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 pre-change and post-change Very-High-spatial-Resolution (VHR) Synthetic-Aperture-Radar (SAR) images. However, multi-temporal VHR SAR images are complex as they show high spatial correlation, prone to shadows, and show inhomogeneous signature. Spatial context needs to be taken into account to effectively detect change in such images. Recently, Convolutional-Neural-Network (CNN) based transfer learning techniques have shown strong performance for CD in VHR multi-spectral images. However, its direct use for SAR CD is impeded by the absence of labeled SAR data and thus of pre-trained networks. To overcome this, we exploit the availability of paired unlabeled SAR and optical images to train for sub-optimal 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 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 bi-temporal deep feature extractor to extract optical like features from bi-temporal 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 datasets made up of pairs of bi-temporal VHR SAR images on the city of L'Aquila (Italy) and Trento (Italy).**

***Index Terms*—Change detection, Synthetic Aperture Radar, Very High Resolution images, Multi temporal images, Deep Change Vector Analysis, Generative Adversarial Network, Remote Sensing.**

I. Introduction

Change detection in 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 have been launched, which can acquire images having 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 (SAR) sensors [5], [6], [7], [8]. SAR sensors are particularly useful in the applications that require 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) which can acquire VHR images.

In the CD literature, unsupervised methods [9], [10], [11], [12] are preferred due to the difficulty of collecting multi-temporal labeled data which becomes more severe in case of post-disaster 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 like buildings show inhomogeneous signature at high resolution due to different scattering contributions from sub-objects [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 multi-temporal VHR SAR data [17], [5], [18]. Brett and Guida [17] proposed a method using curvilinear features to detect changes caused by earthquake. 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:

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

Recently, deep neural networks, especially 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 proposed for SAR CD that 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

Sudipan Saha is with Fondazione Bruno Kessler, 38123 Trento, Italy and also with 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.

Manuscript received April 10, 2020; revised May 26, 2020; accepted May 28, 2020.

(RBM) in 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 pre-classification 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 pre-classification scheme. Li *et. al.* [25] incorporated the concept of image saliency in PCANet. Li *et. al.* [26] proposed a method based on pre-classification scheme where initial pseudo labels are produced through unsupervised spatial fuzzy clustering. Similarly, [27] also uses spatial fuzzy clustering for pseudo label generation. Gong *et. al.* [28] uses sparse autoencoder to transform log-ratio difference image into a feature space and those features are clustered to generate pseudo labels. 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 convolutional-wavelet neural network (CWNN) in low resolution images. Similar to [22], the method in [23] uses a pre-classification scheme to generate training samples. In [30] a two-channel CNN is used to estimate the similarity between the bi-temporal patches. The limitations of these methods are as follows:

- Most of them are supervised, and do not account for the lack of labeled multi-temporal data [31], [32] at large scale.
- The methods relying on pre-classification do not need labeled data. However, they are still incapable of utilizing enormous amount of remote sensing data that is currently available and can be used to improve performance. Moreover, their accuracy depends on the accuracy of the pre-classification 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 multi-spectral optical satellite images by exploiting deep features extracted from a pre-trained network [38]. To apply a DCVA framework, a pixel-wise labeled VHR database (that is not multi-temporal) is used to train a deep network that is subsequently used as multi-temporal deep feature extractor. DCVA is unsupervised as it does not require any labeled multi-temporal data. However it successfully exploits the huge amount of available remote sensing data. Inspite of its success for CD in VHR optical images [4], applying DCVA for VHR SAR images is not trivial. Obtaining labeled VHR SAR dataset is very challenging owing to the difficulty of labeling VHR SAR images [39]. Thus there is a need of 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 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 forces GAN to learn deep features for distinguishing between different land surfaces. The mechanism does not require labeled data. Considering 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 sub-optimal/ill-posed as there are features in SAR images that are not present in the optical images and vice-versa [43], [42]. While SAR images emphasize 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 bi-temporal deep feature extractor in the DCVA framework. The method assumes availability of a dataset of SAR and optical images obtained from the same location or similar geographical locations, i.e., images representing similar behaviour. Different variants of GAN are available in the literature [44], [45], [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 presence of paired/co-registered 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 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 deep feature extractor from multi-temporal SAR images. The use of CycleGAN to learn transcoding between SAR and optical images:

- Helps to train a deep network without requirement of any labeled training data;
- Does not assume presence of co-registered SAR and optical images. While SAR and optical images can be potentially collected from similar geographical locations, collecting co-registered pairs is difficult. This significantly relaxes one of the strongest constraints in CD.
- Ingests the knowledge from plethora of unlabeled images used in the training process. Similar to transfer learning, while subsequently using a generator of the CycleGAN as

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 multi-temporal SAR image analysis. Recovery of optical data from SAR data has its limitations [43] [42] and is beyond the scope of this work.

