Supervised Change Detection Using Prechange Optical-SAR and Postchange SAR Data

Sudipan Saha, *Member, IEEE*, Muhammad Shahzad, *Senior Member, IEEE*, Patrick Ebel, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*— Change detection using satellite/aerial images is used to quantify the impacts of many natural and man-made disasters. At the occurrence of such events, both prechange optical and synthetic aperture radar (SAR) images can be obtained by going back in time. However, the availability of the postchange optical image is often hindered by the presence of artifacts like clouds. To circumnavigate this, we propose a novel change detection data setting that uses both optical and SAR images prechange, yet only SAR imagery postchange. For this challenging scenario, we propose a Siamese network that processes the prechange and postchange SAR inputs using a shared set of weights, while the prechange optical input is processed using a network that do not share the weights with the SAR inputs. The encoded weights from the three networks are fused and finally decoded using a common decoder to obtain the change map. Our model effectively fuses multisensor information and can obtain satisfactory result despite the absence of the postchange optical image. Experimental results on a multisensor urban dataset demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Change detection (CD), fusion, multisensor analysis, optical images, Siamese network, synthetic aperture radar (SAR).**

I. Introduction

LAND surface of our Earth is continuously evolving, caused by anthropogenic and natural reasons. Rapid climate change is causing more and more disaster, e.g., fire events [1], floods [2], dam disasters [3], altered vegetation response [4], landslides [5], and earthquakes [6]. Furthermore, events like wars occasionally cause large-scale destruction and human displacement [7]. Aerial and satellite image-based change detection (CD) methods [8] are used to quantify the impact of such events. Optical/multispectral images are used for most such CD applications [9] as they provide visual cues easily comprehensible by us. While synthetic aperture radar (SAR)-based CD methods are also proposed in the literature, they are generally designed to detect changes in specific objects that show structural changes, e.g., buildings [10].

Optical images are severely impacted by artifacts like clouds, fogs, and smokes. Clouds are frequent in some parts of the Earth, depending on latitude and local climate [11], [12], [13]. Furthermore, some incidents (e.g., volcanic eruptions, wildfires, wartime bombing) may themselves induce smoke, thus further reducing the chance of obtaining a clear optical image. Availability of postchange artifact-free optical image may impact the postchange response time. As an example, we may consider the case of recent armed conflict in Ukraine. The first cloud-free optical (Sentinel-2) acquisition of the city of Mariupol after the armed conflict started on February 24, 2022, was only available on March 14, 2022. However, for most applications, even if the immediate prechange optical image is cloudy, it is possible to go back further in time to obtain a suitable prechange optical image.

The SAR sensors are negligibly impacted by the presence of artifacts [14], [15]. This allows the SAR sensors to be used at any weather conditions and any time of the day [16], [17]. However, side-looking geometry and inherent characteristics of SAR images, such as speckle and layover/foreshortening, makes learning discriminative features more challenging. Thus, SAR images generally obtain lower classification accuracy in comparison to the artifact-free optical images for most common remote sensing datasets [18]. Motivated by the complementary nature of their information, some existing works have discussed the importance of fusing optical and SAR data for CD [17], [19]. Saha et al. [20] and Wan et al. [21], [22] consider the case where only prechange optical and postchange SAR images are available. However, in most applications there is no impediment in collecting prechange SAR images. Thus, a more practical data scenario would be using prechange optical image and both prechange and postchange SAR images.

Manuscript received 21 May 2022; revised 30 June 2022 and 10 August 2022; accepted 9 September 2022. Date of publication 23 September 2022; date of current version 28 September 2022. This work was supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme Grant ERC-2016-StG-714087 (Acronym: *So2Sat*), in part by the Helmholtz Association through the Framework of Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research" (W2-W3-100), in part by the German Federal Ministry of Education and Research (BMBF) in the framework of the International Future AI Lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" under Grant 01DD20001, and in part by the German Federal Ministry of Economics and Technology in the framework of the "National Center of Excellence ML4Earth" under Grant 50EE2201C. *(Corresponding author: Xiao Xiang Zhu.)*

Sudipan Saha, Patrick Ebel, and Xiao Xiang Zhu are with the Department of Aerospace and Geodesy, Chair of Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany (e-mail: sudipan.saha@tum.de; patrick.ebel@tum.de; xiaoxiang.zhu.ieee@gmail.com).

Muhammad Shahzad is with the Department of Aerospace and Geodesy, Chair of Data Science in Earth Observation, Technical University of Munich, 85521 Munich, Germany, on leave from the School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan (e-mail: muhammad.shahzad@tum.de).

Code is available at https://gitlab.lrz.de/ai4eo/cd/-/tree/main/optSarSarCd.

Digital Object Identifier 10.1109/JSTARS.2022.3206898

Siamese networks [23], proposed first in the context of image matching [24], [25], are popular in supervised CD [26], [27], [28]. They generally consist of two weight-sharing networks that ingest two different inputs. Considering the inputs use the same modality, weight sharing ensures that the network produces similar feature representation for the two inputs. However, optical and SAR inputs show different characteristics that need to be processed using different networks [29]. Furthermore, in our case, prechange input consists of two different modalities (optical and SAR), while the postchange input contains only SAR data. Thus, this poses an additional challenge of different number of prechange and postchange inputs. Considering these specifications, we design a modified version of the Siamese network that shares weight for the prechange and postchange SAR inputs; however, prechange optical input is processed using a different set of weights. Features learned by the sensor-specific encoders are concatenated and further processed using a decoder.

