Urban-Aware U-Net for Large-Scale Urban Flood Mapping Using Multitemporal Sentinel-1 Intensity and Interferometric Coherence

Jie Zhao, Yu Li, Patrick Matgen, Ramona Pelich, *Member, IEEE*, Renaud Hostache, Wolfgang Wagner, *Senior Member, IEEE*, and Marco Chini, *Senior Member, IEEE*

***Abstract*— Due to the complexity of backscattering mechanisms in built-up areas, the synthetic aperture radar (SAR)-based mapping of floodwater in urban areas remains challenging. Open areas affected by flooding have low backscatter due to the specular reflection of calm water surfaces. Floodwater within built-up areas leads to double-bounce effects, the complexity of which depends on the configuration of floodwater concerning the facades of the surrounding buildings. Hence, it has been shown that the analysis of interferometric SAR coherence reduces the underdetection of floods in urbanized areas. Moreover, the high potential of deep convolutional neural networks for advancing SAR-based flood mapping is widely acknowledged. Therefore, we introduce an urban-aware U-Net model using dual-polarization Sentinel-1 multitemporal intensity and coherence data to map the extent of flooding in urban environments. It uses *a priori* information (i.e., an SAR-derived probabilistic urban mask) in the proposed urban-aware module, consisting of channel-wise attention and urban-aware normalization submodules to calibrate features and improve the final predictions. In this study, Sentinel-1 single-look complex data acquired over four study sites from three continents have been considered. The qualitative evaluation and quantitative analysis have been carried out using six urban flood cases. A comparison with previous methods reveals a significant enhancement in the accuracy of urban flood mapping: the F1 score of flooded urban increased from 0.3 to 0.6 with few false alarms in urban area using our method. Experimental results indicate that the proposed model trained with limited datasets has strong potential for near-real-time urban flood mapping.**

***Index Terms*— Deep learning (DL), multitemporal synthetic aperture radar (SAR), U-Net, urban flood mapping, urban-aware.**

Manuscript received 16 December 2021; revised 21 April 2022 and 14 July 2022; accepted 27 July 2022. Date of publication 16 August 2022; date of current version 31 August 2022. This work was supported by the Luxembourg National Research Fund (FNR) through the Hydrological Sciences (HYDRO-CSI) Project (reference: FNR PRIDE HYDRO-CSI) under Grant 10623093. *(Corresponding author: Jie Zhao.)*

Jie Zhao is with the Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology (LIST), 4362 Esch-sur-Alzette, Luxembourg, and also with the Remote Sensing Research Group, Department of Geodesy and Geoinformation, Vienna University of Technology, 1040 Vienna, Austria (e-mail: jie.zhao@geo.tuwien.ac.at).

Yu Li, Patrick Matgen, Ramona Pelich, and Marco Chini are with the Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology (LIST), 4362 Esch-sur-Alzette, Luxembourg (e-mail: yu.li@list.lu; patrick.matgen@list.lu; ramona.pelich@list.lu; marco.chini@list.lu).

Renaud Hostache is with the Institut de recherche pour le développement, 34394 Montpellier, Occitanie, France (e-mail: renaud.hostache@ird.fr).

Wolfgang Wagner is with the Remote Sensing Research Group, Department of Geodesy and Geoinformation, Vienna University of Technology, 1040 Vienna, Austria (e-mail: wolfgang.wagner@geo.tuwien.ac.at).

Digital Object Identifier 10.1109/TGRS.2022.3199036

I. Introduction

FLOOD inundation is one of the most costly and worldwide natural disasters, having a profound impact on the global population in terms of lives and property [1], [2]. It is estimated that between 2000 and 2018, about 2.23 million square kilometers of land were flooded and 255–290 million people were directly affected [2]. Meanwhile, the losses caused by floods are projected at US$597 billion from 2016 to 2035 by Willner *et al.* [3]. Spaceborne synthetic aperture radars (SARs) are the optimal sensors for detecting floodwater due to their capability to penetrate clouds and their quasi-independence from sun illumination and weather conditions. In recent decades, the application of SAR-based flood mapping has increased rapidly, thanks to the growing number of SAR satellite constellations characterized by different frequencies, polarizations, and resolutions (e.g., Sentinel-1, COnstellation of small Satellites for the Mediterranean basin Observation (COSMO) SkyMed, TerraSAR-X, RADARSAT-2, Advanced Land Observing Satellite (ALOS)-2, and GaoFen (GF) 3). The majority of flood mapping methods presented in the literature exploit the above-mentioned satellite data (e.g., [4]–[8]) and focus on inundated rural areas (i.e., bare soils and sparsely vegetated terrain) and flooded vegetated areas (i.e., flooded forest [9], wetlands [10], rice field [11]), while methods and applications targeting urban areas are less frequent.

Urban flood mapping methods used in the past can be divided into two categories: traditional remote sensing methods (e.g., [5], [12], [13]) and machine learning (ML)/deep learning (DL)-based methods (e.g., [14]–[16]). Compared with the traditional flood mapping methods, the ML/DL methods are able to learn the characteristics of different classes (e.g., nonflood class, flooded open areas, flooded urban (FU) areas, and flooded vegetation areas) from the training dataset directly instead of relying on assumptions based on prior knowledge. As the amount of training data tends to be very small, it remains particularly challenging to adopt DL approaches for SAR-based flood mapping in urban areas. The application of the attention mechanism and normalization technique alleviated the problem without entirely solving it. Thus, we hypothesize that the introduction of prior information derived from multitemporal SAR data to the DL model improves the robustness and generalization of the trained model. To test this hypothesis, we applied an urban/nonurban mask as in [12] to provide additional essential information for the selection of

effective features characterizing the urban class. Here, instead of using the urban mask as an additional layer of the input data, we introduced and evaluated an urban-aware module making use of an urban mask to refine informative features. In addition, we hypothesized that the proposed DL-based model is able to generalize and be robust in large-scale flood mapping with different land cover scenarios, focusing not only on urban areas but also on bare soils and sparsely vegetated areas. The possibility of generalization and the robustness of the proposed model are also evaluated in this study.

This article is organized as follows: Section II introduces several state-of-the-art methods. Section III describes the proposed urban-aware module and urban-aware U-Net in detail. Section IV describes the flood events, datasets, and the experimental setup. Section V evaluates the results via a series of quantitative and qualitative comparisons. Section VI discusses the advantages and disadvantages of our proposed method. Finally, conclusions are drawn in Section VII.

II. Related Work

Due to the complexity of the urban fabric, SAR-based flood mapping within built environments is challenging, which may explain the comparatively low number of studies focusing on urban floods. Urban flood mapping methods can be divided into several categories depending, for example, on the input data that are considered: methods combining SAR data with auxiliary data [e.g., digital surface model (DSM), hydraulic data, world settlement footprint (WSF2015)] (e.g., [17]–[19]), SAR intensity-only methods (e.g., [20]), InSAR coherence-only methods (e.g., [12], [21]), interferometric phase-only method [13], and methods combining InSAR coherence and interferometric phase statistics [22]. Mason *et al.* [17], [18] delineated urban floods with the help of water height maps and high-resolution DSM [i.e., Light Detection and Ranging (LiDAR) DSM and World digital elevation model (DEM)]. Tanguy *et al.* [19] developed a near-real-time flood extent mapping method relying on the combined use of very-high-resolution Radarsat-2 data acquired in the horizontal-horizontal (HH) polarization and hydraulic data (i.e., flood return period). When mapping urban floods using SAR intensity data, [20] proposed a multitemporal SAR intensity [vertical-vertical (VV)]-based method assuming that the presence of urban flooding increases the SAR backscatter. This assumption is not correct when the double-bounce scattering between the floodwater and the façades of the surrounding buildings is either insignificant or the buildings are submerged by floodwater. To reduce the number of underdetected pixels, the InSAR coherence was investigated for flood mapping in urban areas. Pulvirenti *et al.* [21] explored the role of interferometric coherence in complementing intensity SAR data for mapping floods in urban environments. The study showed for the first time that the use of coherence data significantly reduces underdetections. Floodwater in built-up environments is mapped by detecting a decrease in the multitemporal InSAR coherence (VV) in the vicinity of buildings [12]. The study reported in [13] analyzed the interferometric phase information instead of the more frequently used coherence to detect floodwater in built-up areas. By investigating the L-band Phased Array L-band Synthetic Aperture Radar (PALSAR)-2 data in the HH polarization, the study showed that it is beneficial to explore phase statistics rather than coherence in the case of point-like scatterers. In fact, the underdetection of point-like scatterers affected by flooding may occur when the InSAR coherence is dominated by strong scatterers. When the latter remains stable during a flood event, the phase variation caused by the presence of floodwater might be negligible and therefore undetectable. Recently, Pulvirenti *et al.* [22] proposed detecting floodwater in settlements by exploring the InSAR multitemporal coherence (VV) over persistent scatterers (PSs). In their study, the phase statistics and the InSAR coherence of flooded PSs were used together and this led to a decrease in omission errors in urban flood mapping. The joint exploration of both the VV and vertical-horizontal (VH) polarizations has also proven to be informative for urban flood mapping. Pelich *et al.* [23] demonstrated that adding the VH polarization leads to an increase in the accuracy of the urban flood maps from 75.2% when only using the VV polarization, to 82.9% when using dual-polarization information. This could be explained by the fact that multiple-bounce scattering occurs in addition to double-bounce effects in urban areas and the importance of this mechanism depends on the orientation of the building facades with respect to the SAR sensor's line of sight (LoS) [23].

Besides the conventional flood mapping methods mentioned above, applications of ML/DL methods have attracted much attention in the remote sensing field in recent years [24]–[26]. The advantage of these ML/DL approaches is that they only rely on the annotated input data during the training phase, while during the inference stage, no assumptions have to be made regarding land cover classes. Indeed, precise information on land cover classes is not always readily available, especially for complex environments such as urban areas containing different subclasses, e.g., dense buildings, bare soil, and vegetated areas. However, only a few studies using the ML and DL methods for urban flood mapping are available in the literature. Moya *et al.* [16] used a traditional ML approach, i.e., support vector machine (SVM) and historic Japanese flood events as a training dataset to identify urban floods in Koriyama, Iwaki, and Nagano during the 2019 Hagibis Typhoon. The gathering of training data for urban flood mapping is time-consuming and difficult due to the scarcity of available labels. A temporal-ensembling active self-learning convolutional neural network (A-SL CNN) was adopted for a major flood event that occurred in Houston in 2017 using SAR intensity and InSAR coherence to alleviate the negative effects of a limited annotated training dataset [14]. It should be noted that due to the limitation of annotated data, the A-SL CNN is trained using labeled image patches (i.e., 32×32 pixels) rather than labeled pixels. Li *et al.* [15] proposed an unsupervised urban flood mapping method, combining SAR intensity and InSAR coherence in the framework of a Bayesian network, which is a probabilistic graphical model for parameterizing joint probability distribution over a fixed set of random variables. However, due to the limitation of urban flood training and testing datasets, the above-mentioned studies were trained and tested using the data acquired during one or two urban flood events. The possibility of reapplying the trained models to different flood events was not evaluated in any of these

studies. Thus, to propose a DL-based flood mapping method that can be generalized and reapplied to different scenarios, the attention mechanism should be considered as it plays an essential role in identifying the most representative features, especially when training labels are scarce and unbalanced within different classes. The attention mechanism inspired by the human visual system has been widely used in computer vision and natural language processing [27], as it can focus on important features and suppress unnecessary ones. The attention mechanism can be divided into two categories: self-attention and cross-attention. The self-attention computes attention weights based only on the input features. There are several widely used self-attention modules in the computer vision domain. For example, Hu *et al.* [28] used channel-wise attention to highlight important feature dimensions (i.e., layer of feature). Spatial information can be considered and combined with channel attention modules either in a sequence or in parallel. As an example, Woo *et al.* [29] proposed a convolutional block attention module (CBAM) composed of two submodules, i.e., a channel attention followed by a spatial attention, while the dual-attention module introduced in [30] fused the refined features from the spatial and channel attention modules in parallel. Besides the self-attention mechanism, the cross-attention mechanism was introduced to bridge attention across different data. For example, multimodal approaches take advantage of the attention from one modality (e.g., LiDAR) to highlight features in the other modality (e.g., hyperspectral data) [31]. The attention gate proposed by [32], [33] also acts as cross-attention, since the gating signal is used to help the model concentrating on the local spatial regions of input features instead of the global ones. Besides the attention mechanism, the normalization is able to select the robust features by rescaling them in the DL models. Common normalizations in the DL models are: batch normalization (BN) [34], instance normalization [35], layer normalization [36], and group normalization [37]. Moreover, there are some normalizations that consider additional conditions, i.e., spatially adaptive normalization [38], adaptive instance normalization (AdaIN) [39], and attentive normalization [40]. Thus, features could be affine transformed using different information, such as the other features or conditions. However, to the best of our knowledge, the attention mechanism and feature normalization have not yet been applied to SAR-based flood mapping. As shown by the above-mentioned studies, it is advantageous to consider SAR intensity, InSAR coherence, and the cross- and copolarization data acquired before and during the flood events to map floodwater over bare soils and urban areas. Different information is necessary for each specific land cover (e.g., bare soils and urban areas). Thus, DL approaches must select the most adequate modality for each class.