Co-registered pre-change and post-change VHR SAR images are processed through the multi-layered CNN (i.e., the generator) to obtain deep features. Deep features are compared 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 our work is that we propose an unsupervised method to train a deep network that is used as a multi-temporal optical like deep feature extractor from SAR images to be processed in the DCVA framework. In contrary to [22], [27], [28], [26], proposed mechanism is completely unsupervised and can ingest knowledge from plethora of unlabeled images in the training process. Effectiveness of the multi-temporal 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 paper is organized into 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 our paper and discuss 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 dB scale.

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

Let us assume that generic datasets of unlabeled VHR SAR patches $\mathbf{X} = \{\mathbf{x_i}, \forall \mathbf{i} = \mathbf{1}, ..., \mathbf{I}\}$ and VHR optical patches $\mathbf{Z} = \{\mathbf{z_i}, \forall \mathbf{i} = \mathbf{1}, ..., \mathbf{I}\}$ are available. Considering difficulty of the SAR-optical transcoding task, we consider optical patches are panchromatic. $\mathbf{I}$ is the number of patches in the datasets. $\mathbf{X}$ and $\mathbf{Z}$ do not need to be paired/co-registered. However, they must be acquired from similar geographical locations, thus implying similar information content and image distribution. This condition significantly relaxes typical hypothesis of CD methods where images pairs have to be coregistered. Further, $\mathbf{X}$ 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 Figure 1.

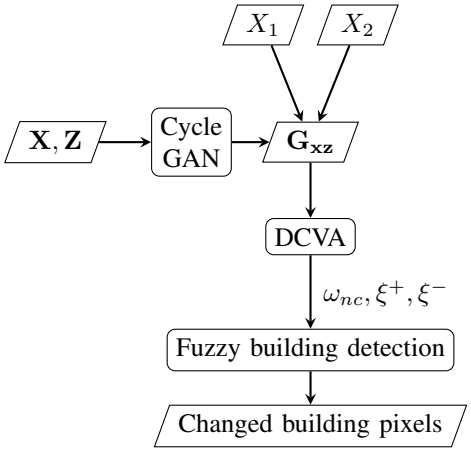


Fig. 1: Proposed CD framework

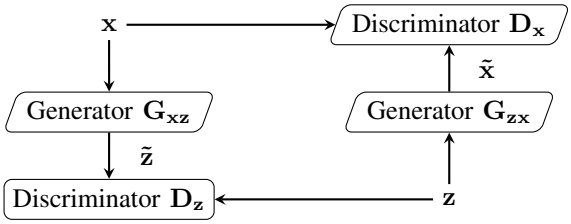


Fig. 2: CycleGAN training process

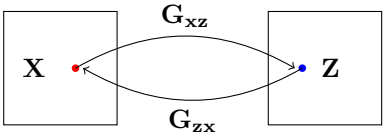


Fig. 3: Cycle-consistency constraint: the left and right rectangles represent the distribution space spanned by $\mathbf{X}$ ($p^*(\mathbf{x})$) and $\mathbf{Z}$ ($q^*(\mathbf{z})$), respectively. Drawing from the distribution $p^*(\mathbf{x})$ and processing twice through $\mathbf{G_{xz}}$ and $\mathbf{G_{zx}}$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

III. Proposed Method

The proposed method is accomplished in the following steps: i) learning transcoding between SAR and optical images using CycleGAN; ii) exploiting the CycleGAN for bi-temporal optical like deep feature extraction from SAR images; and iii) 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 datasets of unlabeled VHR SAR $\mathbf{X} = \{\mathbf{x_i}, \forall \mathbf{i} = \mathbf{1}, ..., \mathbf{I}\}$ and optical $\mathbf{Z} = \{\mathbf{z_i}, \forall \mathbf{i} = \mathbf{1}, ..., \mathbf{I}\}$ 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 non-trivial and sub-optimal 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 iterative fashion. In more details:
   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 $q^*(\mathbf{z})$. The $\mathbf{G_{zx}}$ 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, 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 $\mathbf{G_{zx}}$ 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 geo-located 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].

By 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. Figure 2 shows the adversarial learning mechanism and Figure 3 shows the cycle-consistency criterion.

B. CycleGAN based bi-temporal deep features extraction

After training CycleGAN, the weights of the generator $\mathbf{G_{xz}}$ are frozen and used as bi-temporal 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 input of any spatial size can be fed to it. Pre-change and post-change SAR images X_1 and X_2 are separately processed through $\mathbf{G_{xz}}$ to obtain optical like multi-temporal 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 from 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 change detection. The first convolutional layer of the $\mathbf{G_{xz}}$ captures primitive features like 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 towards 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 up-sampled 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 multi-scale change information from chosen layers.