The contributions of this work are as follows.

1) We introduce the (prechange) optical-SAR and (postchange) only SAR data setting in the context of supervised CD.
2) We propose a supervised Siamese network that can handle different number of prechange and postchange inputs.
3) We show results on a multisensor Sentinel2-Sentinel1 dataset [17] showing the benefits of the proposed data setting and method. We also carry out an additional analysis showing the relative importance of the three inputs (prechange optical, prechange SAR, and postchange SAR).

The rest of this article is organized as follows. Existing works related to multisensor CD are briefly reviewed in Section II. Section III describes the proposed method. We detail the experimental results in Section IV. Finally, Section V concludes this article.

II. Related Work

In this section, we briefly discuss the existing works on SAR and multisensor CD methods.

A. CD in SAR Images

SAR-based CD is highly challenging owing to the issues pertaining to intrinsic speckle noise and deformation sensitivity. To cope with these issues, various approaches have been proposed. For instance, Gong et al. [30] presented an approach based on joint classification of prechange and postchange SAR images. Deep transfer learning is proposed for building CD using SAR images in [10]. Zhang et al. [31] proposed a deep spatiotemporal gray-level co-occurrence aware CNN architecture that takes the 3-D gray-level co-occurrence matrix as an auxiliary feature to better capture the neighborhood relationship. Gao et al. [32] proposed a multiscale capsule network that is able to capture the discriminative characteristics between the changed and unchanged pixels. Similarly, Wang et al. [33] proposed an end-to-end graph level neural network architecture to robustly extract the local neighborhood for more discriminative graph learning for CD. Variants for Siamese networks have also been proposed for SAR CD [34], [35].

B. Multisensor CD

Several architectures have been proposed to perform CD with prechange and postchange images from same sensor [36], [37]. However, there are very few approaches that can tackle the CD problem based on bitemporal multisensor inputs [20], [21], [38], [39], [40]. The reason being that the multisensor images are affected by the differences in spatial resolution [38] and spectral characteristics of the sensors [41]. A straightforward approach to handle multisensor input is to first independently derive the classification/segmentation maps for multisensory images and later compare these maps to extract changed regions [21]. However, such postclassification approaches are prone to errors. The other popular approach is to project the prechange and postchange images into a common feature space such that they become comparable in this new feature space [42], [43]. Such projection into common feature space can be achieved using different ways, e.g., by using generative adversarial network [43], by homogeneous pixel transformation [42], and by using self-supervised learning [20]. Another approach is to learn a mapping function between the prechange and postchange images [38]. Like same-sensor CD, symmetric and Siamese deep neural networks have also been used for multisensor CD. In [39], an approximately symmetric deep neural network was used to project the images into same feature space. Wang et al. [44] proposed a deep CNN-based Siamese network with a hybrid convolutional feature extraction module using multisensor images. Finally, Ebel et al. [17] proposed a novel Siamese architecture for fusion of SAR and optical observations for multimodal CD.

C. Siamese Networks

Due to their capability to process input and output using same set of weights, Siamese networks are preferred in many supervised CD applications. In one of the first works, Zhan et al. [26] proposed a Siamese network for CD in optical aerial images. A Siamese network was used for patch based CD in [45]. Effectiveness of the Siamese networks for CD in Sentinel-2 images was shown in [37]. Zhang et al. [46] proposed a Siamese network for multimodal CD. More complex networks have been designed, e.g., by combining Siamese networks with recurrent neural networks [47]. As already discussed in Sections II-A and II-B, Siamese networks have also been adopted for SAR CD and fusing multisensor CD.

Our work is related to the works in Section II-A, as we also use SAR images as the main data source for CD. However, we also use prechange optical image, which makes our work relatable to Section II-B. Furthermore, we use Siamese networks, as in the works in Section II-C.

III. Proposed Method

When prechange and postchange images are acquired using the same sensor, Siamese (i.e., weight-sharing) networks are popularly used for CD [26], [45]. Following the benefits and practicality of Siamese networks in supervised CD, our method adapts it in our novel challenging data scenario. However, to characterize the differences of multisensor (optical and SAR) input, only prechange and postchange SAR inputs are processed

using weight-sharing networks, as detailed in Section III-A. Optical and SAR images show strong difference, both in terms of characteristics and input bands. Thus, to learn better the sensor-specific features, the optical and SAR inputs are processed using separate encoders. Furthermore, skip connections are used between encoder and decoder modules to propagate fine-grained input specific details, as detailed in Section III-B. The network is trained using weighted cross entropy, as explained in Section III-C. Test time adaptation (see Section III-D) is further applied to enhance the performance of the proposed model.

A. Modality-Specific Weight Sharing Network

Siamese networks are popular in remote sensing CD, as discussed in Section II-C. Generally, prechange and postchange inputs are images acquired using the same sensor with same number of channels. To process such inputs, a Siamese network consists of twin weight-sharing networks. However, in our case only, the SAR data are common in the prechange and postchange inputs. In addition, the prechange input also has optical input that is missing in the postchange input. Thus, we propose a modality-specific weight sharing scheme where prechange and postchange SAR images are processed using weight sharing twin encoders, $\mathbf{E}_{\mathrm{SAR1}}$ and $\mathbf{E}_{\mathrm{SAR2}}$. Furthermore, the prechange optical input is processed using a separate encoder, $\mathbf{E}_{\mathrm{OPT1}}$. Number of features increase in each layer of the encoder and output of the encoder (bottleneck) is the widest part of the network. The bottleneck representations obtained by $\mathbf{E}_{\mathrm{SAR1}}$, $\mathbf{E}_{\mathrm{SAR2}}$, and $\mathbf{E}_{\mathrm{OPT1}}$ are concatenated to form a unified bottleneck representation that is further processed using a decoder network. Using a series of layers, the decoder network reprojects the bottleneck representation to the same spatial dimension as input patch with only two outputs at the final layer, corresponding to changed/unchanged.