III. Methods

In urban areas, there are three distinct scattering mechanisms: specular reflection, the double, and multiple bounces. In the case of specular reflection, the presence of water leads to a decrease in SAR intensity backscattering in both the VV and VH channels. In the case of double-bounce, the intensity is expected to increase mainly in the VV channel, and the magnitude of the effect depends on the orientation of the building façade with respect to the LoS of the sensor. The strength of double-bounce also depends on the height of the building above the water level: the double-bounce effect reduces the difference between the water surface and the height of the surrounding buildings with decreasing elevation and completely disappears when the building is totally submerged. Thus, the interferometric coherence, which indicates the correlation of two complex (amplitude and phase) observations, plays an important role in indicating the surface changes caused by floodwater. In principle, urban settlements are considered as steady targets characterized by high coherence even at large temporal baselines without considering the vegetation with unstable structures resulting in low coherence. Hence, a sudden decrease in the multitemporal InSAR coherence in both the VV and VH polarizations is indicative of the appearance of floodwater between two image acquisitions. However, a significant drop in coherence may also occur because of a precipitation cell covering bare soils, thereby potentially causing false alarms in the urban flood maps. Thus, it is evident that the input features must be carefully selected to tackle floodwater mapping over different land cover classes. Given the scarcity of labels for the urban floodwater class, there is indeed a high risk that self-attention approaches generate less accurate results in terms of the robustness of the trained model, especially when applied to new events in different locations. Thus, we hypothesize that considering an urban mask provides additional information and potentially contributes to more accurate results. An intuitive way to make use of the urban mask is to simply concatenate it with SAR input and then feed them into a semantic segmentation network. However, such a method may suffer from the distribution bias between the training dataset and the application dataset due to the limited volume/diversity of the training dataset. This may lead to inefficient use of the SAR-derived urban mask and thus less accurate predictions. To address this issue, we propose the application of an urban-aware module as a way to leverage an additional urban layer to normalize the features learned and the final predictions. We expect the urban mask to provide information about the built-up area's geometry and the probability of building locations. Such features can guide the neural network toward normalizing features of urban and nonurban flood classes and can thus improve the accuracy of urban flood mapping. This normalization is performed for each input instance adaptively in both the training and testing phases. Then, the urban-aware module is embedded into a multiscale network architecture, i.e., U-Net.

It is worth mentioning that the double- and multiple-bounce scattering mechanisms are also caused by tall vegetation. When such areas are flooded, an increase in the backscattering value is expected, while the specular reflection mechanism dominates in cases of short or sparse vegetation. The latter results in a decrease in the backscattering value [10]. In this study, we consider three classes (i.e., open flood, urban flood, and no flood) because when considering 20 m resolution Sentinel-1 images, it was not easy to clearly identify a class of flooded vegetation in our use cases. Vegetated areas are

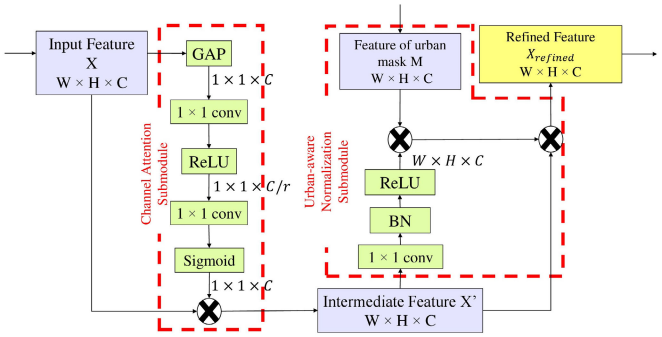


Fig. 1. Structure of the urban-aware module.

mainly composed of isolated trees. Thus, we assume that trees belonging to the same resolution cell as buildings dominate the backscattering. Such areas were assigned to the urban flood class, while low vegetation was associated with open water flooding. We note that isolated trees generating a raise of backscattering values in the presence of floodwater could cause an underdetection of floodwater via the proposed algorithm since vegetation usually shows a low value of multitemporal coherence.

A. Urban-Aware Module

The urban-aware module is applied to facilitate the learning process of the model by focusing on the more representative and robust features for different targets (i.e., different flood classes). Inspired by the work in [29], our module has two submodules that are applied sequentially: i.e., the channel attention and urban-aware normalization submodules (Fig. 1). The feature map is expected to be adaptively refined through each submodule.

1) Channel Attention Submodule: Given a multilayer feature tensor $X \in R^{W\times H\times C}$ where W, H, and C present, respectively, the number of width, height, and channel of the feature map, different attention weights are applied to focus on different channels. Each channel brings different information content. Each channel attention exploits the inter-channel relationship of features using a learnable network that weighs the importance of each channel and generates more informative outputs. In previous studies (e.g., [28], [29], [41]–[43]), the global average pooling (GAP) was selected to squeeze and aggregate the spatial dimension of the input feature map. In the channel attention module, GAP is adopted, followed by the application of two 1×1 convolutions with a rectified linear unit (ReLU) comprised in between. Finally, the sigmoid function is used to obtain a channel attention score (X_c) with values ranging from 0 to 1. The graphical visualization of the channel attention submodule is shown in Fig. 1, and the following equation gives a mathematical expression for the aforementioned channel score (S_c):

$$S_c = \text{sigmoid}(\text{Conv}_{2d}(\text{ReLU}(\text{Conv}_{2d}(\text{GAP}(X))))). \quad (1)$$

Next, the channel attention refined feature X' can be obtained, as follows:

$$X' = S_c \times X. \quad (2)$$

2) Urban-Aware Normalization Submodule: In this section, a normalization was considered by introducing the probabilistic urban mask. The normalization submodule is generally composed of a scale factor and a bias component. The scale factor is derived from the intermediate feature and the "urban mask." We compared the results obtained when normalizing both with and without a bias derived from the "urban mask" and found that in our study, the approach that did not use any bias term achieved slightly better results. Therefore, the normalization submodule without bias is applied in this study.

Given an intermediate feature tensor $X' \in R^{W\times H\times C}$, a block containing 1×1 convolutions, a BN, and a ReLU activation unit are applied. Then, the scale factor W_s is obtained by multiplying it with the urban probabilistic mask feature M

$$W_s = \text{ReLU}\big(\text{BN}\big(\text{conv}_{2d}\big(X'\big)\big)\big) \times M. \quad (3)$$

Finally, the refined feature (X_{refined}), where the urban-aware normalization is applied, can be obtained

$$X_{\text{refined}} = W_s \times X'. \quad (4)$$

As for each input instance, there is an associated "urban mask" that differs from the standard normalization fixed in the testing phase, and this conditional normalization is applied to each instance adaptively in both the training and testing phases. This is done to improve the generalization capability of the model.

3) SAR-Derived Probabilistic Urban Mask: As shown in Fig. 1, an SAR-derived probabilistic urban mask is used as input in the proposed urban-aware normalization submodule. Following [44], we introduced a fuzzy-logic-based probabilistic urban mask. It is generated based on the assumption that both the double-bounce and multiple-bounce effects exist in urban areas and can be detected and measured by the co-polarization and cross-polarization channels, respectively. However, high backscatter values may also be caused by different types of vegetation. To remove vegetation from the analysis and thus reduce the risk of overdetection, the InSAR coherence is considered in addition to SAR intensity. The reason for this is that buildings have stable structures leading to high temporal coherence, while vegetation changes over time lead to low coherence. These assumptions imply that the probabilistic urban mask can be generated using multitemporal mean backscatter in VV and VH and multitemporal mean coherence in VV and VH. In this study, five intensity images and five coherence images, which were acquired during a period without flooding or other changes, are used. We assume that the urban area is stable during this time period and that possible small changes in one image are attenuated by the temporal averaging step of the five intensity and coherence images. To integrate these four layers, a standard S membership function was applied [Fig. 2(a)], where parameters X_1 and X_2 define the minimum and maximum

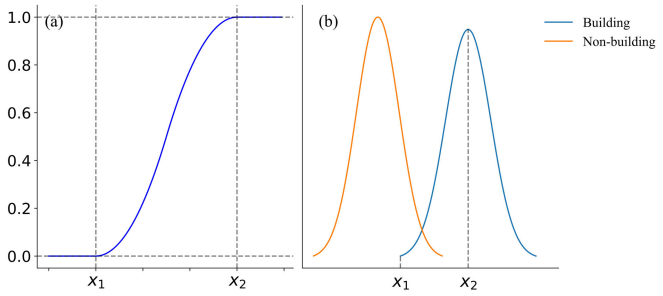


Fig. 2. Illustration of thresholds' selection for the fuzzy-logic-based probabilistic urban mask. (a) Standard S fuzzy membership function. For the S function, values smaller than X_1 have a membership degree equal to 0, while values higher than X_2 have a membership degree equal to 1. (b) Parametric Gaussian distributions of building and nonbuilding in intensity/coherence image.

fuzzy thresholds determining whether a pixel is considered a building or not, respectively. To automatically find suitable X_1 and X_2 for each fuzzy set, a hierarchical split-based approach (HSBA) [7] was applied. Generally, HSBA assumes that the distributions of backscatter values associated with buildings and all other backscatter values, respectively, can be parameterized as two Gaussian distributions in intensity VV and VH, and coherence VV and VH images, respectively. Hence, with the parameterized Gaussian distributions, X_2 is assigned as the mean value of the probability density function (PDF) attributed to backscatter values of buildings, while X_1 is assigned as three standard deviations of the mean (i.e., mean − 3std) [Fig. 2(b)]. The assumption of Gaussian distributions is motivated by the fact that the PDF of a random variable affected by speckle approaches a Gaussian distribution when equivalent number of looks (ENL) increases [45] and that the PDF of the log-transformed noise approaches the Gaussian distribution much faster than on a linear scale [46].

B. Urban-Aware U-Net

In this study, we used a U-Net specifically designed for image segmentation using limited training datasets as a basis. The technique has been widely used in remote sensing image segmentation (e.g., [48], [49]) since the low and high levels of details can be combined to achieve better performances by the skip connections. As shown in Fig. 3, the urban-aware module was adopted at each skip connection of the U-Net. The SAR-derived probabilistic urban mask should be downscaled to each level of the U-Net to be combined with the corresponding feature of the input SAR data. In this study, the downscaling of the SAR-derived probabilistic urban mask was conducted using the same U-Net encoder. The input and SAR-derived probabilistic urban mask features are downsampled at each level and refined by the urban-aware module. Thus, the fine-grained features are successively transmitted to the decoder through skip connections, compensating for the loss of position information in the deep layers of the decoder.

In this study, the focal loss (FL) function was used since it was found to be the most effective solution regarding data imbalance [50] as it considered the imbalance of the data and the difficulty level of the dataset

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{5}$$

where p_t is the class probability. The focusing parameter γ is set to 2 and the weighting factor α_t is set to 0.25 as reported in [50].

IV. Datasets and Experiments

In this section, the datasets and experiments are described in detail. First, several recent urban flood events in four study sites are introduced, and all the relevant data, i.e., SAR data and ground-truth data, are presented. Then, the preprocessing of SAR data and data preparation for the CNN model are explained step by step. Finally, the training of the model using the proposed urban-aware U-Net is presented in detail.

A. Urban Flood Test Cases and Datasets

In this study, several flood events impacting built-up areas in four different regions were selected:

1) The Hurricane Harvey-related flooding which affected the metropolitan area of Houston (USA) on August 30, 2017;
2) The Typhoon Hagibis-related flooding that hit the cities of Iwaki (Japan) and Koriyama (Japan) on October 12, 2019;
3) A rainfall-related flood that impacted the Webi Shebelle River region and the city of Beledweyne (Somalia) in three different years May 6, 2018, October 30, 2019, and May 20, 2020, respectively;
4) Cyclone Ida, which caused massive flooding in Mozambique and in particular in the city of Beira in March 2019.

These urban floods occurred in dense built-up areas composed of residential, commercial, and industrial districts, including roads, parks, and parking lots. Sentinel-1 single-look complex (SLC) data with the VV and VH polarizations are used for the four case studies. Since the proposed algorithm aims to detect floods over bare soils and urbanized areas, both the intensity and multitemporal InSAR coherence information are used. All the information about the dataset used are summarized in Table I.

1) Reference Flood Extent Extraction: Generally, the collection of ground-truth data for this kind of analysis remains a very challenging task. The little information that is available may come from different data sources and only cover a fraction of the main area of interest. Moreover, there might be significant issues of representativeness if SAR and reference data were not collected at the same time. However, this kind of information is crucial to assess the quality of the flood extent maps extracted using automatic algorithms and to provide reliable labels in the training phase of the supervised classification algorithms. The reference maps are usually manually derived from a visual inspection of high-resolution optical images acquired during flood events. However, since the necessary high-resolution data may not always be available, other information sources also have to be considered. If no independent data are available, the flood extent generated using a different SAR-based flood mapping method might be used as a last resort to evaluate the results. Since several different flood events have been considered in our study, the reference

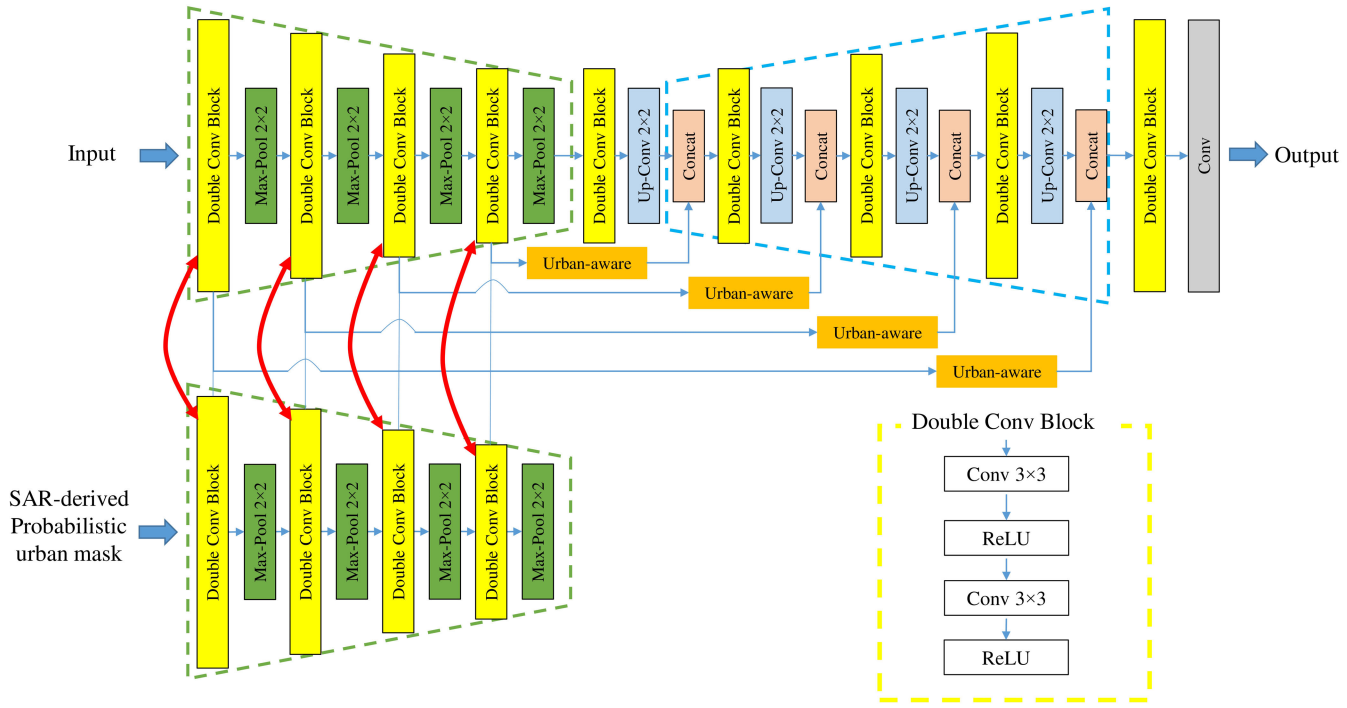


Fig. 3. Proposed urban-aware U-Net architecture: the block in orange represents the urban-aware module.