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

C. Changed building detection

Components of deep change hypervector G are represented as g^d ($d = 1, ..., D$). 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 to 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 complementary classes, changed buildings and all other changes. Changed buildings (new/destroyed) generate specific signature in terms of combination of 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 change detection system as proposed by Marin *et al.* [5]. A building generates a signature that is characterized by presence of : i) backscattering contributions coming from the ground, the vertical wall, and the roof of the building (a layover area); ii) multiple scattering between the ground and the vertical wall (a double bounce line); and iii) the occlusion of the sensor due to the building (a shadow area). The appearance/disappearance of a building in the scene causes 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. Dataset and experimental results

The dataset for training CycleGAN based deep transcoder is detailed in section IV-A. Experiments were performed on two datasets described in section IV-B. Choice of the method for comparison is stated in section IV-C. Choice of layers for bi-temporal 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 details.

A. CycleGAN training

The SARptical dataset [39] that is proposed in context of urban analysis is used in this work for CycleGAN training (both $\mathbf{X}$ and $\mathbf{Z}$). The dataset 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 dataset 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 (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.

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

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)

B. CD dataset

Experiments were conducted on two datasets. One is related to the 2009 L'Aquila earthquake 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 dataset [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, pre-change image is acquired before the earthquake and the post-change image is acquired after immediate relief operation is finished. Figures 4(a), (b) show the pre-change and post-change optical images corresponding to the area of interest. A multi-temporal false color composition of the dataset (red channel: September 12, 2009; green channel: April 5, 2009; blue channel: September 12, 2009) is shown in Figure 4(c). Unchanged pixels appear in gray scale while pixels with an increase in the value of backscattering appears in magenta tone and pixels with a decrease in the value of backscattering appear in green tone. Cadastral map of the area is shown in Figure 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, b, c, and d in Figure 4(d). Number of other small changes exist in the analyzed scene that do not correspond to buildings.

Trento dataset [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 of the University of Trento. Figures 5(a), (b) show the pre-change and post-change optical image corresponding to the area of interest. A multi-temporal false color composition of the dataset (red channel: April 3, 2013; green channel: January 21, 2011; blue channel:April 3, 2013) is shown in Figure 5(c). Unchanged pixels appear in gray scale while pixels with an increase in the value of backscattering appears in magenta tone and pixels with a decrease in the value of backscattering appear in 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 dataset is not homogeneous.

C. Methods for comparison

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

1) The deep learning based building CD methods [33], [34] are supervised. Comparison of the proposed unsupervised method with such supervised methods is unfair.
2) Most SAR CD methods in the literature are supervised [21], [22], [23]. Moreover, they are not designed to handle building CD.
3) The methods based on pre-classification scheme [29], [26], [28], [27] can work without supervision. However, their performance depend on pre-classification schemes.

Moreover, building change is generally sparsely distributed in the image and hence generation of pseudo labeled 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 bi-temporal 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 pre-training, 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.

E. Deep feature visulaization

Figure 6 demonstrates the features learned in the 2nd convolutional layer by visualizing the difference map obtained by individual features for L'Aquila earthquake dataset. Figure 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. Figure 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 3rd convolutional layer (Figure 7). However, even the top features of the layer 14 (Figure 8) do not highlight the changed buildings in the difference map. This is an 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 sake of brevity, deep feature visualization is shown for L'Aquila dataset only, but similar results have been obtained for Trento dataset.

F. CD result

1) L'Aquila dataset: After distinguishing the changed pixels from the unchanged ones, they are further clustered into two types: ξ^+ and ξ^- that are shown in Figure 4(f) in magenta and green. Due to the disappearance of the buildings caused by earthquake, we observe structured patterns made up of $\delta^+ \in \xi^+$ and $\delta^- \in \xi^-$. In addition to that some other isolated occurrences of ξ^+ and ξ^- are observed. By following fuzzy building detection method, destroyed buildings are identified that are shown in Figure 4(g). The proposed method correctly identifies all the six buildings (Figure 4(g)). No false alarm is produced. Thus the proposed method outperforms the state-of-the-art method [5] that produces one false alarm (Figure 4(e)). Consistent with the previous works of SAR based building CD [5], 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 (Figure 4(d)). For better visualization of building "b", Figure 9 (a)-(c) show the zoomed view of the pre-earthquake optical image, result obtained by Marin *et al.* [5], and result obtained by the proposed method, respectively.

2) Trento dataset: The result obtained by the state-of-the-art method [5] is shown in Figure 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 Figure 5(e) in magenta and green. By following fuzzy building detection method, new buildings are identified that are shown in Figure 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 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 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 additional 81 seconds (averaged over 10 executions) running time in comparison to the [5] in a machine equipped with GPU NVidia Geforce GTX 1080 Ti and Intel I7 CPU(3.2 GHz).