We design each encoder using ten convolution layers. Each convolution layer uses kernels of spatial size 3*3 and a stride of 1 pixel. Postprocessing units (batch normalization, rectified linear unit, and dropout) follow the convolution layers. Furthermore, a max pooling (2×2) follows each convolution layer, thus shrinking the spatial size as we progress through the encoder. The decoder unit consists of 14 transposed convolution layers that are also followed by the same postprocessing units as the input. A schematic of the proposed architecture is shown in Fig. 1.

B. Skip Connections From Encoder to Decoder

U-Net [48], a popular semantic segmentation architecture, uses skip connections between encoder and decoder to propagate the fine-grained details learned in the encoder part to construct an image in the decoder part. Following this, we pass the features from the three encoders ($\mathbf{E}_{\mathrm{SAR1}}$, $\mathbf{E}_{\mathrm{SAR2}}$, and $\mathbf{E}_{\mathrm{OPT1}}$) and connect them to the appropriate depth of the decoder. We argue that in this fashion, we are able to retain the fine-grained input-specific details while decoding the bottleneck representation to obtain the final output map. Skip connections are shown as "concat" units in Fig. 1.

C. Weighted Loss

Generally, the unchanged pixels are significantly more frequent than the changed pixels in the training data. To account for this, we use weighted cross-entropy loss to train the model $\mathcal{M}$ on the training data. Weighted cross-entropy loss is a variant of the cross-entropy loss function weighted by class, varying the relative penalty of a probabilistic false negative for an individual class [49], [50]. In our case, we deal with two classes—changed and unchanged, weights of which (β_c and β_{nc}) are derived from their inverse ratio in the training data. For a certain pixel x, given reference label whether the pixel is changed as y and softmax output corresponding to the changed class as $p(x)$, loss function is given as

$$\ell(x) = -\beta_c y \log(p(x)) - \beta_{\mathrm{nc}}(1-y)\log(1-p(x)). \quad (1)$$

Weighted loss is aggregated over all samples in a training mini-batch [51].

D. Test Time Adaptation

The model trained in the above step may be suboptimal for applying on the test regions unseen during training time [52], [53]. This is furthermore challenging in the considered data scenario, as the postchange optical image is not seen in our case. Dong et al. [54] showed that the model activation's mean can be used to effectively model the domain differences. Inspired by this, we propose a simple test time adaptation strategy to adapt the trained model to the test cities. For each pixel, the model $\mathcal{M}$ produces two unnormalized predictions (also called logits) θ_{uc} and θ_c, corresponding to the unchanged and changed class, respectively. Let the mean of θ_{uc}, estimated on the entire training data, be $\theta_{\mathrm{uc}}^{Tr}$. For a given test patch, if the mean of θ_{uc} is estimated as $\theta_{\mathrm{uc}}^{Te}$, then the θ_{uc} values for each pixel in this patch are added to $(\theta_{\mathrm{uc}}^{Tr} - \theta_{\mathrm{uc}}^{Te})/\theta_{\mathrm{uc}}^{Tr}$. This ensures that the neural mean activations are similar for the training and test data. The impact of this adaptation may vary depending on the size of the test patch on which the adaptation is performed. Let say a given test scene is divided into $p \times p$ patches. A smaller value of p performs adaptation only at a global scale, while a larger value of p implies adaptation at more local scale.

IV. Experimental Validation

A. Dataset

The Onera Satellite CD (OSCD) dataset [55] is a popular urban CD dataset. While originally proposed for only optical CD, a multisensor version of this is available in [17]. The dataset uses ascending orbit Sentinel-1 SAR observations coordinate-transformed via GDAL [56] to match the coordinate system of the original optical data.

The dataset consists of 24 cities distributed across the world. Originally, the OSCD dataset [17], [55] and works using this dataset [57] split 24 cities into a training set of 14 cities and a test set of ten cities, with no validation set. However, the use of separate validation set to optimize the hyperparameters is generally recommended in deep learning [58]. So, departing