TABLE I
DETAILED INFORMATION OF URBAN FLOOD EVENTS AND SAR (I.E., SENTINEL-1) IMAGERIES CHARACTERISTICS. THE FLOODED ACQUISITIONS ARE MARKED IN RED

Event	Location	Acquisition data (DD/MM/YYYY/)	Orbit	Ground truth data		Sentinel-1 acquisition date for probabilistic urban mask generation
				Data source	Areas (number of pixels)	
Hurricane Harvey	Houston, US	18/08/2017 24/08/2017 30/08/2017	143	35 cm optical image from NOAA [51]	2,813,921	25/07/2017 31/07/2017 06/08/2017 12/08/2017
Typhoon Hagibis	Iwaki/Koriyama, Japan	24/09/2019 06/10/2019 12/10/2019	46	Iwaki Flood maps from EOS [52]; Koriyama Flood maps from GSI [53]	3,334,841	07/08/2019 19/08/2019 31/08/2019 12/09/2019
-	Beledweyne, Somalia	14/04/2018 26/04/2018 08/05/2018	35	Flood maps from UNOSAT [54]	47,185	25/02/2018 09/03/2018 21/03/2018 02/04/2018
		24/09/2019 06/10/2019 30/10/2019	35		47,185	-
		26/04/2020 08/05/2020 20/05/2020	116		47,185	09/03/2020 21/03/2020 02/04/2020 14/04/2020
Cyclone Ida	Beira, Mozambique	02/03/2019 14/03/2019 20/03/2019	6	Flood maps generated from SAR data [7], [12]	86,343,680	13/01/2019 25/01/2019 06/02/2019 18/02/2019

floodwater extent maps are obtained using different approaches and data sources:

- High-resolution optical data were available for the Houston event. Thus, the binary flood extents were manually derived by analyzing the 35 cm resolution optical data provided by National Oceanic and Atmospheric Administration (NOAA) [51], distinguishing floods over sparsely vegetated and urbanized areas.

- The reference maps for the Iwaki flood and Koriyama flood were provided by the Earth Observation of Singapore (EOS), which used Sentinel-1 data [52], and by the Geospatial Information Authority of Japan (GSI) [53], which used

TABLE II
CHARACTERISTICS OF THE INPUT DATASETS: TEMPORAL BASELINE (B_t) AND PERPENDICULAR BASELINE ($B_\perp$)

Flood cases	Image pair	Acquisition data (DD/MM/YYYY)	Characteristics	
			B_t (days)	$B_\perp$ (m)
Houston (US) flood in 2017	Pre-flood	18 & 24/08/2017	6	80.10
	Co-flood	24 & 30/08/2017	6	52.49
Iwaki & Koriyama (Japan) flood in 2019	Pre-flood	24/09/2019 & 06/10/2019	12	52.46
	Co-flood	06 & 12/10/2019	6	14.2
Beledweyne (Somalia) flood in 2018	Pre-flood	14 & 26/04/2018	12	32.71
	Co-flood	26/04/2018 & 08/05/2018	12	1.32
Beledweyne (Somalia) flood in 2019	Pre-flood	24/09/2019 & 06/10/2019	12	11.05
	Co-flood	06 & 30/10/2019	24	85.38
Beledweyne (Somalia) flood in 2020	Pre-flood	26/04/2020 & 08/05/2020	12	43.21
	Co-flood	08 & 20/05/2020	12	66.05
Beira (Mozambique) flood in 2019	Pre-flood	02 & 14/03/2019	12	39.37
	Co-flood	14 & 20/03/2019	6	89.13

airborne optical images, respectively. These binary flood maps were combined with an additional urban mask—WSF2015 [55]—to distinguish flooded bare soils/sparsely vegetated areas and FU areas.

– For the Beira test case, no independent flood extent maps were available. Hence, flood extent references indicating flooded bare soils/sparsely vegetated areas and FU areas were generated using SAR data. In particular, the flood extent maps were generated over bare soils using a state-of-the-art SAR intensity-based algorithm [7], while the floods in urbanized regions were delineated using a semiautomatic algorithm based on multitemporal InSAR coherence and land cover maps [12]. The resulting reference maps were refined by visual comparison with a land cover map and a very high-resolution optical image at 0.5-m spatial resolution acquired by Planet's Skysat satellite, presented in Fig. 10(c). Due to the limitations caused by the differences in their acquisition times (i.e., the optical image was acquired on March 23, 2019, three days after the acquisition of S1 data) and cloud cover, it was not possible to create a representative ground-truth map based on optical data. On March 23, the flood had already partially receded, and clouds were still affecting the area. Hence, the presence of water could only be observed in a few regions. Therefore, the optical image was used to carry out a visual inspection over areas not affected by cloud cover and for which the results provided by the InSAR coherence-based algorithm indicated the presence of floodwater. From Fig. 10(c), we note the presence of water around buildings indicated in dark green color in the optical image.

- Concerning the Somalia flood events that occurred in 2018, 2019, and 2020 [54], United Nations Satellite Centre (UNOSAT) provides the flood maps at district scale after visually inspecting optical images (i.e., WorldView-1/2, GeoEye-1) made available during the International Disaster Charter activations. These maps do not provide information at the pixel scale, but rather a degree of the flood (partially or fully) at the district scale.

2) Data Preprocessing and Preparation: This study was carried out using Level-1 SLC Sentinel-1 data acquired during several flood events, as shown in Table I. For each flood event, six SLC images were acquired before flooding started and one SLC flood image was acquired during flooding. SLC data were used to extract interferograms and were then calibrated and

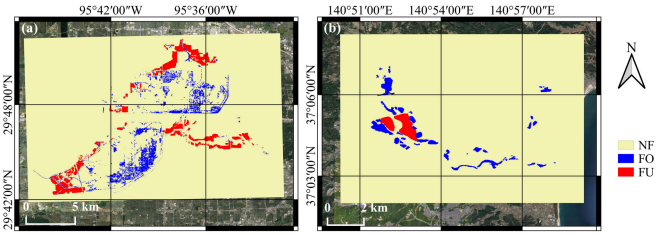


Fig. 4. Location of the training dataset. (a) ROI covering parts of Houston, USA. (b) ROI covering Iwaki, Japan.

transformed into intensity (in dB). Using SARscape software, the interferometric processing was performed to obtain a slant range interferogram with a square pixel of 20 m by applying a 4 × 1 multilook step (4 and 1 looks in the range and azimuth, respectively). A Goldstein filter with a size of 9 × 9 pixels was applied to the interferogram to reduce noise in the phase, and then the interferometric coherence was estimated by a moving window of 9 × 9 pixels. Finally, all the intensity and coherence data were geocoded to World Geodetic System (WGS) 1984 Universal Transverse Mercator (UTM) with 20-m spatial resolution. To attenuate the speckle in the intensity image, a Lee filter with a 9 × 9 window was also applied. It should be noted that the intensity and coherence were generated from the SLC image pairs for each polarization channel separately. Next, an SAR-derived probabilistic urban mask was generated for each location using six multitemporal intensity images (VV and VH) and five multitemporal coherence images (VV and VH). Besides the probabilistic urban mask, the input data of each flood event in our study had eight images: preevent and coevent coherence in VV and VH, and preevent and postevent intensity in VV and VH. Table II summarizes the characteristics of the input data of each flood case.

Given that one of the objectives of this study is to assess the possibility of applying the proposed approach to different flood events, only part of the available reference data is used to train the DL model, while the remaining part is used for testing the model. In particular, only one region of interest (ROI) with 1304 × 957 pixels in Houston (USA) and one ROI with 667 × 577 pixels in Iwaki (Japan) are selected as the training datasets (Fig. 4), while other ROIs from these two test cases and all other study sites are used only for testing.

As shown in Fig. 4, all the pixels are labeled as flooded open areas (i.e., bare soils/sparsely vegetated areas) (FO), FU areas, and nonflooded (NF) areas, respectively.

B. Model Training

First, the ROIs of the training dataset were split into patches with a size of 134 × 134 pixels, and each patch has 50% overlapping with its adjacent patches. Then, data augmentation, i.e., a random cropping with a size of 128 × 128 pixels, a random vertical and horizontal flipping, and a random rotation (i.e., 90°, 180°, 270°) were applied during the training phase. From this dataset, 70% of the samples were used for training and 30% of the samples were used for validation. To generate the dataset for testing the approach, the preprocessed images were split into nonoverlapping patches with a size of 128 × 128 pixels. The CNN architecture (section III-B) is trained with a batch size of eight and using an Adam Optimizer with an initial learning rate of 1e-6, momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and a weight decay with a coefficient of 1e-4. The FL was adopted to deal with the imbalanced datasets [50]. The models were trained for 100 epochs, thereby delaying the learning rate with ReduceLROnPlateau. Instead of using early stopping, the model weights are saved every five epochs and then we chose the best model, i.e., the model with the highest accuracy in terms of F1 score during the training phase for inferencing over several different test sites. All the models were implemented in PyTorch[1].

V. Results

The results of all the case studies are qualitatively and quantitatively evaluated. For the qualitative interpretation, red green blue (RGB) combinations of multitemporal intensity and coherence were generated. The Kappa coefficient, precision, recall, and F1 score are reported for quantitative evaluation. In Section V-A, the results of the evaluation for each region are reported.

A. Houston (USA) Flood Case

Fig. 5(a) shows the intensity (VV) composite (R = preevent, G = B = coevent) covering a subset of the Houston flood case in an area outside that used for training the model. The red color indicates FO, i.e., areas without any substantial double-bounce effect between the surface water and adjacent building facades or tree trunks. The cyan color depicts the FU and submerged vegetation where backscatter increases due to the double-bounce effect. The coherence (VV) RGB composite (R = preevent, G = B = coevent) is presented in Fig. 5(b). The white color indicates NF urban areas which are characterized by high coherence (VV) in both preevent and coevent acquisitions. The appearance of floodwater between buildings results in a significant drop in the coevent coherence (VV), as illustrated by the red color. It should be noted that the coherence (VV) in vegetated areas is low in both pre- and coevent acquisitions. In Fig. 5(c), the RGB composite of intensity (VV) and coherence (VV) is generated as R = postevent intensity, G = preevent coherence VV, and B = coevent coherence VV. The FU is recognizable in yellow (i.e., high postevent intensity, high preevent coherence, and low coevent coherence). The green color can be related to FO which is characterized by low postevent intensity, medium preevent coherence, and low coevent coherence. Fig. 5(d) and (f) depicting the data acquired in the VH polarization can be interpreted in a similar way. By comparing the FU in the yellow rectangular areas of Fig. 5(c) and Fig. 5(f) [zoom-in in Fig. 6(c) and Fig. 6(d)], it is clear that FU is recorded differently in the VV and VH polarizations, especially when considering the intensity. The reason for this is that the double-bounce (VV) or multiple bounces (VH) can vary depending on the orientation of a building with respect to the LoS of the satellite [23].

Table III provides an overview of the quantitative evaluation of the flood extents obtained for the Houston (USA) test case, for which high-resolution optical images are available. The Kappa coefficient is 0.83, and the F1 scores for FO and FU are 0.87 and 0.63, respectively. In Fig. 5(g) and (h), our FO (shown in blue) is consistent with the reference, while the FU of the flood map (represented in red) is underestimated with a recall of 0.54. Regions for which FU is underestimated (squares in yellow) correspond to dense built-up areas with heavy vegetation. To have a better understanding of the backscatter behavior leading to underdetected FU areas, a zoom-in of the regions highlighted by yellow squares is provided in Fig. 6(a)–(d). It can be seen that the underdetected areas are visible in cyan in the RGB combination of intensity and coherence [Fig. 6(c) and Fig. 6(d)]. This is indicative of a medium postevent intensity VV and VH, a medium preevent coherence VV and VH, and a medium/low coevent coherence VV and VH. Hence, the underdetection of FU can be explained by a comparatively small change in the InSAR coherence, arguably due to the medium value of the preevent image caused by the presence of vegetation (e.g., 0.61 in underdetected FU areas and 0.75 in correctly identified FU). Moreover, the spatial resolution difference between SAR data (i.e., 20 m) and optical-derived reference map (35 cm) may also lead to the underdetection of FU.

B. Koriyama (Japan) Flood Case

Fig. 7 shows the SAR data, flood reference map, and our flood map of the Koriyama (Japan) flood case in the exact same way as depicted in Fig. 5 and interpreted in Section V-A. From this figure, it can be observed that the reference [Fig. 7(g)] is consistent with our flood map [Fig. 7(h)] in FO, while the underestimation of FU is also depicted.