V. Conclusions

In this paper, an unsupervised deep learning based method for building CD in multi-temporal 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 multi-temporal 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 CycleGAN framework to train a SAR-Optical deep transcoder using an unlabeled SAR-Optical dataset that is easier to obtain compared to a pixelwise labeled SAR dataset. After training the CycleGAN, a generator network is used to obtain optical like deep features from pre-change and post-change SAR images and used for change detection 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 deep network for multi-temporal analysis. This is an important

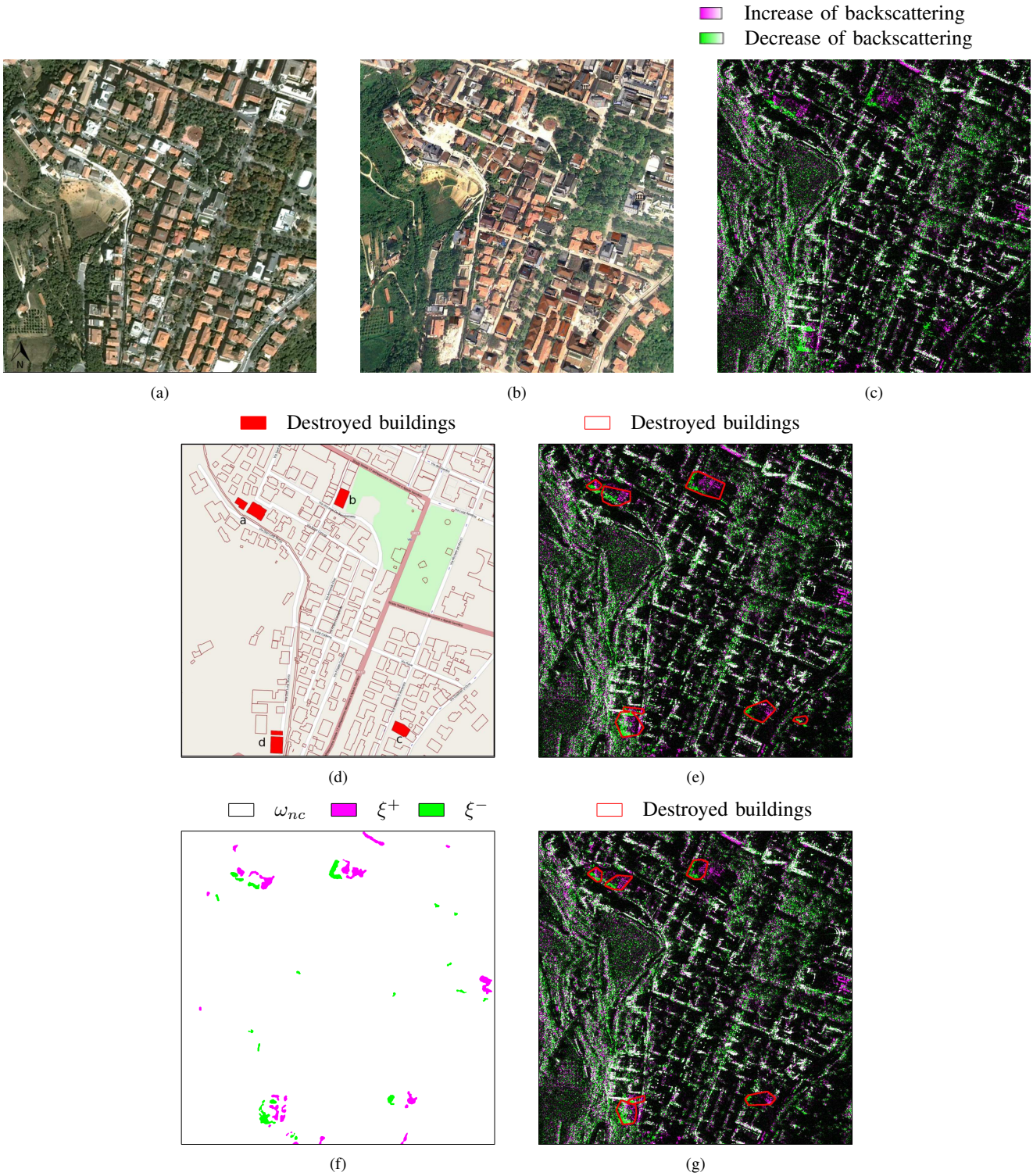


Fig. 4: L'Aquila (Italy) dataset: (a) optical image (September 4, 2006 [5]) ([58]); (b) optical image (May 8, 2009 ([58]); (c) RGB multi-temporal composition of COSMO-SKyMed images (R: September 12, 2009; G: April 5, 2009; B: September 12, 2009) (©Agenzia Spaziale Italiana, 2009. All Rights Reserved.) (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.