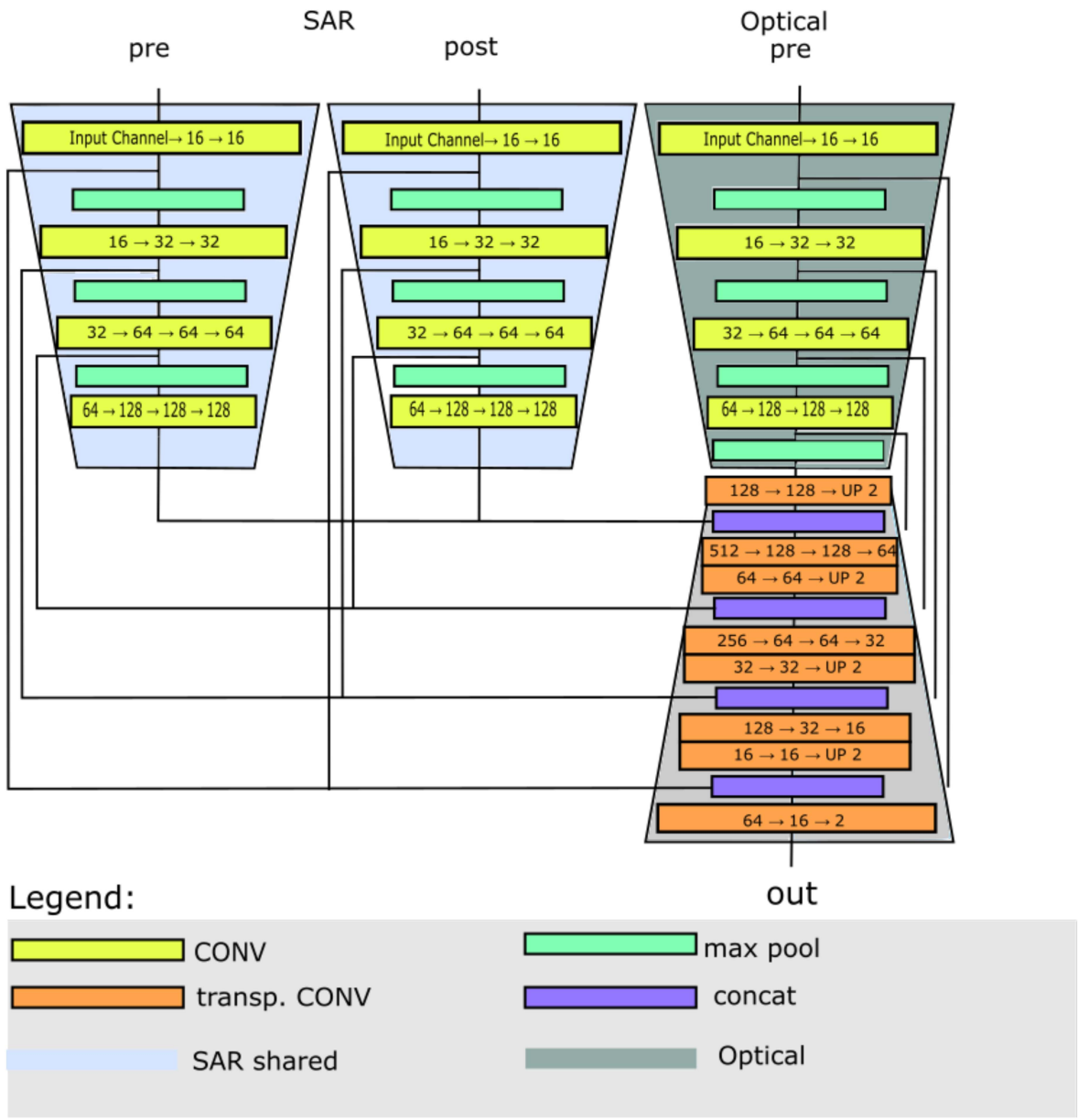

Fig. 1. Proposed Architecture. A $->$ B $->$ C refers to 2 Conv. (or transposed conv.) layers that projects feature dimension from A to B and B to C, respectively. UP refers to the spatial upsampling.

from previous usual practice w.r.t. this dataset, we divide the 24 cities into three sets:

1) *Training:* Aguasclaras, Beihai, Beirut, Bercy, Bordeaux, Hong Kong, Nantes, Abudhabi, Rennes, Saclay E, Pisa, Mumbai.
2) *Validation:* Cupertino, Paris.
3) *Test:* Brasilia, Chongqing, Dubai, Lasvegas, Milano, Montpellier, Norcia, Rio, Saclay W, Valencia.

B. Metrics and Settings

To measure the performance, we use precision (TP/(TP+FP)), recall (TP/(TP+FN)), F1 score (harmonic mean of the precision and recall), and accuracy, where TP indicates true positive, FP indicates false positive, and FN indicates false negative.

We set the number of training epochs by inspecting performance evolution on validation subset (see Section IV-A). Models are trained for 50 epochs with a learning rate of 0.0001 and Adam optimizer [60]. Result of the proposed method is shown as an average of three runs with random seeds.

C. Compared Methods and Ablation Studies

To verify the effectiveness of the proposed setting and method, we need to investigate the following two aspects:

1) whether the proposed data setting (prechange: optical+SAR, postchange: SAR) is outperformed by more simpler data settings (prechange: SAR, postchange: SAR) or (prechange: optical, postchange: SAR) or (prechange: SAR, postchange: optical);

TABLE I
CD RESULTS FOR PROPOSED METHOD/SETTING VERSUS EXISTING METHODS/SETTINGS

Method	Data setting	Precision	Recall	Accuracy	F1*100
Early fusion [37]	Pre: SAR, post: SAR	0.14	0.61	79.58	22.88
Siamese	Pre: SAR, post: SAR	0.17	0.36	88.19	23.08
Early fusion [37]	Pre: optical, post: SAR	0.10	0.60	71.26	17.13
Encoder-decoder [59]	Pre: optical, post: SAR	0.07	0.34	73.55	11.61
Early fusion [37]	Pre: SAR, post: optical	0.07	0.46	66.88	12.18
Encoder-decoder [59]	Pre: SAR, post: optical	0.07	0.24	79.99	10.87
Early fusion [37]	Pre: optical+SAR, post: SAR	0.22	0.39	89.70	27.72
Encoder-decoder [59]	Pre: optical+SAR, post: SAR	0.08	0.38	75.11	13.64
Vanilla Siamese	Pre: optical+SAR, post: SAR	0.18	0.40	88.09	25.24
Proposed (w/o test-time adaptation)	Pre: optical+SAR, post: SAR	0.22	0.47	89.13	30.25
Proposed	Pre: optical+SAR, post: SAR	0.27	0.41	91.28	32.14

2) whether the proposed Siamese architecture can be outperformed by other popular CD architectures, e.g., fully convolutional network or Vanilla Siamese.

