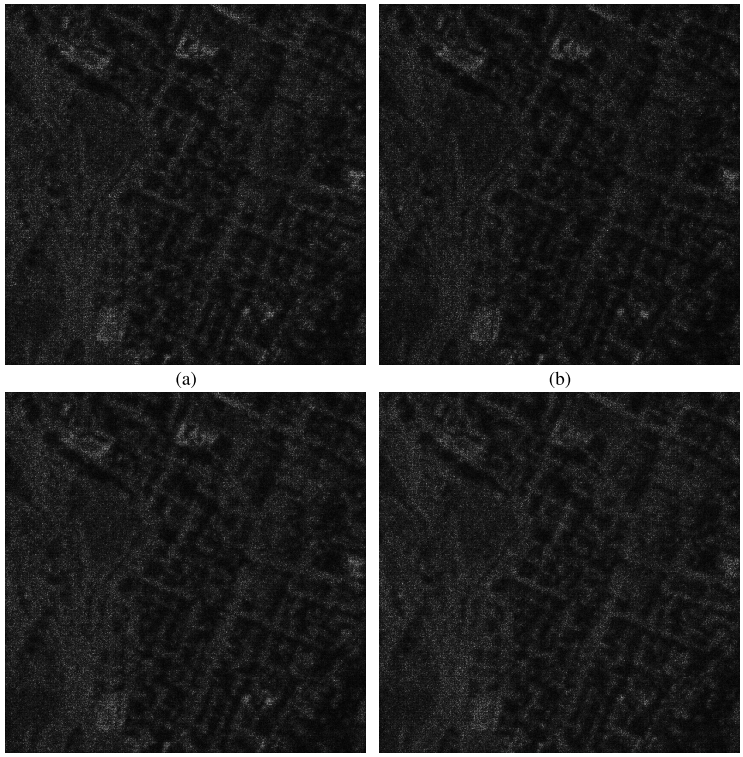


Fig. 8. Visualization of features from layer 14 on L'Aquila (Italy) data set. (a)–(c) Top three features. (d) Bottom feature.

E. Deep Feature Visulaization

Fig. 6 demonstrates the features learned in the second convolutional layer by visualizing the difference map obtained by individual features for the L'Aquila Earthquake data set. Fig. 6(a)–(c) shows the top three features selected according to the variance criterion [4]. All three features highlight the changed buildings. This shows that those features have learned semantic relevant information for building detection. Fig. 6(d)–(f) shows the bottom three features selected according to the variance criterion. They are agnostic to the building and do not highlight the changed buildings in the difference map.

The same phenomenon can be observed in the third convolutional layer (see Fig. 7). However, even the top features of the layer 14 (see Fig. 8) do not highlight the changed buildings in the difference map. This is evidence of our hypothesis in Section IV-D that deeper layers of $\mathbf{G_{xz}}$ are not suitable for deep feature extraction for CD, thus showing a poor contrast among ground features.

For the sake of brevity, deep feature visualization is shown for the L'Aquila data set only, but similar results have been obtained for the Trento data set.

F. CD Result

1) L'Aquila Data Set: After distinguishing the changed pixels from the unchanged ones, they are further clustered into two types: ξ^+ and ξ^- that are shown in Fig. 4(f) in magenta and green. Due to the disappearance of the buildings caused by the earthquake, we observe structured patterns made up of $\delta^+ \in \xi^+$ and $\delta^- \in \xi^-$. In addition, some other isolated occurrences of ξ^+ and ξ^- are observed. By following the fuzzy building detection method, destroyed buildings are identified, which are shown in Fig. 4(g). The proposed method correctly identifies all the six buildings [see Fig. 4(g)]. No false alarm is produced. Thus, the proposed method outperforms the state-of-the-art method [5] that produces one false alarm [see Fig. 4(e)]. Consistent with the previous works of SAR-based building CD [5], the result is discussed here in terms of objects (buildings) instead of pixels. The quantitative result is shown in Table III.

Furthermore, the proposed method models the shape of the destroyed buildings more accurately than [5]. This is evident for the buildings marked as "b" and "c" in the cadastral map [see Fig. 4(d)]. For better visualization of building "b," Fig. 9(a)–(c) shows the zoomed-in view of the preearthquake optical image, the result obtained by Marin *et al.* [5], and the result obtained by the proposed method, respectively.