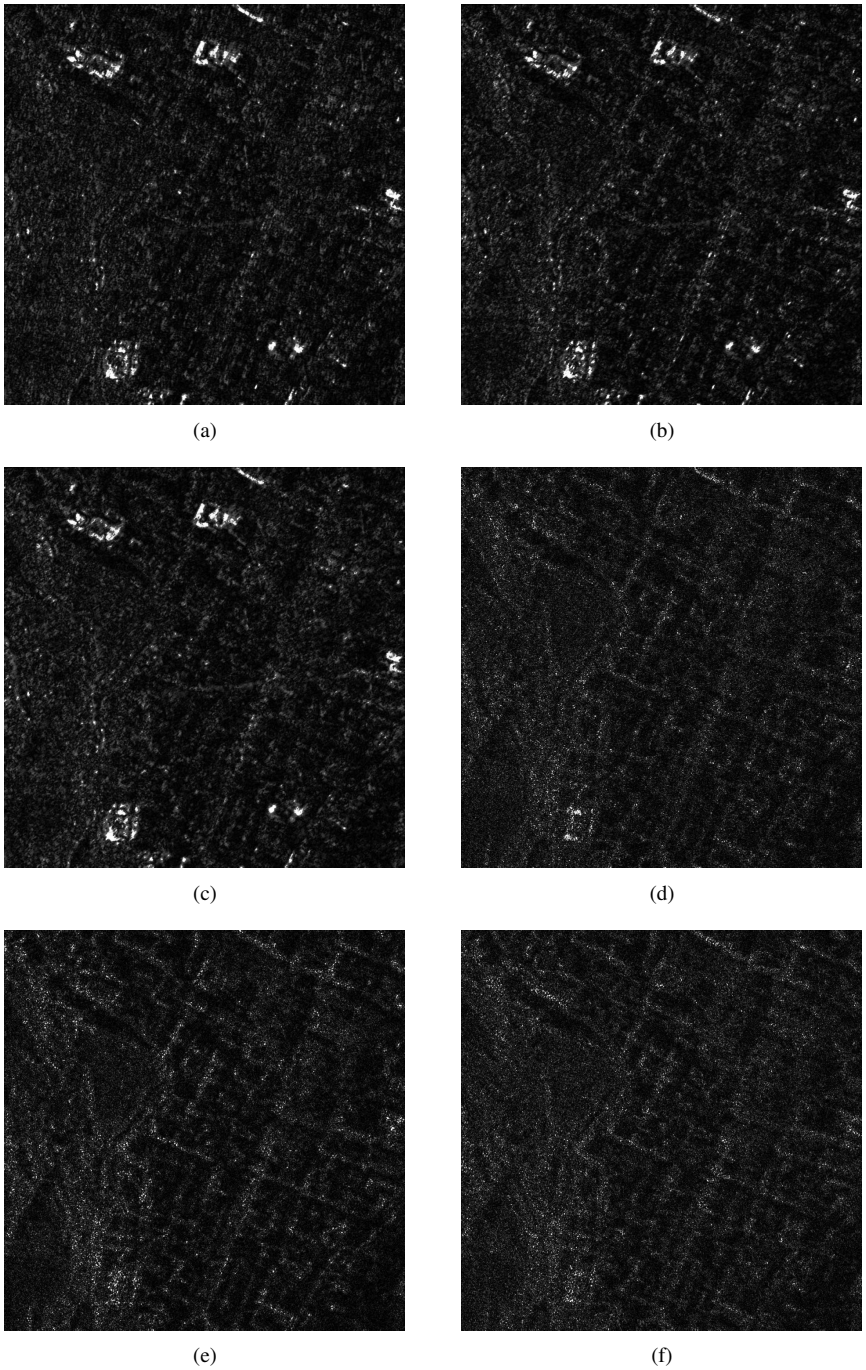
Fig. 5: Trento (Italy) dataset: (a) optical image (2011 [58]); (b) optical image (2014 [58]). (c) RGB multitemporal composition of spotlight TerraSAR-X and TanDEM-X images (R: April 3, 2013; G: January 21, 2011; B: April 3, 2013). (d) Changed buildings detected by Marin *et al.* [5]. (e) Increase and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

TABLE III: Performance on the L'Aquila dataset (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 dataset (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) (d) (e) (f)

Fig. 6: Visualization of features from 2nd convolutional layer on L'Aquila (Italy) dataset: (a)-(c) Top 3 features, (d)-(f) Bottom 3 features.