We investigate the abovementioned two aspects by comparing to the following methods (first six are for the first aspect and the remaining are for the second aspect).

1) An Early Fusion fully convolutional network using only prechange and postchange SAR images. Input to the network is a stacked version of the prechange and postchange SAR data.
2) Data setting same as above, however using a Siamese network similar to the proposed approach.
3) An early fusion fully convolutional network using only prechange optical and postchange SAR images.
4) Data setting same as above, however using an encoder–decoder network-based method that projects prechange optical image into SAR image [59].
5) An early fusion fully convolutional network using only prechange SAR and postchange optical images.
6) Data setting same as above, however using an encoder–decoder network-based method that projects prechange SAR image into optical image [59].
7) Data setting same as the proposed method, i.e., prechange optical and SAR images and postchange SAR image are used. However, instead of the proposed Siamese setting, a fully convolutional network [37] is used where the all three input images are stacked and fed to the network. For fairness of comparison, the fully convolutional network uses same number of layers as the proposed method.
8) Data setting same as the proposed method, however using an encoder–decoder-based approach, similar to [59], where input of encoder is stacked prechange optical and SAR images.
9) Data setting same as the proposed method, however using a Vanilla Siamese network.
10) Proposed method without test time adaptation.

In addition, we also perform the following additional ablation studies:

1) By varying the value of p for test time adaptation.
2) By inserting a test-time dropout layer with high dropout rate (probability 0.9) after the first convolution layer in the encoder processing the prechange optical input or prechange SAR input or postchange SAR input. This study helps us to understand the relative importance of every of the three inputs, e.g., hypothesizing that the postchange SAR input is vital for the proposed CD architecture, then applying test-time dropout to its encoder will severely impact the CD performance.

D. Results

Table I shows the result from different data settings and methods.

Prechange: SAR, postchange: SAR: For this data setting, Siamese network outperforms the early fusion-based method, both in terms of F1 score (difference of 0.2) and accuracy (difference of 8.61%).

Prechange: Optical, postchange: SAR: While early fusion outperforms the encoder–decoder-based method, overall this data setting performs poorly compared to the SAR–SAR setting. This shows that merely setting up correspondence between prechange and postchange images from two different modalities is difficult.

Prechange: SAR, postchange: Optical: This data setting also performs poorly compared to the SAR–SAR setting. This leads to the same conclusion as above that merely setting up correspondence between prechange and postchange images from two different modalities is challenging.

Proposed, prechange: Optical+SAR, postchange: SAR: Benefiting from the availability of both prechange optical and prechange SAR images, the proposed data setting outperforms other data settings irrespective of architecture. As an example, early fusion using proposed setting obtains an F1 score of 27.72 in comparison to 22.88 obtained using prechange SAR and postchange SAR. Similarly, the encoder–decoder-based approach obtains a better F1 score than the same approach for other data settings.

Remarkably, early fusion obtains slightly better scores than Vanilla Siamese. The early fusion obtains an F1 score of 27.72 and accuracy of 89.70%. The proposed architecture outperforms both early fusion and Vanilla Siamese by a significant margin.