Table III also reports the quantitative evaluation of the flood extent in the Koriyama flood case. The Kappa coefficient (i.e., 0.51), F1 score of FO (i.e., 0.5), and F1 score of FU (i.e., 0.51) demonstrate moderate consistency between the reference and our result. The low precision (i.e., 0.5) and low recall (i.e., 0.5) of FO are indicative of the existence of both over- and underestimation. Fig. 8(a) depicts the overestimation of FO in green and the underestimation of FO in yellow. However, as shown in Fig. 8(c), the backscatter of those overdetected FO [i.e., green in 8(a)] is quite low. Thus,

[1]http://pytorch.org/

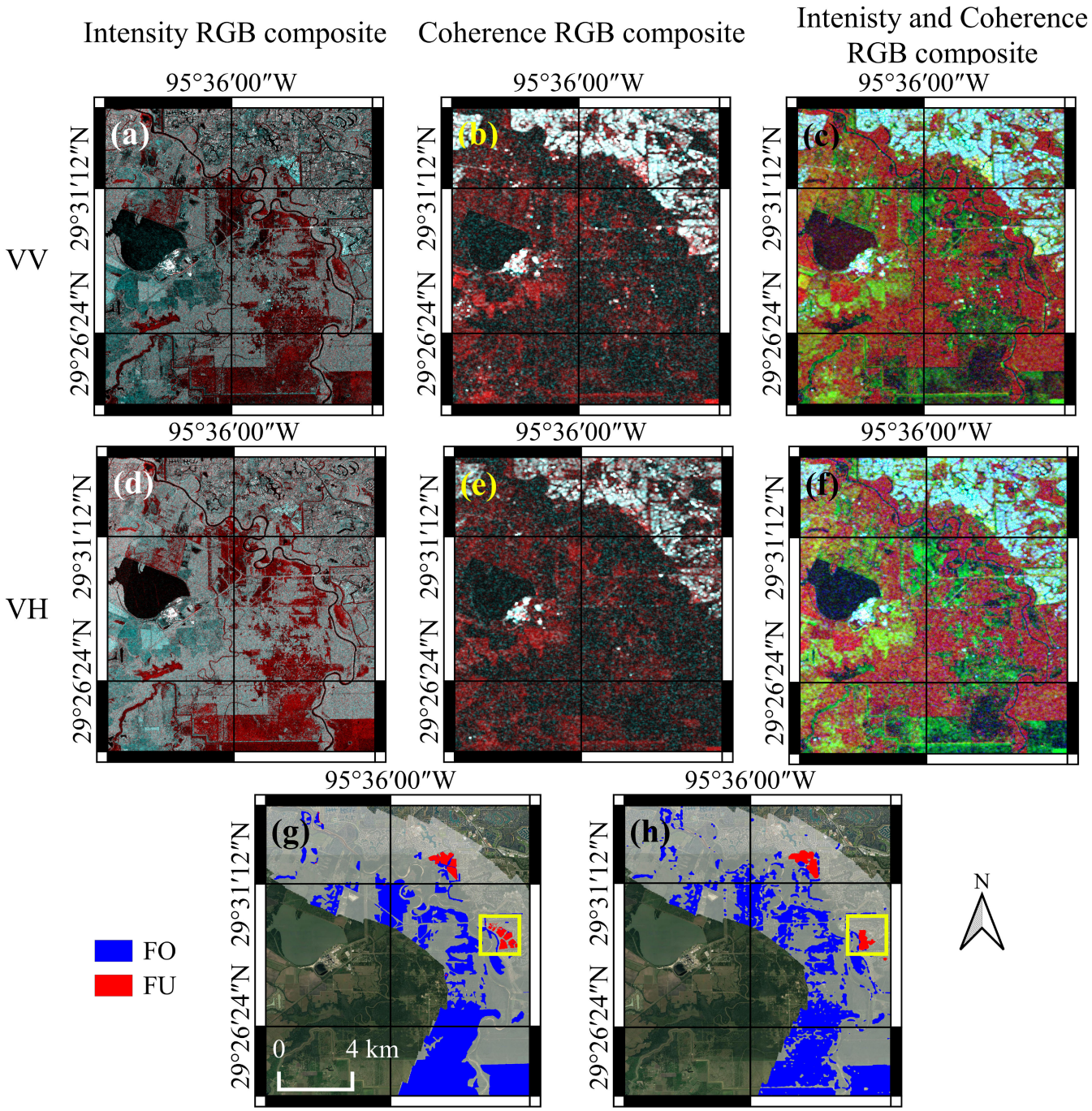


Fig. 5. RGB color composites of the Houston (USA) case study, corresponding reference, and our flood map. (a) Intensity RGB composite, R = intensity VV of August 24, 2017, and G = B = intensity VV of August 30, 2017. (b) Coherence RGB composite, R = coherence VV between August 18 and 24, 2017, and $G = B$ = coherence VV between August 24 and 30, 2017. (c) Intensity and coherence RGB composite, R = intensity VV of August 30, 2017, G = coherence VV between August 18 and 24, 2017, and B = coherence VV between August 24 and 30, 2017. (d)–(f) Using the data acquired in VH as the same RGB combinations of (a)–(c), respectively. (g) Reference. (h) Flood map inferred by the proposed model.

we argue that the overdetection of FO originates from a difference in the acquisition times between the SAR data and the reference. The reference for Koriyama (Japan) is derived from aerial photographs taken by GSI on October 13, 2019, while the Sentinel-1 data were acquired on October 12, 2019. Areas of underdetection of FO (yellow) are usually located near the edge of the flood extents where the terrain may not be totally submerged, and the backscatter tends to be slightly higher. Areas with overdetected FU (i.e., FU precision = 0.71) (black areas in Fig. 8) tend to be located near the edge of the flood. We assume that this overdetection of FU is due to a difference in the acquisition time of the SAR data and the reference and the fact that floodwater had receded in the meantime. Finally, the underdetection of FU is shown in cyan

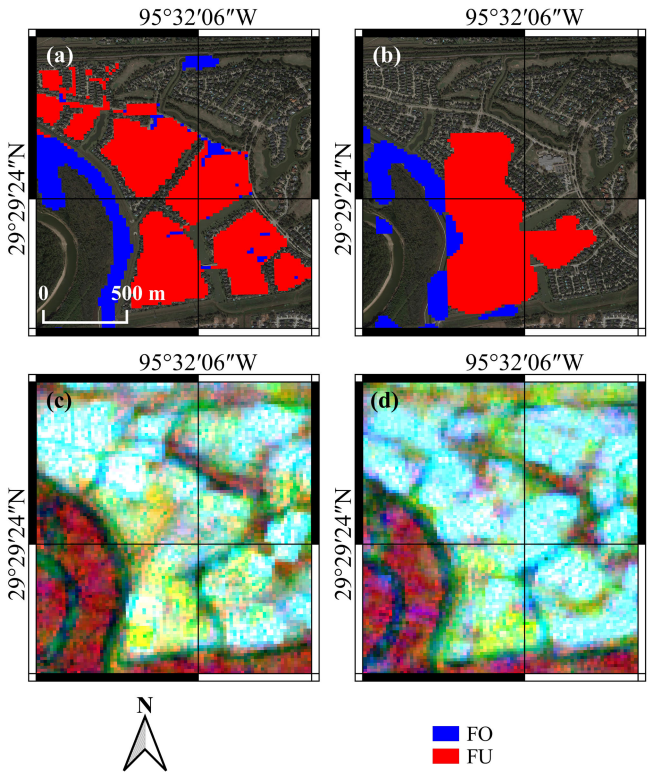


Fig. 6. Flood extent maps and the corresponding SAR data of the Houston flood case covering the areas in yellow rectangles in Fig. 5. (a) Reference. (b) Flood map. (c) Intensity and coherence RGB composite. R = intensity VV of August 30, 2017, G = coherence VV between August 18 and 24, 2017, and B = coherence VV between August 24 and 30, 2017. (d) Intensity and coherence RGB composite: R = intensity VH of August 30, 2017, G = coherence VH between August 18 and 24, 2017, and B = coherence VH between August 24 and 30, 2017.

in Fig. 8(a), with an FU recall value of 0.4. According to Fig. 7(b) and (e), the decrease in the coherence (VV and VH) in underdetected areas is not sufficiently high to enable the mapping of FU. This is because the general assumption in FU detection is that the appearance of floodwater in urban areas leads to a drop in coherence. However, the decrease in coherence is associated with pixels surrounding the buildings that give rise to the double- or multiple-bounce scattering mechanisms. Thus, the coherence may remain stable during flooding when it comes to the FU areas containing buildings that have large and flat roofs and relatively narrow streets [e.g., industrial areas shown in Fig. 8(b)]. Theoretically, the coherence is composed of amplitude and phase differences in InSAR data [56], and comparatively small changes in InSAR coherence may occur if the magnitude of the amplitude in InSAR coherence dominates the phase. In this case, no significant drop in coherence can be observed since the phase variations due to the presence of floodwater are only minor. As a result, FU may be underestimated as shown in Fig. 8(a). A similar occurrence of FU underdetections using InSAR coherence has been reported in [13], [22].

C. Beira (Mozambique) Flood Case

The RGB color composition of the Beira flood case is presented in Fig. 9, and the color interpretation is similar as described in Section V-A for the Houston (USA) flood case in Fig. 5. In Fig. 9(a) and (d), the cyan color can be attributed to bare soils and sparsely vegetated areas characterized by an increase in backscatter, arguably caused by an increase in soil moisture or the presence of double-bounce effects. For the RGB composite of coherence in VV [Fig. 9(b)] and VH [Fig. 9(e)], the decrease in coherence shown in red exists not only in FU but also in sparsely vegetated areas due to the random backscatter effects caused by vegetation. When it comes to the combination of intensity and coherence using the VV [Fig. 9(c)] and VH polarizations [Fig. 9(f)], areas in red represent high postevent intensity and low coherence in both preevent and coevent images. This behavior is typical for vegetated regions. The black to green areas represent low postevent intensity, moderate preevent coherence, and low coevent coherence. These areas are expected to cover FO. We argue that very dry soils with strong variations in soil moisture caused by the cyclone-related rainfall are in indicated in yellow-green. According to the quantitative evaluation for the Beira flood case in Table III, the Kappa coefficient (i.e., 0.81) suggests a good consistency between these two maps, while the F1 scores for FO and FU are 0.82 and 0.66, respectively. The relatively low recall (i.e., 0.74) indicates the existence of underdetection for FO. According to Fig. 10(a), the underdetection of FO (yellow) is mainly due to the inaccurate reference since wind-affected permanent water bodies are shown as floodwater in the reference. The moderate consistency with respect to FU (i.e., 0.59) demonstrates the overdetection of urban floods. The area in green rectangles in Fig. 10(a) is shown in Fig. 10(b), Fig. 10(d), and Fig. 10(e). As shown in Fig. 10(d) and Fig. 10(e), the overdetection of FU (black areas) in Beira is apparent in the areas characterized by small changes in coherence between preevent and coevent (<0.3) (shown in white and cyan). It is important to note that this area is not included in the manually derived reference map since the presence of water cannot be ascertained. The 0.75 recall of FU reveals underdetection, which is shown in orange in Fig. 10(d) and Fig. 10(e). The orange color is the result of a high postevent intensity, a moderate preevent coherence, and a low coevent coherence. A moderate preevent coherence was observed only in Beira. We assume that this is due to the fact that the urban areas of this study site contain buildings and bare soils, while the urban areas in Houston and Koriyama mainly correspond to buildings and impervious surfaces.

D. Beledweyne (Somalia) Flood Cases

Due to the unavailability of reference maps at the pixel scale for the Somalia flood cases (2018, 2019, 2020), only a qualitative evaluation at district scale has been carried out. Reference maps provided by UNOSAT classify different districts as totally flooded, partially flooded, or potentially flooded by visually inspecting optical and SAR data [Fig. 11(a)–(c)], while our results corresponding to flood maps are classified as flooded or not for each individual pixel [Fig. 11(d) and (f)]. To enable a comparison between our results and the UNOSAT reference maps, we generated similar categories by calculating

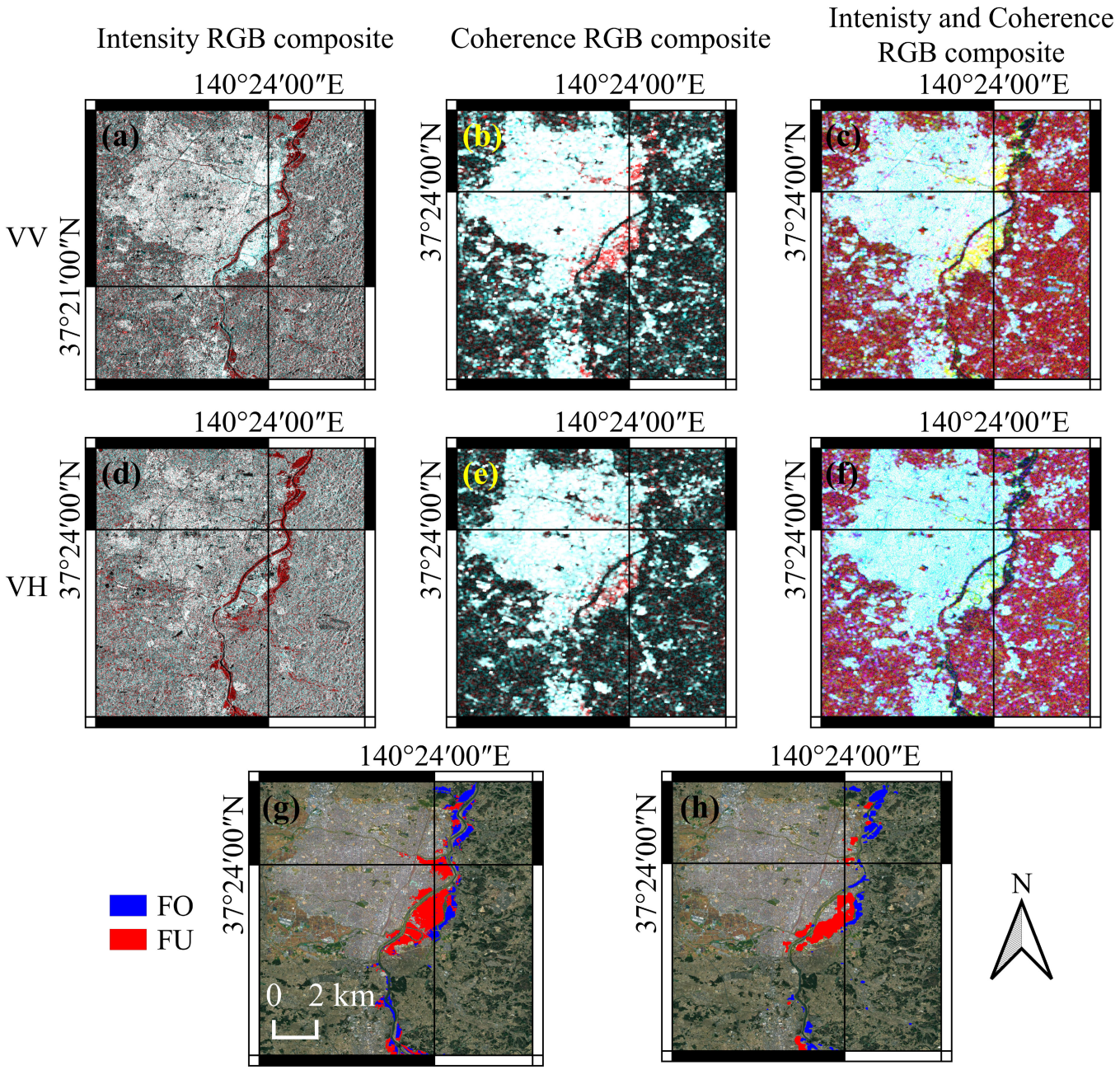


Fig. 7. RGB color composites of the Koriyama case study, the corresponding reference, and our flood map. (a) Intensity RGB composite, R = intensity VV of October 6, 2019, and G = B = intensity VV of October 12, 2019. (b) Coherence RGB composite, R = coherence VV between September 24 and October 6, 2019, and G = B = coherence VV between October 6 and 12, 2019. (c) Intensity and coherence RGB composite, R = intensity VV of October 12, 2019, G = coherence VV between September 24 and October 6, 2019, and B = coherence VV between October 6 and 12, 2019. (d)–(f) Using the data acquired in VH for the same RGB composites. (g) Reference. (h) Flood map inferred by the proposed model.

the percentage of pixels classified as flooded with respect to the total number of pixels (F_{perc}) comprising each of the polygons provided by UNOSAT. It is worth mentioning that the "potential flooded regions" class only exists in two of the polygons for the event in 2018. For this reason, we did not consider the two corresponding polygons in our F_{perc}-derived category flood maps. We plotted F_{perc} values, which for the different polygons varied between 1% and 71% for areas affected by the flooding, and defined a threshold to identify these two classes.