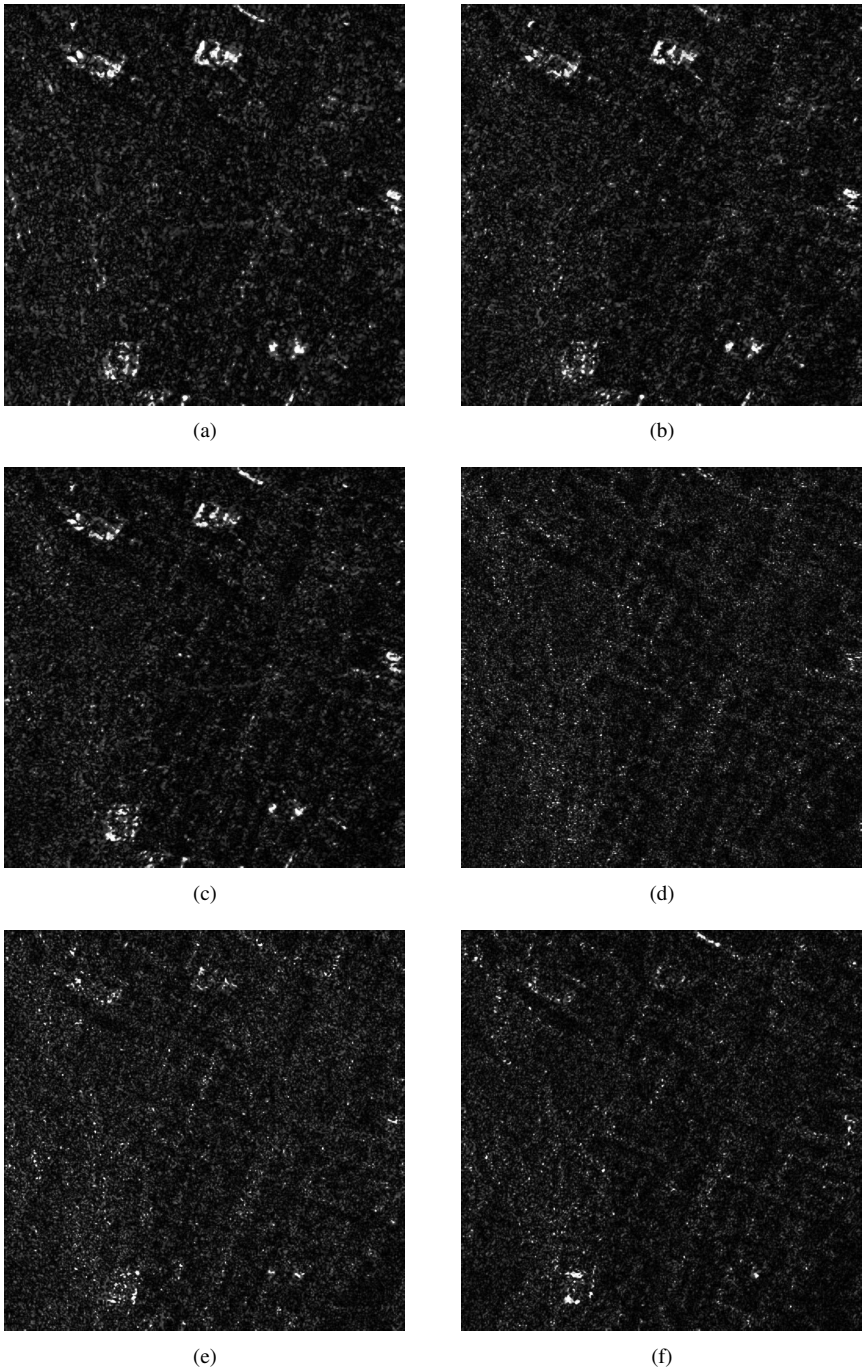
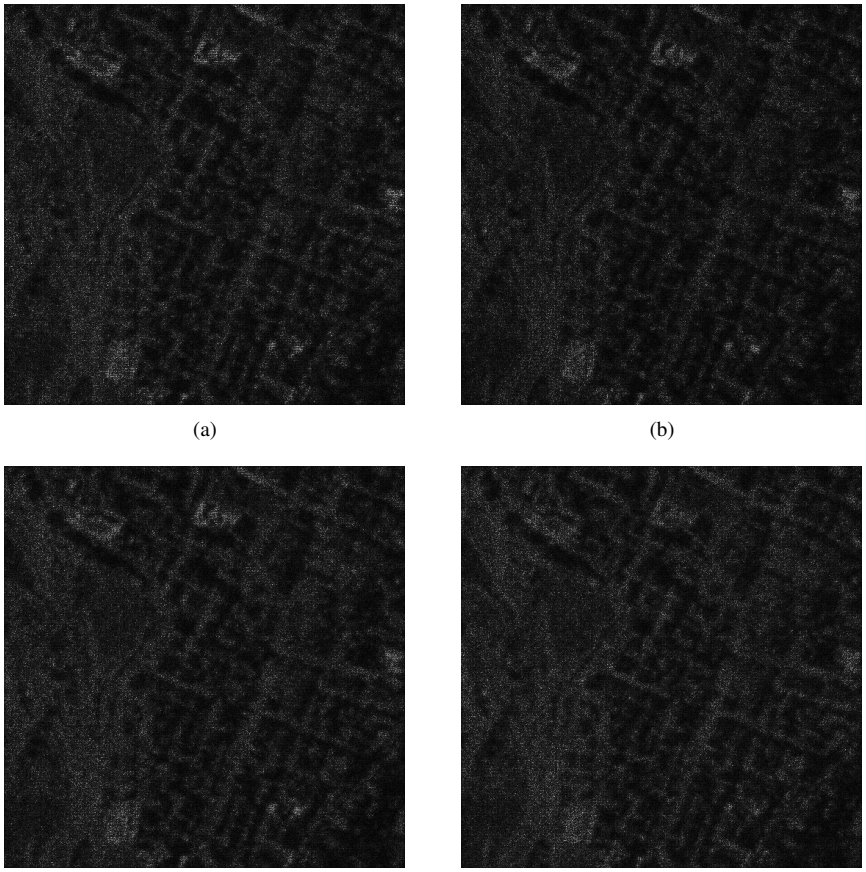