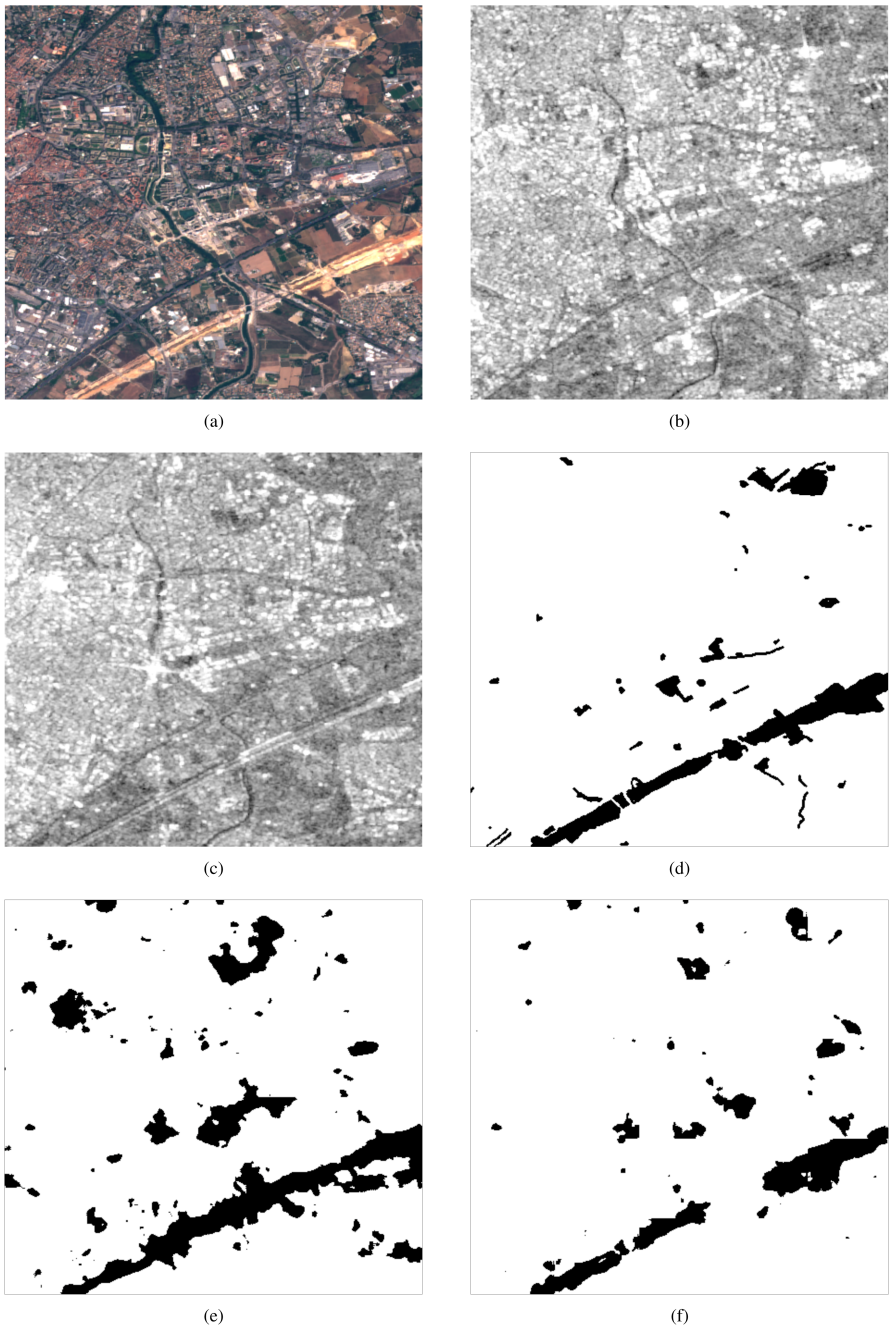

Fig. 2. Montpellier (one of the test cities). (a) Prechange (optical). (b) Prechange (SAR). (c) Postchange (SAR). (d) Reference image (white—unchanged, black—changed), and CD maps. (e) Early fusion. (f) Proposed. It is evident that early fusion produces much more false alarms compared to the proposed Siamese approach.

The proposed method (with $p = 5$) obtains an F1 score of 32.14 and accuracy of 91.28%.

Qualitative result for the early fusion and proposed architecture (both using proposed data setting) for the Montpellier city are shown in Fig. 2(e) and (f), respectively. It is evident that the proposed approach is less prone to false alarms than the early fusion approach.

Performance of the proposed method is better with proposed test time adaptation (F1 score: 32.14 in comparison to 30.25). This is because, by homogenizing the neural mean activation between test and training data, model makes the test features more similar to training features, for which it was originally trained.

Variation of p: In Table II, we show the variation of the performance of the proposed method as p is varied during test time adaptation. While the best F1 score is obtained at $p = 5$, we observe a relatively stable performance w.r.t. variation of p. This is an advantage of the proposed method, as we would not need to focus on setting its value for practical applications.

Weighted loss: Due to the strong class imbalance between changed and unchanged class, we have used weighted cross-entropy loss, similar to previous works on Siamese supervised

TABLE II
VARIATION OF THE PERFORMANCE OF THE PROPOSED METHOD AS TEST TIME ADAPTATION IS PERFORMED AT MORE LOCAL SCALE BY INCREASING VALUE OF p

Split (p)	Precision	Recall	Accuracy	F1*100
2	0.26	0.40	91.43	31.68
3	0.26	0.41	91.22	32.06
4	0.27	0.41	91.31	32.09
5	0.27	0.41	91.28	32.14
10	0.27	0.40	91.35	31.98

TABLE III
PERFORMANCE DROP OF THE PROPOSED METHOD AS A HIGH TEST-TIME DROPOUT IS APPLIED TO THE FIRST CONVOLUTION LAYER IN THE ENCODER OF THE DIFFERENT INPUTS

Dropout applied to	Accuracy drop	F1*100 drop
Prechange Optical	5.26	7.04
Prechange SAR	0.98	4.81
Both prechange	6.85	10.9
Postchange SAR	14.40	9.04

Notes: If an input is more important to the network, then the performance drop is expected to be high when test-time dropout attenuates that input.

CD [55]. To further validate this, we experimented with the nonweighted version of cross-entropy loss and found that it fails to obtain satisfactory result (F1 score 11.54 only).

Relative importance of three inputs: Table III shows the fall in performance (accuracy and F1 score) as test-time dropout is applied to the first convolution layer of encoder of some of the three inputs. We observe that most significant decrease in performance is observed when dropout is applied to postchange SAR input. This is intuitive as we have only one postchange input, and thus, the information from it is essential to make decision about change. Similar decrease in performance is also observed if dropout is applied simultaneously to both prechange inputs. Among the prechange optical and prechange SAR inputs, decrease in performance is more in the case of prechange optical input. While dropout applied to the prechange optical input leads to drop in F1 score of 7.04, the drop for the prechange SAR is 4.81. This shows that for the proposed data setting, the network extracts the prechange information more from the optical input than the SAR input.