TABLE III

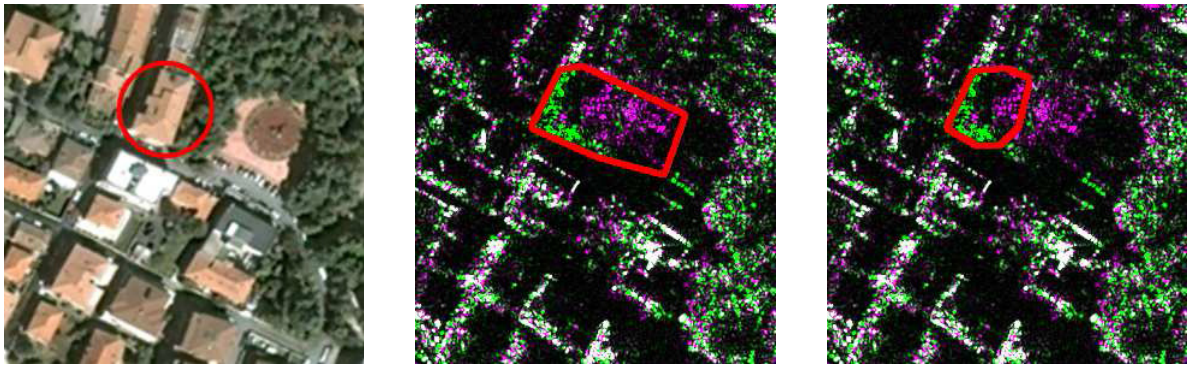
PERFORMANCE ON THE L'AQUILA DATA SET (TOTAL NUMBER OF BUILDINGS = 200)

Method	Correctly detected destroyed buildings	Missed destroyed buildings	Falsely detected destroyed buildings
Marin *et. al.* [5]	6	0	1
Proposed	6	0	0

TABLE IV

PERFORMANCE ON THE TRENTO DATA SET (TOTAL NUMBER OF BUILDINGS = 187)

Method	Correctly detected new buildings	Missed new buildings	Falsely detected destroyed buildings
Marin *et. al.* [5]	3	0	1
Proposed	3	0	0



(a) (b) (c)

Fig. 9. Zoomed-in view of the building "b" in cadastral map of L'Aquila data set. (a) Optical image (September 4, 2006) [58]. (b) Destroyed building detected by Marin *et al.* [5]. (c) Destroyed building detected by the proposed method.

2) Trento Data Set: The result obtained by the state-of-the-art method [5] is shown in Fig. 5(d). Though [5] detects all three new buildings, it misclassifies one building as destroyed. The proposed method clusters changed pixels into two types: ξ^+ and ξ^- that are shown in Fig. 5(e) in magenta and green. By following the fuzzy building detection method, new buildings are identified, which are shown in Fig. 5(f). The proposed method correctly identifies all the three new buildings, despite their inhomogeneity in size. We recall from Section IV-B that the building in the left center of the image is significantly larger than the other buildings in the scene. Moreover, some part of it is still under construction during the second acquisition, thus creating a discontinuity in the building footprint. The reconstructed footprint of this building is divided into three different parts. Nevertheless, the three parts considered jointly correctly locate the building footprint. The reconstructed building footprints are accurate for the small- and medium-size buildings. No false alarm is produced by the proposed method, despite many buildings were subject to minor renovations in the analyzed images. Thus, the proposed method outperforms [5] that misclassifies one building as destroyed. This demonstrates that the proposed method is able to work under the heterogeneous condition and can discriminate between demolished and standing buildings despite the high density of buildings (187) in the area. The quantitative result is shown in Table IV.

The proposed method takes an additional 81 seconds (averaged over ten executions) running time in comparison to [5] in a machine equipped with GPU NVidia Geforce GTX 1080 Ti and Intel I7 CPU (3.2 GHz).

V. Conclusion

In this article, an unsupervised deep learning-based method for building CD in multitemporal VHR SAR images has been proposed. CNN is known to be effective in dealing with VHR images as deep learning-based features are suitable to capture the contextual information. However, its application in unsupervised multitemporal VHR SAR analysis is limited due to the difficulty of obtaining pixelwise labeled SAR data. To address these problems, we propose a novel unsupervised CD technique that exploits the CycleGAN framework to train a SAR-optical deep transcoder using an unlabeled SAR-Optical data set that is easier to obtain compared with a pixelwise labeled SAR data set. After training the CycleGAN, a generator network is used to obtain optical-like deep features from prechange and postchange SAR images and used for CD in DCVA framework [4] originally developed for optical images. The proposed method further uses the fuzzy building detection rules [5] to identify the changed building pixels. The proposed method demonstrates that the proxy task of SAR-optical transcoding is an effective way to train a deep network for multitemporal analysis. This is an important take away considering the difficulty of labeling data in VHR SAR image analysis. Furthermore, this opens up a way of using knowledge from the unlabeled data in the unsupervised multitemporal analysis. In this fashion, an unsupervised CD method does not need to be restricted to the analyzed scene but can use the knowledge from a huge amount of remote sensing data currently being collected to process unknown scenes. Experiments conducted on a data set containing preearthquake and postearthquake images and another data set containing newly constructed buildings demonstrated the effectiveness of the proposed approach. Though demonstrated for building CD, the proposed method can be employed for other applications. By further taking advantage of the SAR-optical transcoding process, in our future work, we plan to devise a CD method for multisensor CD admitting images from both optical and SAR sensors. We also plan to extend our work for time-series analysis consisting of more than two images.