Fig. 7: Visualization of features from 3rd convolutional layer on L'Aquila (Italy) dataset: (a)-(c) Top 3 features, (d)-(f) Bottom 3 features.

take away considering difficulty of labeling data in VHR SAR image analysis. Furthermore, this opens up a way of using knowledge from the unlabeled data in unsupervised multi-temporal analysis. In this fashion, an unsupervised CD method does not need to be restricted to the analyzed scene, but can use the knowledge from huge amount of remote sensing data currently being collected to process unknown scenes. Experiments conducted on a dataset containing pre and post-earthquake images and another dataset 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 multi-sensor 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.

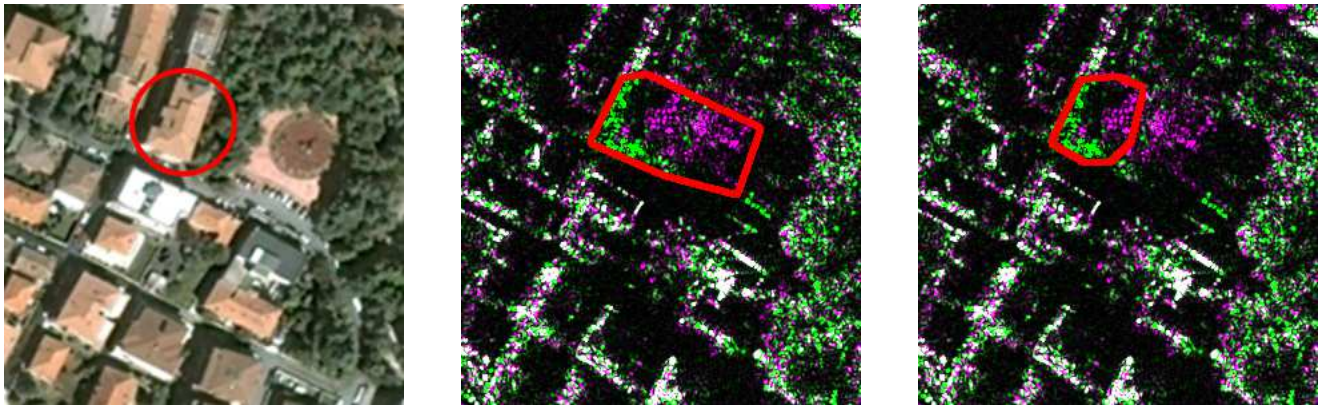
REFERENCES

[1] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proceedings of*



(a) (b) (c) (d)

Fig. 8: Visualization of features from layer 14 on L'Aquila (Italy) dataset: (a)-(c) Top 3 features, (d) Bottom feature.



(a) (b) (c)

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

the IEEE, vol. 104, no. 11, pp. 2207–2219, 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 *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, 2015, pp. 2723–2726.

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

[4] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3677–3693, 2019.

[5] C. Marin, F. Bovolo, and L. Bruzzone, "Building change detection in multitemporal very high resolution SAR images," *IEEE transactions on geoscience and remote sensing*, vol. 53, no. 5, pp. 2664–2682, 2015.

[6] K. Jiang, C. Wang, H. Zhang, W. Chen, B. Zhang, Y. Tang, and F. Wu, "Damage analysis of 2008 Wenchuan earthquake using SAR images," in *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, vol. 5. IEEE, 2009, pp. V–108.

[7] P. Uprety and F. Yamazaki, "Damage detection using high resolution TerraSAR-X imagery in the 2009 L'Aquila earthquake," in *8th International workshop on remote sensing for disaster management, Tokyo, Tokio Institute of Technology*, vol. 9, 2010.

[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 *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 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 Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874–887, 2005.

[10] F. Chatelain, J.-Y. Tourneret, and J. Inglada, "Change detection in multisensor sar images using bivariate gamma distributions," *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 249–258, 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 transactions on geoscience and remote sensing*, vol. 51, no. 4, pp. 2022–2031, 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 Transactions on Fuzzy Systems*, vol. 22, no. 1, pp. 98–109, 2013.

[13] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2963–2972, 2005.