1) $F_{\text{perc}} \leq 29\%$: partially flooded.
2) $F_{\text{perc}} > 29\%$: completely flooded.

Assuming that only two classes are present in this area, the threshold is set by the OTSU automatic method [57] for F_{perc} values. For the Somalia flood in 2018, the reference maps, the flood extent maps, the different category flood maps, and

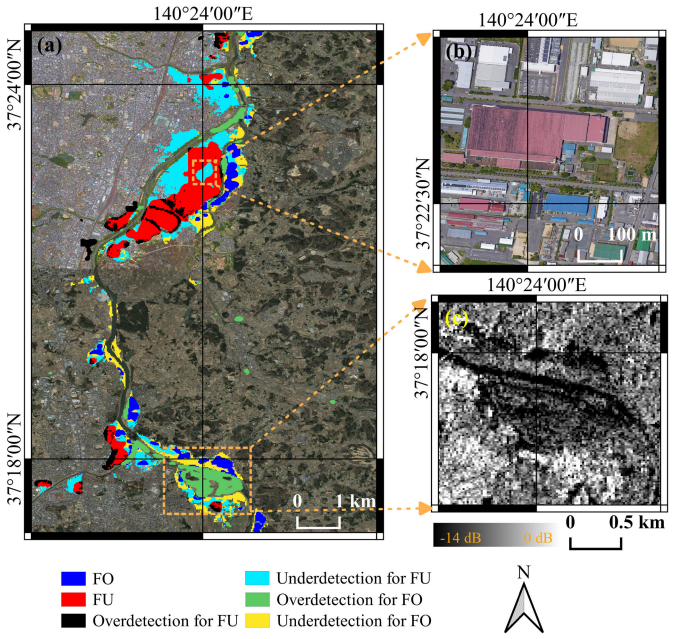


Fig. 8. (a) Comparison between reference flood category map and flood map inferred by the proposed model of the Koriyama test case. (b) Google Earth image. (c) Postevent Sentinel-1 image (VV).

the SAR data are shown in Fig. 11(a), (d), (g), and (j), respectively. In Fig. 11(a), yellow represents the areas that are potentially flooded, orange corresponds to partially flooded areas, while brown indicates completely flooded areas. By comparing Fig. 11(a), (d), and (g), it is clear that the largest part of the identified flood extent in Fig. 11(d) belongs to the "completely flooded areas" class (brown) in Fig. 11(a). As expected, the percentage of the area covered by flooding is much lower in the "partially flooded areas" class (orange) than in the "potentially flooded area" class (yellow), where almost no floods are detected. Fig. 11(g) and Fig. 11(a) shows a rather good consistency based on visual inspection. Moreover, areas in yellow in Fig. 11(j) have high postevent intensity (VV), high preevent coherence (VV), and low coevent coherence (VV). These are the expected characteristics of FU, and as a result, the corresponding areas are successfully identified as FU in Fig. 11(d). The areas in green in Fig. 11(j) have low postevent intensity (VV), moderate preevent coherence (VV), and low coevent coherence (VV). Again, this corresponds to the behavior characteristic of FO and the areas successfully detected as FO in Fig. 11(d). Similar phenomena were observed for the data of the Somalia flood that occurred in 2019 and 2020. Thus, we argue that our flood maps have identified flood extents satisfactorily.

VI. Discussion

To investigate the effectiveness of the proposed urban-aware module in the network, we first conducted a comparison between our proposed method and two previous methods (i.e., U-Net and CBAM U-Net). Then, two ablation studies for the urban-aware module and the input data were carried out separately. The detailed discussion is described in Section VI-A.

A. Comparing Three Different U-Net-Based Models

The U-Net [47], CBAM U-Net [29], and the urban-aware U-Net were trained as described in Section IV-B, and the experimental results are listed in Table III. The CBAM U-Net is considered since it uses an attention module based on the SAR data itself, while the proposed Urban-aware U-Net involves prior information (i.e., SAR-derived probabilistic urban mask), in addition to the SAR data. The results show that the two networks with implemented CBAM and urban-aware modules are able to generate flood maps achieving a higher Kappa coefficient when compared with the results inferred by the standard U-Net: our proposed model has the highest kappa (i.e., 0.83 and 0.81) in the Houston and Beira flood cases, while the CBAM U-Net reaches the highest kappa (i.e., 0.54) in the Koriyama flood case. In the proposed model result of the Koriyama case, underdetections of FU in the urban-aware U-Net exist because some bare soils and sparsely vegetated areas in the city of Koriyama were not classified as urban areas according to the SAR-derived probabilistic urban mask. Thus, only flooded buildings and not FU areas (i.e., mixture of buildings, vegetation, and bare soils) are classified as FU in the results of urban-aware U-Net. In terms of the F1 scores of FO, the CBAM variant of U-Net provides the best results. However, the difference between the results achieved with the three models is rather small and does not exceed 0.03. Hence, the performances of the three networks with respect to FO are rather similar at all the three study sites. On the other hand, the F1 score of FU significantly improves when using the proposed network. It reduces the overdetection in two test cases, i.e., 0.63 for Houston, 0.66 for Beira, while in Koriyama (Japan), the accuracy of FU is similar for the three models. Moreover, it is worth noting that the precision scores of FU in three test cases were increased by applying the proposed model. The difference in behavior in terms of FO and FU detections demonstrates that the identification of FU is a more challenging application. This is not surprising as the SAR characteristics of FU may vary according to different floodwater levels and the configuration of the urban fabric with respect to the SAR sensor. In comparison, the mechanisms determining the SAR backscatter in FO tend to be more stable. It is particularly challenging to learn the representative and robust features of the urban flood class with a rather limited training dataset. Our results indicate that in such cases with very limited data availability, the application of the "urban mask" enables the extraction of more relevant features for different flood classes. As a result, the proposed urban-aware module allowed us to achieve a better classification accuracy in terms of FU at all the three test sites.

To facilitate the intercomparison, the four most representative results covering the entire Sentinel-1 image frame are shown together in Fig. 12: the Houston (USA) flood, the Japan flood caused by the Hagibis Typhoon, the Beira (Mozambique) flood, and the Beledweyne (Somalia) flood that occurred in 2020. Due to the unavailability of independent references at the image-frame scale, the references were generated using a semiautomatic method [7], [12] according to the changes in SAR intensity and InSAR coherence. We are aware that

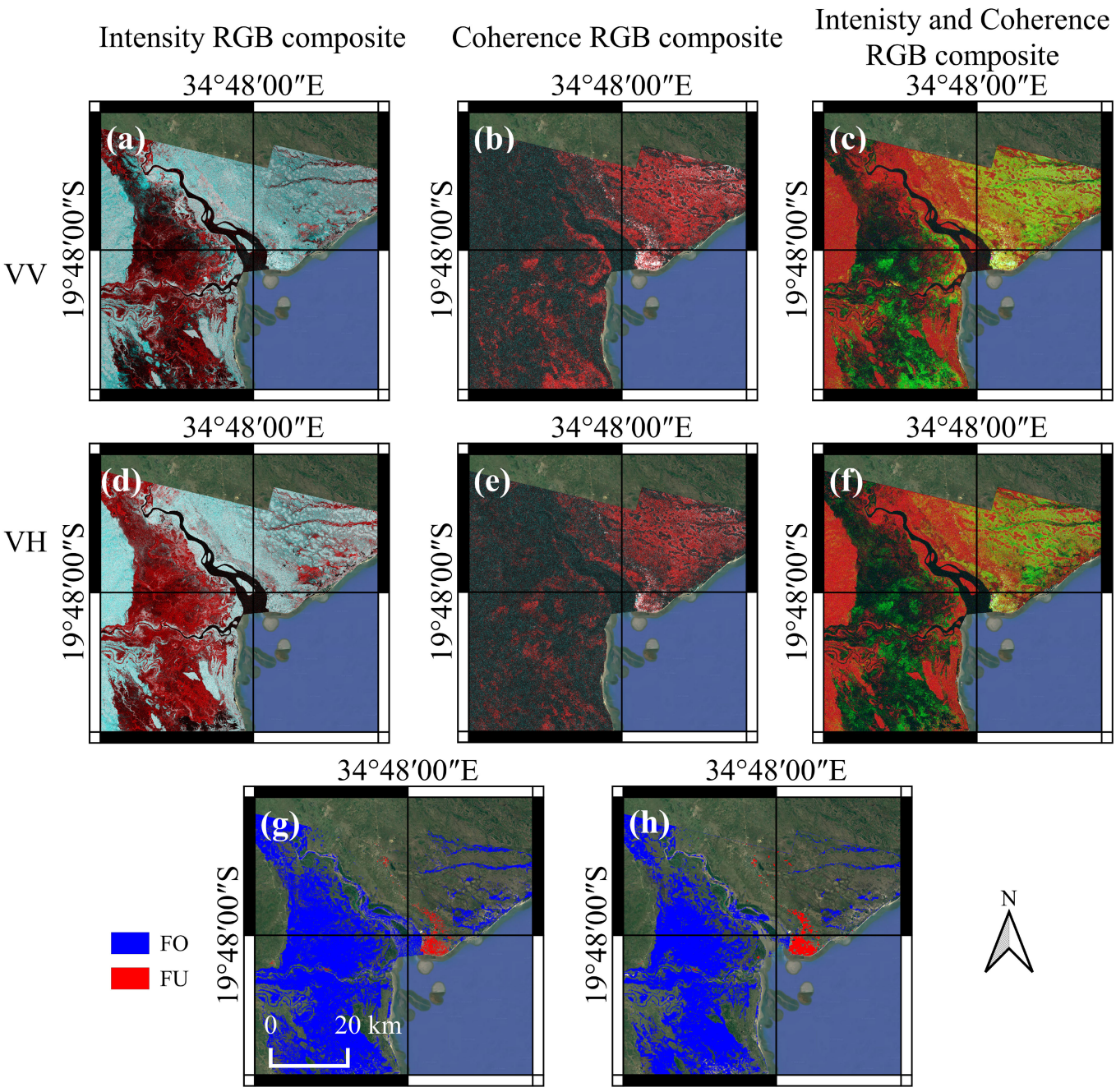


Fig. 9. RGB color composites of the Beira case study, the corresponding reference, and our flood map. (a) Intensity RGB composite, R = intensity VV of March 14, 2019, and G = B = intensity VV of March 20, 2019. (b) Coherence RGB composite, R = coherence VV between March 2 and 14, 2019, and G = B = coherence VV between March 14 and 20, 2019. (c) Intensity and coherence RGB composite, R = intensity VV of March 20, 2019, G = coherence VV between March 2 and 14, 2019, and B = coherence VV between March 14 and 20, 2019. (d)–(f) Using the data acquired in VH for the same RGB composites. (g) Reference. (h) Flood map inferred by the proposed model.

the reference datasets are not of equal quality. This hampers the intercomparison and makes it more difficult to draw general conclusions. In Fig. 12(a)–(d), the overestimation of FO that we observed for the Houston (USA) case study (in the yellow circle) could be reduced by applying the self-attention module (i.e., CBAM). Both the overdetections have been removed by applying the proposed urban-aware U-Net. Similar phenomena are observed for the Japan and Beira case studies: the overdetections of FU over the nonurban areas that are visible (in the cyan circle) in Fig. 12(f) and (g) were substantially reduced and almost completely disappeared in the flood map that was derived from the proposed urban-aware U-Net. Moreover, false alarms of FU in the north of Beira [in the green circle of Fig. 12(j) and (k)] were removed from the flood maps generated by urban-aware U-Net [Fig. 12(l)]. However, in the results covering the Beledweyne (Somalia) flood, the overdetections of FU still exist in flood maps generated via the CBAM U-Net and the proposed model