V. CONCLUSION

This article presents a new data setting for remote sensing CD, using prechange optical and SAR images and postchange SAR image. Using a set of experiments, we show that this setting obtains superior performance to only SAR scenario. This data setting is especially useful, since at occurrence of any disaster/incident, it is indeed possible in most cases to obtain an artifact-free prechange optical image; however, it is not always practical to wait for acquisition of an artifact-free postchange optical image. The article also presents a novel Siamese architecture for this data setting, where prechange and postchange SAR images share the same encoder, while prechange optical data are processed using a separate encoder. The proposed architecture effectively fuses information from different data modalities and processes them using a common decoder to obtain the change map. Our analysis also shows the relative importance of three inputs for our data setting, ordered as postchange SAR, prechange optical, and prechange SAR. Our future work will focus on designing more robust architecture for the proposed data setting. We will furthermore investigate the combination with image reconstruction methods for spatio-temporal cloud removal [61]. We will also like to extend the method for fine grained changed objection detection [62], [63].

REFERENCES

[1] M. Zanetti et al., "A system for burned area detection on multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5404315.

[2] A. El-Zein, T. Ahmed, and F. Tonmoy, "Geophysical and social vulnerability to floods at municipal scale under climate change: The case of an inner-city suburb of Sydney," *Ecological Indicators*, vol. 121, 2021, Art. no. 106988.

[3] T. Waisberg, "The emergence of a security approach to climate change governance: Prospects and challenges to enhance environmental security in Brazil and the global south," *Revista Brasileira de Estudos Latino Americanos-REBELA*, vol. 11, no. 2, pp. 358–377, 2021.

[4] L. Höhn, M. Leunda, E. Gobet, W. Tinner, and C. Schwörer, "Vegetation response to rapid climate change during the Late Glacial–Early Holocene transition at Gola di Lago, Southern Switzerland," *Boreas*, vol. 51, no. 3, pp. 606–620, 2022.

[5] T. Sivasankar, S. Ghosh, and M. Joshi, "Exploitation of optical and SAR amplitude imagery for landslide identification: A case study from Sikkim, northeast India," *Environ. Monit. Assessment*, vol. 193, no. 7, pp. 1–17, 2021.

[6] B. Sadhukhan, S. Chakraborty, and S. Mukherjee, "Investigating the relationship between earthquake occurrences and climate change using RNN-based deep learning approach," *Arabian J. Geosci.*, vol. 15, no. 1, 2022, Art. no. 31.

[7] M. N. Hamoodi, "Investigating the effects of armed and political conflicts on the land use/cover change and surface urban heat islands: A case study of Baghdad, Iraq," *J. Indian Soc. Remote Sens.*, vol. 49, no. 7, pp. 1493–1506, 2021.

[8] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 871.

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

[10] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1917–1929, Mar. 2021.

[11] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3826–3852, Jul. 2013.

[12] V. Prudente et al., "SAR data for land use land cover classification in a tropical region with frequent cloud cover," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 4100–4103.

[13] J. D. Bermudez, P. N. Happ, R. Q. Feitosa, and D. A. Oliveira, "Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1220–1224, Aug. 2019.

[14] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[15] J. Ling, H. Zhang, and Y. Lin, "Improving urban land cover classification in cloud-prone areas with polarimetric SAR images," *Remote Sens.*, vol. 13, no. 22, 2021, Art. no. 4708.

[16] S. Saha, F. Bovolo, and L. Bruzzone, "Destroyed-buildings detection from VHR SAR images using deep features," *Proc. SPIE*, vol. 10789, 2018, Art. no. 107890Z.

[17] P. Ebel, S. Saha, and X. X. Zhu, "Fusing multi-modal data for supervised change detection," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 243–249, 2021.

[18] J. Gawlikowski, M. Schmitt, A. Kruspe, and X. X. Zhu, "On the fusion strategies of Sentinel-1 and Sentinel-2 data for local climate zone classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2081–2084.

[19] S. Hafner, A. Nascetti, H. Azizpour, and Y. Ban, "Sentinel-1 and Sentinel-2 data fusion for urban change detection using a dual stream U-NET," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4019805.

[20] S. Saha, P. Ebel, and X. X. Zhu, "Self-supervised multisensor change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 4405710.

[21] L. Wan, Y. Xiang, and H. You, "A post-classification comparison method for SAR and optical images change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1026–1030, Jul. 2019.

[22] L. Wan, Y. Xiang, and H. You, "An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9941–9959, Dec. 2019.

[23] D. Chicco, "Siamese neural networks: An overview," in *Artificial Neural Networks*. Berlin, Germany: Springer, 2021, pp. 73–94.

[24] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a Siamese time delay neural network," *Adv. Neural Inf. Process. Syst.*, vol. 6, pp. 737–744, 1993.

[25] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *Proc. IEEE 23rd Int. Conf. Pattern Recognit.*, 2016, pp. 378–383.

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

[27] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 8007805.

[28] Q. Zhu et al., "Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 63–78, 2022.

[29] W. Kang, Y. Xiang, F. Wang, and H. You, "CFNet: A cross fusion network for joint land cover classification using optical and SAR image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1562–1574, 2022.

[30] 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.

[31] X. Zhang, X. Su, Q. Yuan, and Q. Wang, "Spatial–temporal gray-level co-occurrence aware CNN for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4018605.

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

[33] R. Wang, L. Wang, X. Wei, J.-W. Chen, and L. Jiao, "Dynamic graph-level neural network for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4501005.

[34] Y. Gao, F. Gao, J. Dong, Q. Du, and H.-C. Li, "Synthetic aperture radar image change detection via Siamese adaptive fusion network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10748–10760, 2021.

[35] G. Zhao and Y. Peng, "Semisupervised SAR image change detection based on a Siamese variational autoencoder," *Inf. Process. Manage.*, vol. 59, no. 1, 2022, Art. no. 102726.