[14] M. Gong, Y. Cao, and Q. Wu, "A neighborhood-based ratio approach for change detection in sar images," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 2, pp. 307–311, 2011.

[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 Proceedings-Radar, Sonar and Navigation*, vol. 153, no. 3, pp. 294–300, 2006.

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

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

[18] O. Yousif and Y. Ban, "A novel approach for object-based change image generation using multitemporal high-resolution SAR images," *International journal of remote sensing*, vol. 38, no. 7, pp. 1765–1787, 2017.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in neural information processing systems*, 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 transactions on neural networks and learning systems*, vol. 27, no. 1, pp. 125–138, 2015.

[22] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1792–1796, 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 Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1240–1244, 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 Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1655–1659, 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 Geoscience and Remote Sensing Letters*, vol. 16, no. 3, pp. 402–406, 2018.

[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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5751–5763, 2019.

[27] H. M. Keshk and X.-C. Yin, "Change detection in SAR images based on deep learning," *International Journal of Aeronautical and Space Sciences*, pp. 1–11, 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 Journal of Photogrammetry and Remote Sensing*, vol. 129, pp. 212–225, 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 transactions on neural networks and learning systems*, vol. 29, no. 3, pp. 545–559, 2016.

[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 Sensing*, vol. 11, no. 11, p. 1314, 2019.

[31] F. Bovolo and L. Bruzzone, "The time variable in data fusion: a change detection perspective," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 8–26, 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 Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[33] F. Chen and B. Yu, "Earthquake-induced building damage mapping based on multi-task deep learning framework," *IEEE Access*, vol. 7, pp. 181 396–181 404, 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 Sensing*, vol. 11, no. 9, p. 1091, 2019.

[35] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 806–813.

[36] O. A. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 44–51.

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

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

[39] Y. Wang and X. X. Zhu, "The SARptical dataset for joint analysis of SAR and optical image in dense urban area," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 6840–6843.

[40] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.

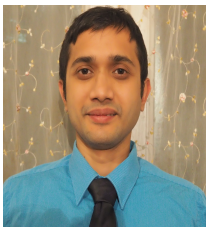
[41] S. Roy, E. Sangineto, N. Sebe, and B. Demir, "Semantic-fusion GANs for semi-supervised satellite image classification," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 684–688.

[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 *EUSAR 2018; 12th European Conference on Synthetic Aperture Radar*. VDE, 2018, pp. 1–6.

[43] M. Fuentes 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 Sensing*, 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," *arXiv preprint*, 2017.

[45] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. 34th International Conference on Machine Learning (ICML)-Vol. 70*. JMLR. org, 2017, pp. 2642–2651.
[46] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks: A survey and taxonomy," *arXiv preprint arXiv:1906.01529*, 2019.
[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 Sensing*, vol. 41, no. 3, pp. 244–266, 2004.
[48] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 5, pp. 770–787, 2009.
[49] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 447–456.
[50] A. M. El Amin, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images," in *First International Workshop on Pattern Recognition*. International Society for Optics and Photonics, 2016, pp. 100 110W–100 110W.
[51] W. Zhang, R. Li, T. Zeng, Q. Sun, S. Kumar, J. Ye, and S. Ji, "Deep model based transfer and multi-task learning for biological image analysis," *IEEE transactions on Big Data*, 2016.
[52] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for sar target classification with limited labeled data," *Remote Sensing*, vol. 9, no. 9, p. 907, 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 Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1658–1670, 2007.
[54] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multi-scale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4124–4137, 2017.
[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[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*. 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717–1724.
[58] Google maps. [Online]. Available: https://maps.google.com



Sudipan Saha (S'16) received the Bachelor of Technology degree in Electronics and Communication engineering from Institute of Engineering & Management, Kolkata, India in 2011 and Master of Technology degree in Electrical Engineering from Indian Institute of Technology Bombay, Mumbai, India in 2014. He worked as an engineer at TSMC Limited, Taiwan (2015 - 2016). He is currently a Ph.D. student in Information and Communication Technologies at the University of Trento, Trento, Italy and Fondazione Bruno Kessler, Trento, Italy. In 2019, he was a guest researcher at Technical University of Munich (TUM), Munich, Germany for 3 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. He is a reviewer for several international journals.

Francesca Bovolo (S'05–M'07-SM'13) 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 currently 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 CoChair 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 for several international journals.

Lorenzo Bruzzone (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, and digital communications.

Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current 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 259 scientific publications in referred international journals (193 in IEEE journals), more than 330 papers in conference proceedings, and 22 book chapters. He is editor/co-editor of 18 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven from the total number of citations (more than 31600) and the value of the h-index (83) (source: Google Scholar). He was invited as 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 is Vice-President for Professional Activities. Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since that he was 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. Dr. Bruzzone was a Guest Co-Editor of many Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal 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 Editor-in-Chief between 2013-2017.Currently he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has been Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society between 2012-2016.