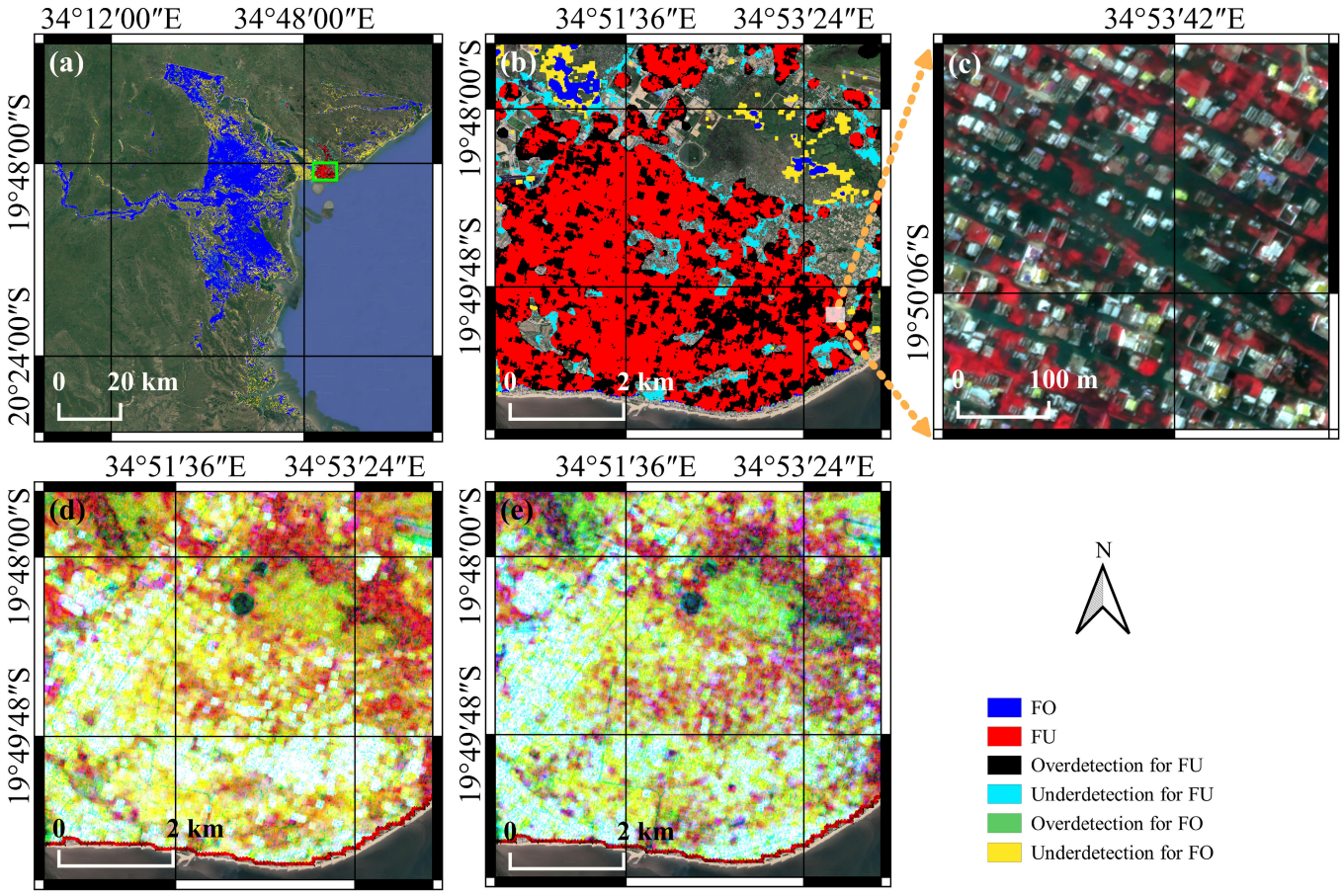


Fig. 10. Flood extent maps of the Beira test case. (a) Comparison between the flood map inferred by the proposed model and the reference. (b) Zoom-in of the green rectangular in (a). (c) False color composite of the Planet Skysat data for the white rectangular in (b). (d) Corresponding RGB combination of intensity and coherence R = intensity VV of March 20, 2019, G = coherence VV between March 2 and 14, 2019, and B = coherence VV between March 14 and 20, 2019. (e) Corresponding RGB combination of intensity and coherence R = intensity VH of March 20, 2019, G = coherence VV between March 2 and 14, 2019, and B = coherence VH between March 14 and 20, 2019.

TABLE III
EVALUATION OF INFERENCE FLOOD MAPS USING DIFFERENT CNN MODELS. THE HIGHEST VALUES IN KAPPA, PRECISION, RECALL, AND F1 FOR EACH CLASS ARE MARKED IN RED

Study site	Model	Kappa	FO			FU		
			Precision	Recall	F1	Precision	Recall	F1
Houston(US)	U-Net	0.80	0.89	0.84	0.86	0.22	0.57	0.32
	CBAM U-Net	0.82	0.83	0.94	0.88	0.27	0.65	0.38
	Urban-Aware U-Net (proposed)	0.83	0.87	0.88	0.87	0.76	0.54	0.63
Koriyama(Japan)	U-Net	0.53	0.45	0.55	0.50	0.52	0.59	0.55
	CBAM U-Net	0.54	0.51	0.54	0.52	0.52	0.58	0.55
	Urban-Aware U-Net (proposed)	0.51	0.50	0.50	0.50	0.71	0.40	0.51
Beira(Mozambique)	U-Net	0.56	0.95	0.73	0.83	0.02	0.98	0.04
	CBAM U-Net	0.57	0.92	0.79	0.85	0.04	0.96	0.08
	Urban-Aware U-Net (proposed)	0.81	0.93	0.74	0.82	0.59	0.75	0.66

[Fig. 12(o) and (p)]. To investigate this overdetection further, the corresponding SAR-based probabilistic urban mask and SAR data are analyzed.

Fig. 13 presents the flood map produced by applying the urban-aware U-Net, the SAR-derived probabilistic urban mask, and the RGB composite of intensity and coherence in VV and VH for the 2020 Beledweyne (Somalia) flood. In the probabilistic urban mask of Fig. 13(b), the probabilistic values of overdetected FU areas are slightly below the probabilistic values of urban areas, but higher than the probabilistic values of other nonurban areas without false alarms. This is to be expected to guide the model toward the selection of the distinctive features of FU and NF. However, the regions of overdetected FU and the regions with correctly mapped FU share similar values of intensity and coherence in VV and VH [Fig. 13(c) and (d)]. More specifically, the yellow color in Fig. 13(c) and (d) represents high postevent intensity, high preevent coherence, and low coevent coherence. Regarding

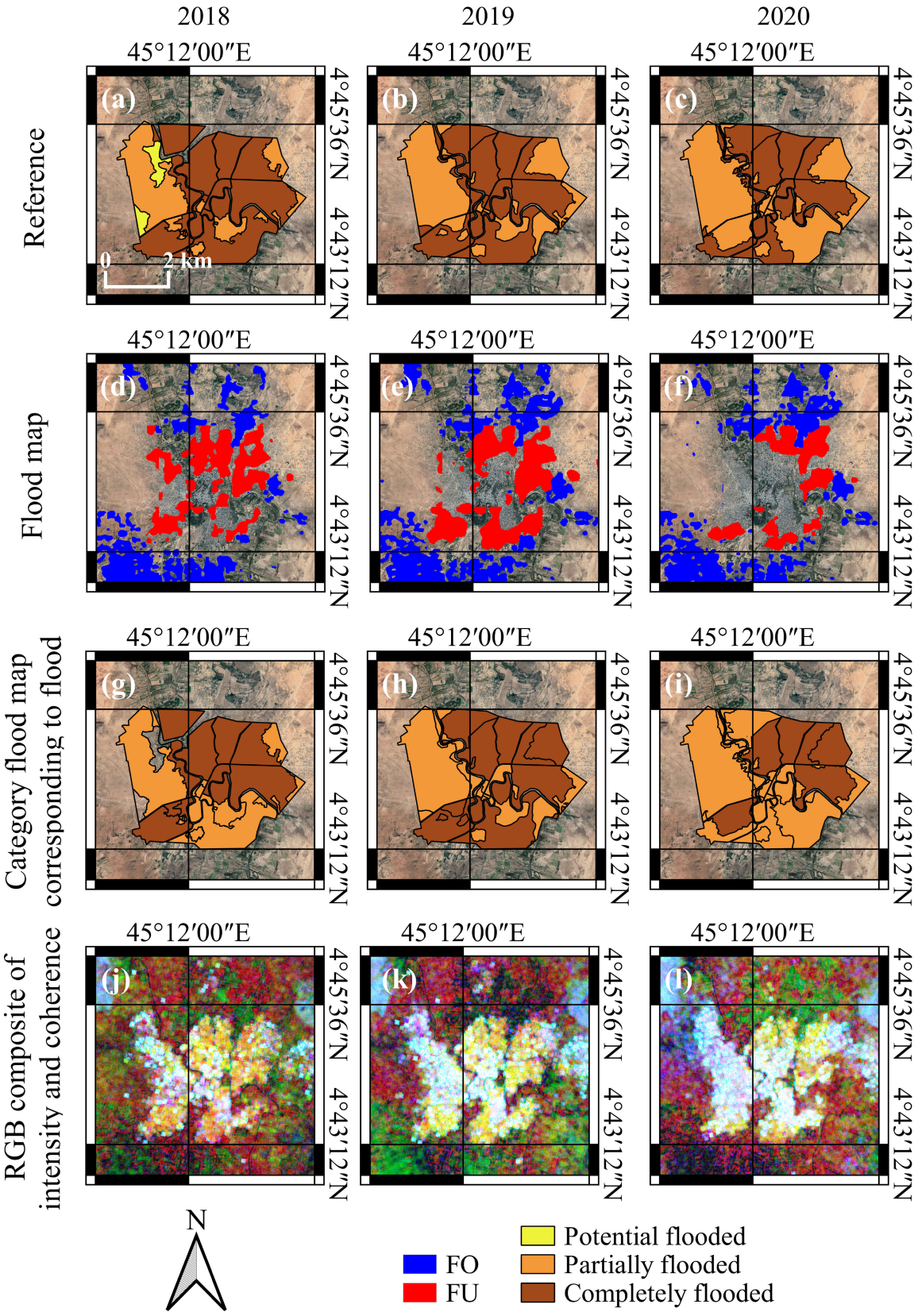


Fig. 11. Comparison of references, flood maps, and RBG composites of intensity and coherence for Somalia flood cases. (a)–(c) Satellite detected water distribution of Somalia flood in 2018, 2019, and 2020, respectively. (d)–(f) Our flood maps corresponding to (a)–(c). (g)–(i) Category flood maps generated according to the percentages of identified flood pixels in each polygon corresponding to (d)–(f). (j)–(l) RGB composite of intensity and coherence for Somalia flood in 2018, 2019, and 2020, respectively, R = postevent intensity VV, G = preevent coherence VV, and B = coevent coherence VV.

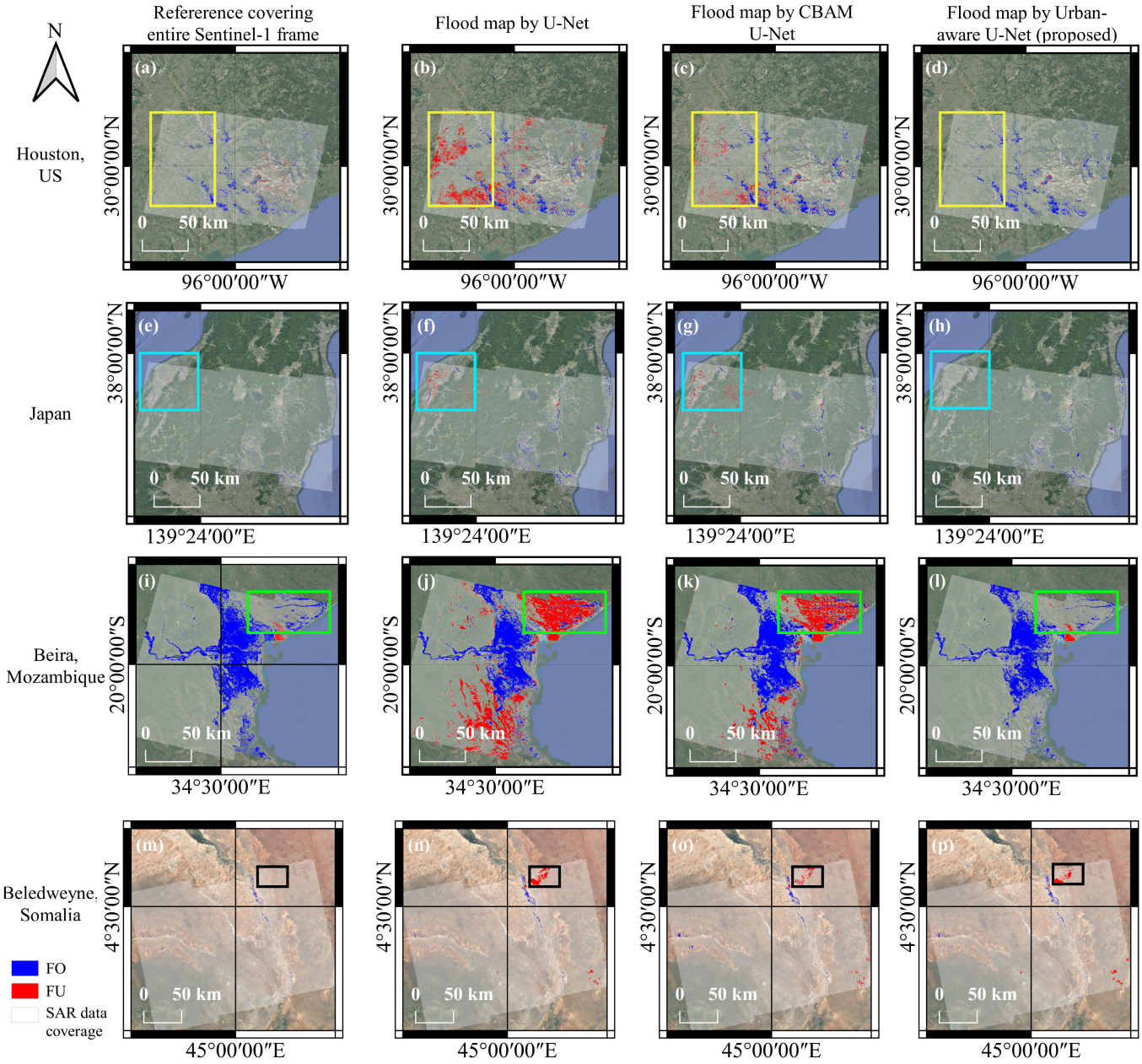


Fig. 12. Comparison of flood maps covering the entire Sentinel-1 image frame inferred by different models over three flood cases. (a), (e), (i), and (m) References generated by manual thresholding for the Houston (2017) flood case, Japan (2019) flood case, Beira (2019) flood case, and the Somalia (2020) flood case, respectively. (b), (f), (j), and (n) Flood map generated by U-Net corresponding to (a), (e), (i), and (m). (c), (g), (k), and (o) Flood map generated by CBAM U-Net corresponding to (a), (e), (i), and (m). (d), (h), (l), and (p) Flood map generated by Urban-aware U-Net (proposed) corresponding to (a), (e), (i), and (m).

high preevent/postevent intensity, the rough topography (high topographic variations) could be one of the reasons, as reported in [58]. Moreover, Wagner *et al.* [59] reported that rough karst rocks may cause strong backscatter returns in the C-band Advanced Scatterometer (ASCAT) data due to the presence of strong subsurface scatterers. Since our study site in Somalia has the exact same rough karst rocks, according to the World Karst Aquifer Map (WOKAM)/Karst aquifers map [59], and considering that we are using the C-band Sentinel-1 data, this specific soil type may explain the high pre- and postevent intensity in this case. In addition, high soil moisture could also be the reason for the high postevent intensity. The decrease in coherence may be due to scattering changes caused by ponds in the rough terrain. Thus, the quality of the results provided by the proposed urban-aware module may be lower when using an inaccurate SAR-derived probabilistic urban mask caused by "FU-lookalike" regions (e.g., extremely dry soils with shrublands) with similar SAR characteristics to FU as input. Such ambiguities lead to false alarms that we observed.