References

[1] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, Nov. 2016.

[2] M. Chini, R. Anniballe, C. Bignami, N. Pierdicca, S. Mori, and S. Stramondo, "Identification of building double-bounces feature in very high resoultion SAR data for earthquake damage mapping," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 2723–2726.

[3] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jan. 2009.

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

[5] C. Marin, F. Bovolo, and L. Bruzzone, "Building change detection in multitemporal very high resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2664–2682, May 2015.

[6] K. Jiang *et al.*, "Damage analysis of 2008 Wenchuan earthquake using SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 5, Jul. 2009, p. V-108.

[7] P. Uprety and F. Yamazaki, "Damage detection using high resolution TerraSAR-X imagery in the 2009 L'Aquila Earthquake," in *Proc. 8th Int. Workshop Remote Sens. Disaster Manage.*, vol. 9. Tokyo, Japan: Tokio Inst. Technol., 2010, pp. 1–9.

[8] F. Bovolo, C. Marin, and L. Bruzzone, "A novel hierarchical approach to change detection with very high resolution SAR images for surveillance applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 1992–1995.

[9] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.

[10] F. Chatelain, J.-Y. Tourneret, and J. Inglada, "Change detection in multisensor SAR images using bivariate gamma distributions," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 249–258, Mar. 2008.

[11] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and C. Zoppetti, "Nonparametric change detection in multitemporal SAR images based on mean-shift clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2022–2031, Apr. 2013.

[12] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98–109, Feb. 2014.

[13] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2963–2972, Dec. 2005.

[14] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 307–311, Mar. 2012.

[15] U. Soergel, U. Thoennessen, A. Brenner, and U. Stilla, "High-resolution SAR data: New opportunities and challenges for the analysis of urban areas," *IEE Proc.-Radar, Sonar Navigat.*, vol. 153, no. 3, pp. 294–300, Jun. 2006.

[16] A. R. Brenner and L. Roessing, "Radar imaging of urban areas by means of very high-resolution SAR and interferometric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2971–2982, Oct. 2008.

[17] P. T. B. Brett and R. Guida, "Earthquake damage detection in urban areas using curvilinear features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4877–4884, Sep. 2013.

[18] O. Yousif and Y. Ban, "A novel approach for object-based change image generation using multitemporal high-resolution SAR images," *Int. J. Remote Sens.*, vol. 38, no. 7, pp. 1765–1787, Apr. 2017.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[20] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 487–495.

[21] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.

[22] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, Dec. 2016.

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

[24] Y. Gao, F. Gao, J. Dong, and S. Wang, "Transferred deep learning for sea ice change detection from synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1655–1659, Oct. 2019.

[25] M. Li, M. Li, P. Zhang, Y. Wu, W. Song, and L. An, "SAR image change detection using PCANet guided by saliency detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 402–406, Mar. 2019.

[26] 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.

[27] H. M. Keshk and X.-C. Yin, "Change detection in SAR images based on deep learning," *Int. J. Aeronaut. Space Sci.*, vol. 21, pp. 549–559, Oct. 2019.

[28] M. Gong, H. Yang, and P. Zhang, "Feature learning and change feature classification based on deep learning for ternary change detection in SAR images," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 212–225, Jul. 2017.

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

[30] B. Cui, Y. Zhang, L. Yan, J. Wei, and H. Wu, "An unsupervised SAR change detection method based on stochastic subspace ensemble learning," *Remote Sens.*, vol. 11, no. 11, p. 1314, Jun. 2019.

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

[32] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[33] F. Chen and B. Yu, "Earthquake-induced building damage mapping based on multi-task deep learning framework," *IEEE Access*, vol. 7, pp. 181396–181404, 2019.

[34] L. Li, C. Wang, H. Zhang, B. Zhang, and F. Wu, "Urban building change detection in SAR images using combined differential image and residual U-Net network," *Remote Sens.*, vol. 11, no. 9, p. 1091, May 2019.

[35] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 806–813.

[36] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 44–51.

[37] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, Jan. 2017.

[38] M. Volpi and D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, Feb. 2017.

[39] Y. Wang and X. X. Zhu, "The SARptical dataset for joint analysis of SAR and optical image in dense urban area," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 6840–6843.

[40] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2016, *arXiv:1605.09782*. [Online]. Available: http://arxiv.org/abs/1605.09782

[41] S. Roy, E. Sangineto, N. Sebe, and B. Demir, "Semantic-fusion gans for semi-supervised satellite image classification," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 684–688.