[36] Y. Lin, S. Li, L. Fang, and P. Ghamisi, "Multispectral change detection with bilinear convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1757–1761, Oct. 2020.

[37] R. C. Daudt, B. Le Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[38] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 24–41, 2016.

[39] W. Zhao, Z. Wang, M. Gong, and J. Liu, "Discriminative feature learning for unsupervised change detection in heterogeneous images based on a coupled neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7066–7080, Dec. 2017.

[40] S. Hafner, "Multi-modal deep learning with Sentinel-1 and Sentinel-2 data for urban mapping and change detection," Ph.D. dissertation, School Architecture Built Environ. (ABE), Urban Planning Environ., Geoinformatics KTH Royal Inst. Technol., Stockholm, Sweden, 2022.

[41] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm Remote Sens.*, vol. 107, pp. 50–63, 2015.

[42] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1822–1834, Apr. 2018.

[43] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised multiple-change detection in VHR multisensor images via deep-learning based adaptation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5033–5036.

[44] M. Wang, K. Tan, X. Jia, X. Wang, and Y. Chen, "A deep Siamese network with hybrid convolutional feature extraction module for change detection based on multi-sensor remote sensing images," *Remote Sens.*, vol. 12, no. 2, 2020, Art. no. 205.

[45] F. Rahman, B. Vasu, J. Van Cor, J. Kerekes, and A. Savakis, "Siamese network with multi-level features for patch-based change detection in satellite imagery," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2018, pp. 958–962.

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

[47] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.

[49] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2019.

[50] Z. Zhou, H. Huang, and B. Fang, "Application of weighted cross-entropy loss function in intrusion detection," *J. Comput. Commun.*, vol. 9, no. 11, pp. 1–21, 2021.

[51] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, vol. 32 pp. 1–12.

[52] S. Saha, S. Zhao, and X. X. Zhu, "Multitarget domain adaptation for remote sensing classification using graph neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6506505.

[53] S. Zhao, S. Saha, and X. X. Zhu, "Graph neural network based open-set domain adaptation," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 1407–1413, 2022.

[54] X. Dong, J. Guo, A. Li, W.-T. Ting, C. Liu, and H. Kung, "Neural mean discrepancy for efficient out-of-distribution detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 19217–19227.

[55] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2115–2118.

[56] F. Warmerdam, "The geospatial data abstraction library," in *Open Source Approaches Spatial Data Handling*. Berlin Germany: Springer, 2008, pp. 87–104.

[57] S. Saha, Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone, "Unsupervised deep transfer learning-based change detection for HR multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 856–860, May 2021.

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

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

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[61] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu, "SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5222414.
[62] X. Sun et al., "FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 116–130, 2022.
[63] J. Kim, J. Bae, G. Park, D. Zhang, and Y. M. Kim, "N-ImageNet: Towards robust, fine-grained object recognition with event cameras," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2146–2156.

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

From 2015 to 2016, he was an Engineer with TSMC Limited, Hsinchu, Taiwan. In 2019, he was a Guest Researcher with the Technical University of Munich, Munich, Germany, where he has been a Postdoctoral Researcher since 2020. His research interests include multitemporal remote sensing image analysis, domain adaptation, self-supervised learning, image segmentation, uncertainty quantification, deep learning, and image processing.

Dr. Saha was the recipient of Fondazione Bruno Kessler Best Student Award 2020. He is a Reviewer for several international journals. He was a Guest Editor for the special issue on Advanced Artificial Intelligence for Remote Sensing: Methodology and Application of *Remote Sensing* (MDPI).

Muhammad Shahzad (Senior Member, IEEE) received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2004, the M.Sc. degree in autonomous systems (robotics) from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany, in 2011, and the Ph.D. degree on radar remote sensing and image analysis from the Department of Signal Processing in Earth Observation, Technische Universität München (TUM), Munich, Germany, in 2016. His Ph.D. dissertation was titled "Automatic 3-D Reconstruction of Objects From Point Clouds Retrieved From Spaceborne Synthetic-Aperture-Radar (SAR) Image Stacks."

He possesses rich experience in multimodal remote sensing data processing. Besides, he also attended twice two-week professional thermography training course at Infrared Training Center (ITC), North Billerica, MA, and Portland, OR, USA, in 2005 and 2007, respectively. In 2015–2016, he was a Visiting Scientist with the Institute for Computer Graphics and Vision, Technical University of Graz, Austria. From 2016 to 2021 (currently on leave), he was an Assistant Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad. Furthermore, he also remained Co-Principal Investigator of the established Deep Learning Laboratory (DLL) under the umbrella of National Center of Artificial Intelligence (NCAI), Islamabad. He is currently a Core Project Scientist and Guest Professor with the AI4EO Future Lab, Data Science in Earth Observation, TUM. His research interests include application of deep learning for processing unstructured/structured 3-D point clouds, optical RGBD data, and very high-resolution radar images.

Patrick Ebel received the B.Sc. degree in cognitive science from the University of Osnabrück, Osnabrück, Germany, in 2015, and the M.Sc. degree in cognitive neuroscience from Radboud University Nijmegen, Nijmegen, The Netherlands, in 2018. He is currently working toward the Ph.D. degree with the Department of Aerospace and Geodesy, Technical University of Munich, Munich, Germany.

His research interests include deep learning and its applications in computer vision and to remote sensing data.

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

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

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