In addition to the study cases mentioned above, the algorithm was applied to four additional areas with no flooding to test the robustness of the proposed model with respect to the overdetection: data acquired on June 9, 2020, over Houston (USA), data acquired on June 26, 2020, over Koriyama and

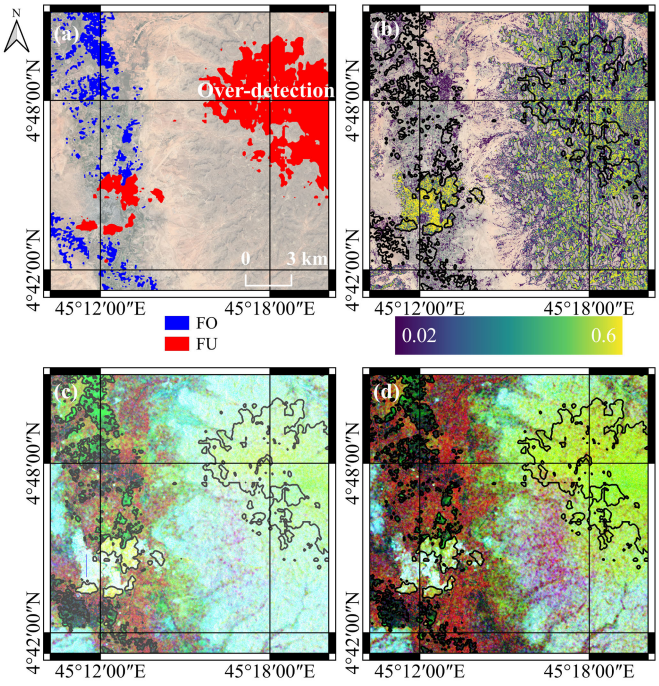


Fig. 13. Analysis of the overdetected FU in the flood map generated by the urban-aware U-Net for Beledweyne, (Somalia, 2020) flood case. (a) Flood map generated by the urban-aware U-Net. (b) Corresponding SAR-derived probabilistic urban mask, whose value is represented by different colors as shown in the color scale. (c) RGB composite: R = postevent intensity VV (May 20, 2020), G = preevent coherence VV (April 26–May 8, 2020), and B = coevent coherence VV (May 8–20, 2020). (d) RGB composite: R = postevent intensity VH (May 20, 2020), G = preevent coherence VH (April 26–May 8 2020), and B = coevent coherence VH (May 8–20, 2020). The boundaries of identified floods are shown in black in (b)–(d).

TABLE IV
COMPARISON OF THE TRAINING TIME AND COMPLEXITY OF EACH MODEL

Model	Training time for 100 epochs (min)	Number of parameters
U-Net	173	17,270,403
CBAM U-Net	207	22,846,675
Urban-aware U-Net	220	24,606,595

Iwaki (Japan), data acquired on February 6, 2019, over Beira (Mozambique), and data acquired on June 30, 2017, over Beledweyne (Somalia). According to our results (Fig. 14), only a few pixels located on the wind-affected permanent water bodies, bare soils, and image boundaries were identified as floods, while no large regions were identified as floods. Thus, the proposed urban-aware U-Net shows high robustness.

Moreover, the complexity and efficiency of the three models (i.e., U-Net, CBAM U-Net, and the urban-aware U-Net) were compared (Table IV). It appears that the time needed to complete the training phase increased from 207 to 220 min. Meanwhile, 1757k additional trainable parameters are computed in urban-aware U-Net when compared against CBAM U-Net. To understand the influence of the model's complexity on the flood maps' accuracy, a larger U-Net containing 25 606 275 parameters was tested. It generates similar results to U-Net: the Kappa is 0.72, 0.46, and 0.57 for Houston, Koriyama, and Beira, respectively; F1 for FO is 0.79, 0.36, and 0.72 for Houston, Koriyama, and Beira, respectively; F1 for FU is 0.42, 0.56, and 0.08 for Houston, Koriyama, and Beira, respectively. Therefore, we conclude that it is the urban-aware module itself rather than the complexity of the model that contributes to the improvement of flood map accuracy. Thus, with a large increase in urban flood mapping accuracy (Table III), the proposed urban-aware module does not lose much in terms of efficiency and increased computational complexity.

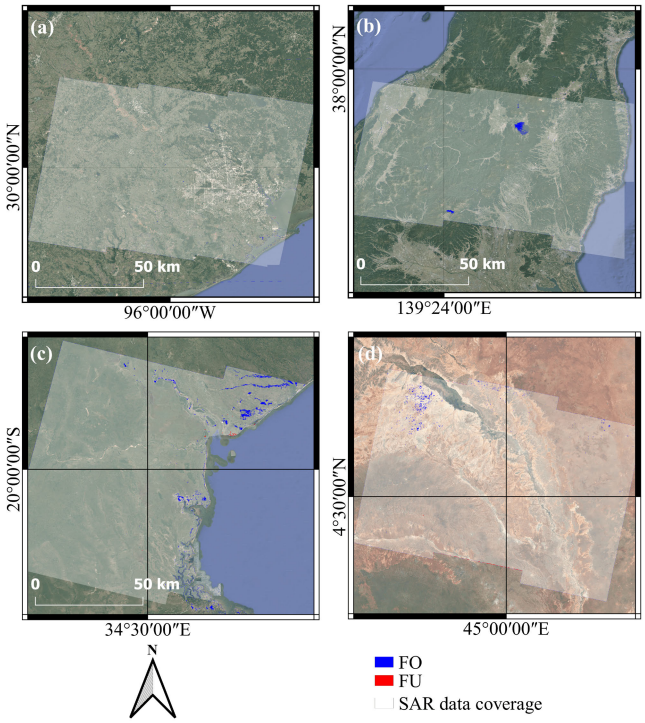


Fig. 14. Flood maps of four nonflood cases. (a) Houston (USA). (b) Japan. (c) Beira (Mozambique). (d) Beledweyne (Somalia).

B. Ablation Study for Urban-Aware Module

To investigate the effectiveness of each individual submodule in the proposed urban-aware module, we conduct the following ablation study by comparing three different model structures.

1) U-Net with the channel attention submodule.
2) U-Net with the urban-aware normalization submodule.
3) U-Net with the urban-aware module.

The three models were trained as described in Section IV-B, and the quantitative evaluation of three flood test cases (i.e., Houston test case, Koriyama test case, and Beira test case) is reported in Table V: for the Houston and Beira test cases, U-Net with the urban-aware normalization submodule achieved a higher F1 score in terms of FU (i.e., 0.65 and 0.57, respectively) than the results of U-Net with the channel attention submodule; for the Koriyama test case, U-Net with the channel attention submodule and U-Net with the urban aware normalization submodule have similar results. However, the combination of two submodules, i.e., urban-aware U-Net, always has the highest Kappa and the highest F1 score for both

TABLE V
COMPARISON OF THE ACCURACY OF FLOOD EXTENT MAPS USING DIFFERENT COMBINATIONS OF SUBMODULES IN THE URBAN-AWARE MODULE. THE HIGHEST VALUES IN KAPPA, PRECISION, RECALL, AND F1 FOR EACH CLASS ARE MARKED IN RED

Study site	Model	Kappa	FO			FU		
			Precision	Recall	F1	Precision	Recall	F1
Houston(US)	U-Net + channel attention submodule	0.80	0.91	0.81	0.86	0.44	0.54	0.48
	U-Net + urban aware normalization submodule	0.79	0.9	0.8	0.85	0.64	0.67	0.65
	Urban-Aware U-Net (proposed)	0.83	0.87	0.88	0.87	0.76	0.54	0.63
Koriyama(Japan)	U-Net + channel attention submodule	0.49	0.54	0.37	0.44	0.65	0.43	0.52
	U-Net + urban aware normalization submodule	0.48	0.53	0.37	0.44	0.64	0.43	0.51
	Urban-Aware U-Net (proposed)	0.51	0.50	0.50	0.50	0.71	0.40	0.51
Beira(Mozambique)	U-Net + channel attention submodule	0.70	0.97	0.65	0.78	0.08	0.78	0.15
	U-Net + urban aware normalization submodule	0.77	0.96	0.67	0.79	0.47	0.73	0.57
	Urban-Aware U-Net (proposed)	0.81	0.93	0.74	0.82	0.59	0.75	0.66

TABLE VI
EVALUATION OF U-NET AND CBAM U-NET USING NINE BANDS OF DATA: EIGHT BANDS OF SAR DATA AND A PROBABILISTIC URBAN MASK

Study site	Model	Kappa	FO			FU		
			Precision	Recall	F1	Precision	Recall	F1
Houston(US)	U-Net (9 bands)	0.82	0.82	0.95	0.88	0.3	0.66	0.41
	CBAM U-Net (9 bands)	0.66	0.95	0.61	0.74	0.26	0.49	0.34
Koriyama(Japan)	U-Net (9 bands)	0.53	0.48	0.56	0.52	0.62	0.47	0.53
	CBAM U-Net (9 bands)	0.41	0.6	0.24	0.34	0.7	0.38	0.49
Beira(Mozambique)	U-Net (9 bands)	0.69	0.93	0.81	0.87	0.04	0.91	0.08
	CBAM U-Net (9 bands)	0.55	0.99	0.51	0.67	0.06	0.8	0.11

the FO and FU classes. Thus, we conclude that our proposed U-Net with the urban-aware normalization plays an important role in improving the flood mapping accuracy in FU, while the combination of the channel attention submodule and the urban-aware normalization submodule produces the highest Kappa coefficient for the entire flood map.

C. Ablation Study for Input Data

To evaluate the effect of the way the urban mask is integrated, we also carried out an ablation study of inputs by stacking the urban mask to the original eight bands of SAR data (i.e., from eight bands of input data to nine bands of input data). We compared the different inputs for U-Net and CBAM U-Net with the same training process as presented in Section IV-B. The results of three flood test cases are reported in Table VI. By comparing the results in Table III and VI, U-Net with nine bands of input data achieves an overall improvement in comparison to U-Net with eight bands of input data in terms of Kappa coefficients and F1 scores, except for the Koriyama test case. No obvious improvements are found for CBAM U-Net by adding the urban mask as an additional input band. Instead, the performance degrades overall. Moreover, by comparing the results of Urban-aware U-Net, U-Net with nine bands, and U-Net with eight bands, the integration of the urban mask via the proposed urban-aware module provides greater improvements than the simple stacking of the urban mask together with SAR bands, especially for the FU class. For instance, a 0.31 improvement versus a 0.09 improvement in terms of FU F1 for the Houston test case, and a 0.62 improvement versus a 0.04 improvement in terms of FU F1 for the Beira test case. Hence, it is the way of integrating the urban mask information rather than the urban mask itself that plays the most essential role in the model performance enhancement. The same conclusion was reported in [60].

VII. CONCLUSION AND OUTLOOK

In this study, we proposed an urban-aware module to enable U-Net applications for large-scale SAR-based urban flood mapping. It should be noted that many supervised CNN models and available attention modules (e.g., attention-gate, channel attention, spatial attention, CBAM) cannot obtain informative features when only limited training datasets are available. This is often the case when applying such methods to map the extent of flooding and the issue becomes even more critical when aiming for the detection of floods in built-up environments. To solve this problem, we proposed an urban-aware module making use of an SAR-based probabilistic urban mask to improve the accuracy of flood extent maps by providing additional guidance. More specifically, the new module applies channel attention to extract informative features from each channel. Next, an urban-aware normalization submodule involving an SAR-based probabilistic urban mask was used to calibrate intermediate features with a learnable scale factor.

The proposed urban-aware module was tested in the U-Net image segmentation framework, i.e., urban-aware U-Net. The network was trained using datasets covering a small subset of Houston (USA) and the entire city of Iwaki (Japan). Then, the trained model was tested and evaluated in six flood cases and four nonflood cases with varying topography and land cover. In all the flood cases, our model produced flood maps that were consistent with the references, while the results in nonflood cases demonstrated the robustness of the proposed model. Moreover, a comparison between our urban-aware module and CBAM was performed. The qualitative and quantitative results obtained by the newly proposed approach confirm

that the urban-aware module is able to guide the network toward the most informative features, thereby leading to more accurate flood maps with a similar computation cost with respect to the available state-of-the-art attention modules applied for SAR-based urban flood mapping by considering the multitemporal SAR-derived prior information. However, some limitations remain because the proposed urban-aware module depends on the availability of a highly accurate urban mask. When an SAR-derived probabilistic urban mask of lower quality was considered, urban floods were overdetected. Overall, the results obtained in a total of ten cases (i.e., six flood cases and four nonflood cases) indicate that the well-trained urban-aware U-Net model shows a high potential for monitoring floods in near-real-time, especially when only limited annotation datasets are available in the training phase. Finally, through a series of experimental comparisons, our method outperformed other state-of-the-art methods (i.e., U-Net, CBAM U-Net) in terms of the F1-score of FU while maintaining a similar computational cost.