[42] A. Ley, O. Dhondt, S. Valade, R. Haensch, and O. Hellwich, "Exploiting GAN-based SAR to optical image transcoding for improved classification via deep learning," in *Proc. 12th Eur. Conf. Synth. Aperture Radar*, Jun. 2018, pp. 1–6.

[43] M. F. Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "SAR-to-optical image translation based on conditional generative adversarial networks—Optimization, opportunities and limits," *Remote Sens.*, vol. 11, no. 17, p. 2067, 2019.

[44] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv:1703.10593*. [Online]. Available: http://arxiv.org/abs/1703.10593

[45] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 70, 2017, pp. 2642–2651.

[46] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," 2019, *arXiv:1906.01529*. [Online]. Available: http://arxiv.org/abs/1906.01529

[47] F. Qiu, J. Berglund, J. R. Jensen, P. Thakkar, and D. Ren, "Speckle noise reduction in SAR imagery using a local adaptive median filter," *GIScience Remote Sens.*, vol. 41, no. 3, pp. 244–266, Sep. 2004.

[48] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 770–787, May 2010.

[49] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 447–456.

[50] A. M. El Amin, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images," *Proc. SPIE*, vol. 10011, Jul. 2016, Art. no. 100110W.

[51] W. Zhang *et al.*, "Deep model based transfer and multitask learning for biological image analysis," *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 322–333, Jun. 2020.
[52] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sens.*, vol. 9, no. 9, p. 907, Aug. 2017.
[53] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, Jun. 2007.
[54] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multiscale 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.
[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[56] D.-H. Lee, Y. Lee, and B.-S. Shin, "Mid-level feature extractor for transfer learning to small-scale dataset of medical images," in *Advances in Computer Science and Ubiquitous Computing*. Singapore: Springer, 2018, pp. 8–13.
[57] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.
[58] *Google Maps*. Accessed: Apr. 10, 2020. [Online]. Available: https://maps.google.com

Sudipan Saha (Graduate Student Member, IEEE) received the Bachelor of Technology degree in electronics and communication engineering from the Institute of Engineering and Management, Kolkata, India, in 2011, and the Master of Technology degree in electrical engineering from IIT Bombay, Mumbai, India, in 2014. He is pursuing the Ph.D. degree in information and communication technologies with the University of Trento, Trento, Italy, and Fondazione Bruno Kessler, Trento.

He was an Engineer at TSMC Limited, Hsinchu, Taiwan, from 2015 to 2016. In 2019, he was a Guest Researcher with the Technical University of Munich (TUM), Munich, Germany, for three months. His research interests are related to multitemporal remote sensing image analysis, domain adaptation, time-series analysis, image segmentation, deep learning, image processing, and pattern recognition.

Mr. Saha is a reviewer of several international journals.

Francesca Bovolo (Senior Member, IEEE) received the Laurea (B.S.) degree, 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.

Until 2013, she was a Research Fellow with the University of Trento. She is the Founder and the Head of the Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler, Trento, and a member of the Remote Sensing Laboratory, Trento. She is one of the co-investigators of the Radar for Icy Moon Exploration Instrument of the European Space Agency Jupiter Icy Moons Explorer. Her research interests include remote sensing image processing, multitemporal remote sensing image analysis, change detection in multispectral, hyperspectral, synthetic aperture radar images, 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 research 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. She was the Technical Chair of the International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp 2011 and 2019). She has been the Co-Chair of the SPIE International Conference on Signal and Image Processing for Remote Sensing since 2014. She was 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 a Guest Editor of the Special Issue on Analysis of Multitemporal Remote Sensing Data of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She is also a referee of several international journals.

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

He is a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is the Founder and Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. His research interests are in the areas of remote sensing, radar and SAR, signal processing, machine learning, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the principal investigator of many research projects. Among the others, he is the Principal Investigator of the Radar for Icy Moon exploration (RIME) instrument in the framework of the JUpiter ICy moons Explorer (JUICE) mission of the European Space Agency (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 259 scientific publications in referred international journals (193 in IEEE journals), more than 330 articles in conference proceedings, and 22 book chapters.

Dr. Bruzzone has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2009, where he has been the Vice President for Professional Activities since 2019. He was ranked First Place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, in 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. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He was a co-guest editor 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 a member of the Permanent Steering Committee of this series of workshops. He is an 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 31 600) and the value of the H-index (83) (source: Google Scholar). He was invited as a keynote speaker in more than 40 international conferences and workshops. He is the Founder of the *IEEE Geoscience and Remote Sensing Magazine* for which he was the Editor-in-Chief from 2013 to 2017. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE GRSS from 2012 to 2016.