Further investigations are needed to better understand the advantages and limitations of the proposed method and to improve the classification accuracy and robustness of the approach: 1) analyzing the interferometric phase component instead of InSAR coherence to improve the accuracy of the classification in challenging land cover classes such as extremely dry areas; 2) adapting the proposed urban-aware module to other state-of-the-art image segmentation DL architectures with the aim of evaluating the flexibility and performance of the urban-aware module; and 3) applying the proposed module to unsupervised self-learning image segmentation paradigms to further mitigate the issue of insufficient annotated training datasets in urban flood mapping.

ACKNOWLEDGMENT

The authors would like to thank the two anonymous reviewers and editors for insightful comments which helped us to improve and clarify the article. The Skysat data are provided by Planet.

REFERENCES

[1] N. Najibi and N. Devineni, "Recent trends in the frequency and duration of global floods," *Earth Syst. Dyn.*, vol. 9, no. 2, pp. 757–783, Jun. 2018.

[2] B. Tellman *et al.*, "Satellite imaging reveals increased proportion of population exposed to floods," *Nature*, vol. 596, no. 7870, pp. 80–86, Aug. 2021.

[3] S. N. Willner, C. Otto, and A. Levermann, "Global economic response to river floods," *Nature Climate Change*, vol. 8, no. 7, pp. 594–598, Jul. 2018.

[4] N. I. Ulloa, S.-H. Chiang, and S.-H. Yun, "Flood proxy mapping with normalized difference Sigma-Naught index and Shannon's entropy," *Remote Sens.*, vol. 12, no. 9, p. 1384, Apr. 2020.

[5] J. Zhao, R. Pelich, R. Hostache, P. Matgen, W. Wagner, and M. Chini, "A large-scale 2005–2012 flood map record derived from ENVISAT-ASAR data: United Kingdom as a test case," *Remote Sens. Environ.*, vol. 256, Apr. 2021, Art. no. 112338.

[6] S. Martinis, J. Kersten, and A. Twele, "A fully automated TerraSAR-X based flood service," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 203–212, Jun. 2015.

[7] M. Chini, R. Hostache, L. Giustarini, and P. Matgen, "A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6975–6988, Dec. 2017.

[8] S. Grimaldi, J. Xu, Y. Li, V. R. N. Pauwels, and J. P. Walker, "Flood mapping under vegetation using single SAR acquisitions," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111582.

[9] L. Landuyt, F. M. B. Van Coillie, B. Vogels, J. Dewelde, and N. E. C. Verhoest, "Towards operational flood monitoring in flanders using Sentinel-1," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11004–11018, 2021.

[10] B. Zhang, S. Wdowinski, D. Gann, S.-H. Hong, and J. Sah, "Spatiotemporal variations of wetland backscatter: The role of water depth and vegetation characteristics in Sentinel-1 dual-polarization SAR observations," *Remote Sens. Environ.*, vol. 270, Mar. 2022, Art. no. 112864.

[11] L. Chang, Y.-T. Chen, J.-H. Wang, and Y.-L. Chang, "Rice-field mapping with Sentinel-1A SAR time-series data," *Remote Sens.*, vol. 13, no. 1, p. 103, Dec. 2020.

[12] M. Chini, R. Pelich, L. Pulvirenti, N. Pierdicca, R. Hostache, and P. Matgen, "Sentinel-1 InSAR coherence to detect floodwater in urban areas: Houston and hurricane Harvey as a test case," *Remote Sens.*, vol. 11, p. 107, Jan. 2019.

[13] M. Ohki, T. Tadono, T. Itoh, K. Ishii, T. Yamanokuchi, and M. Shimada, "Flood detection in built-up areas using interferometric phase statistics of PALSAR-2 data," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1904–1908, Nov. 2020.

[14] Y. Li, S. Martinis, and M. Wieland, "Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 178–191, Jun. 2019.

[15] Y. Li, S. Martinis, M. Wieland, S. Schlaffer, and R. Natsuaki, "Urban flood mapping using SAR intensity and interferometric coherence via Bayesian network fusion," *Remote. Sens.*, vol. 11, p. 2231, Jan. 2019.

[16] L. Moya, E. Mas, and S. Koshimura, "Learning from the 2018 western Japan heavy rains to detect floods during the 2019 Hagibis typhoon," *Remote Sens.*, vol. 12, no. 14, p. 2244, Jul. 2020.

[17] D. C. Mason, S. L. Dance, S. Vetra-Carvalho, and H. L. Cloke, "Robust algorithm for detecting floodwater in urban areas using synthetic aperture radar images," *J. Appl. Remote Sens.*, vol. 12, no. 4, 2018, Art. no. 045011.

[18] D. C. Mason, S. L. Dance, and H. L. Cloke, "Floodwater detection in urban areas using Sentinel-1 and WorldDEM data," *J. Appl. Remote Sens.*, vol. 15, no. 3, Feb. 2021, Art. no. 032003.

[19] M. Tanguy, K. Chokmani, M. Bernier, J. Poulin, and S. Raymond, "River flood mapping in urban areas combining RADARSAT-2 data and flood return period data," *Remote Sens. Environ.*, vol. 198, pp. 442–459, Sep. 2017.

[20] Y. N. Lin, S. H. Yun, A. Bhardwaj, and E. M. Hill, "Urban flood detection with Sentinel-1 multi-temporal synthetic aperture radar (SAR) observations in a Bayesian framework: A case study for hurricane Matthew," *Remote Sens.*, vol. 11, no. 15, p. 1778, Jul. 2019.

[21] L. Pulvirenti, M. Chini, N. Pierdicca, and G. Boni, "Use of SAR data for detecting floodwater in urban and agricultural areas: The role of the interferometric coherence," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1532–1544, Mar. 2016.

[22] L. Pulvirenti, M. Chini, and N. Pierdicca, "InSAR multitemporal data over persistent scatterers to detect floodwater in urban areas: A case study in Beletweyne, Somalia," *Remote. Sens.*, vol. 13, p. 37, Jan. 2021.

[23] R. Pelich, M. Chini, R. Hostache, P. Matgen, L. Pulvirenti, and N. Pierdicca, "Mapping floods in urban areas from dual-polarization InSAR coherence data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4018405.

[24] J. E. Ball, D. T. Anderson, and C. S. Chan, Sr., "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, Sep. 2017, Art. no. 042609.

[25] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2018.

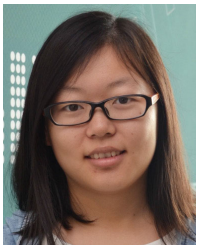
[26] X. X. Zhu *et al.*, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[27] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[29] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[30] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.
[31] S. Mohla, S. Pande, B. Banerjee, and S. Chaudhuri, "FusAtNet: Dual attention based SpectroSpatial multimodal fusion network for hyperspectral and LiDAR classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 92–93.
[32] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
[33] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.
[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[35] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.
[36] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
[37] Y. Wu and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[38] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2337–2346.
[39] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1501–1510.
[40] X. Li, W. Sun, and T. Wu, "Attentive normalization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 70–87.
[41] P. Zhao, J. Zhang, W. Fang, and S. Deng, "SCAU-Net: Spatial-channel attention U-Net for gland segmentation," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 670, Jul. 2020.
[42] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, and A. Plaza, "Remote sensing image superresolution using deep residual channel attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9277–9289, Nov. 2019.
[43] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[44] M. Chini, R. Pelich, R. Hostache, P. Matgen, and C. Lopez-Martinez, "Towards a 20 m global building map from Sentinel-1 SAR data," *Remote Sens.*, vol. 10, p. 1833, Nov. 2018.
[45] F. Ulaby and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*. Norwood, MA, USA: Artech House, 1989.
[46] H. Xie, L. E. Pierce, and F. T. Ulaby, "Statistical properties of logarithmically transformed speckle," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 3, pp. 721–727, Mar. 2002.
[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[48] I. K. Lee, J. C. Trinder, and A. Sowmya, "Application of U-Net convolutional neural network to bushfire monitoring in Australia with Sentinel-1/-2 data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 573–578, Aug. 2020.
[49] L. Hakim, N. Yudistira, M. Kavitha, and T. Kurita, "U-Net with graph based smoothing regularizer for small vessel segmentation on fundus image," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2019, pp. 515–522.
[50] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[51] NOAA. *Hurricane Harvey: Response Imagery of the Surrounding Regions*. Accessed: Aug. 23, 2022. [Online]. Available: https://storms.ngs.noaa.gov/storms/harvey/index.html#7/28.400/-96.690
[52] EOS. (Oct. 2019). *Flood Extent Map Generated for Japan Typhoon Hagibis*. Accessed: Aug. 23, 2022. [Online]. Available: http://eos-rs-products.earthobservatory.sg/EOS_ARIA-SG_201910_Japan_Typhoon_Hagibis/
[53] *Flood Extent Map Covering Koriyama for Japan Typhoon Hagibis*. Accessed: Aug. 23, 2022. [Online]. Available: https://www.gsi.go.jp/BOUSAI/R1.taihuu19gou.html
[54] UNOSAT. *Satellite Detected Water and IDP Distribution Over Belet Weyne Town*. Accessed: Aug. 23, 2022. [Online]. Available: https://unitar.org/maps/countries/93
[55] M. Marconcini *et al.*, "Outlining where humans live, the world settlement footprint 2015," *Sci. Data*, vol. 7, no. 1, pp. 1–14, Dec. 2020.
[56] F. T. Ulaby *et al.*, *Microwave Radar and Radiometric Remote Sensing*, vol. 4. Ann Arbor, MI, USA: Univ. Michigan Press, 2014.
[57] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[58] J. Zhao *et al.*, "Deriving exclusion maps from C-band SAR time-series in support of floodwater mapping," *Remote Sens. Environ.*, vol. 265, Nov. 2021, Art. no. 112668.
[59] W. Wagner *et al.*, "Widespread occurrence of anomalous C-band backscatter signals in arid environments caused by subsurface scattering," *Remote Sens. Environ.*, vol. 276, Jul. 2022, Art. no. 113025.
[60] Y. Sun, Y. Hua, L. Mou, and X. X. Zhu, "CG-Net: Conditional GIS-aware network for individual building segmentation in VHR SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.



Jie Zhao received the bachelor's degree in engineering of surveying and mapping from the Beijing University of Civil Engineering and Architecture (BUCEA), Beijing, China, in 2013, and the master's degree in Earth-oriented space science and technology (ESPACE) from Technische Universität München (TUM), Munich, Germany, in 2016. She is pursuing the Ph.D. degree with the Technische Universität Vienna (TUW), Vienna, Austria.

From 2017 to 2020, she worked as a Ph.D. Student with the Luxembourg Institute of Science and Technology (LIST), Esch-sure-Alzette, Luxembourg. She is currently a Project Assistant with the Remote Sensing Group, Department of Geodesy and Geoinformation, TUW. Her research interests include deep learning, flood mapping, synthetic aperture data analysis, and remote sensing applications.

Yu Li received the B.Sc. degree in geo-information science and technology and the M.Sc. degree in mathematical geology from the China University of Geosciences, Wuhan, China, in 2012 and 2015, respectively, and the Ph.D. degree in physical geography from Ludwig Maximilian University of Munich (LMU) of Munich, Munich, Germany, in 2020.

From 2016 to 2019, he worked as a Ph.D. Student with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He is currently a Junior Research and Technology Associate with the "Remote Sensing and Natural Resources Modeling" Group, Luxembourg Institute of Science and Technology (LIST), Esch-sure-Alzette, Luxembourg. His research interests include machine/deep learning, multitemporal image analysis, and natural disaster monitoring with SAR data.

Patrick Matgen received the M.Sc. degree in environmental engineering from the Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2002, and the Ph.D. degree in water engineering from the Delft University of Technology, Delft, The Netherlands, in 2011.

As a Project Manager from 2002 and then as a Lead Research and Technology Associate from 2015 until 2020, he was responsible for acquiring funding and managing Research Development and Innovation projects in Earth observation and hydrological-hydraulic modeling within the research unit "Environmental sensing and modelling" of the Luxembourg Institute of Science and Technology, Esch-sure-Alzette, Luxembourg. Since January 2020, he is leading the group "remote sensing and natural resources modelling" of the research unit.

Ramona Pelich (Member, IEEE) received the B.S. degree in signal and image processing from the "Traian Vuia" Polytechnic Institute of Timisoara, Timisoara, Romania, in 2010, the M.S. degrees in signal and image processing from the Institut Mines Télécom, Télécom Bretagne, Brest, France, in 2012, and the Ph.D. degree in remote sensing and image processing from the Institut Mines Télécom, Télécom, Bretagne, in 2015.

From 2012 to 2015, she has been with the Collecte Localization Satellites, France and the Maritime Surveillance Business Unit, Institut Mines Telecom, Télécom Bretagne. In 2016, she joined the Environmental Research and Innovation (ERIN) Department, Luxembourg Institute of Science and Technology, Esch-sure-Alzette, Luxembourg. Her research interests include signal and image processing techniques applied to remote sensing data. The developed scientific algorithms are applied in different fields such as large-scale flood hazard mapping, vessel monitoring, oil spill detection, and urban mapping.

Renaud Hostache received the M.Sc. degree in mechanics of geophysical media and environment from Joseph Fourier University, Grenoble, France, in 2003, the M.E. degree in hydraulics and environmental sciences from the Polytechnic Institute of Grenoble, Grenoble, in 2003, and the Ph.D. degree in water sciences from AgroParisTec, Montpellier, France, in 2006.

From 2007 to 2014, he was with the Centre de Recherche Public Gabriel Lippman, Belvaux, Luxembourg. From 2015 to 2021, he has been a Researcher with the Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg. Since 2022, he has been a Research Professor with the French Research Institute for Sustainable Development (IRD), Espace-Dev Research Unit, Montpellier. His research interests include exploitation of satellite Earth observation for flood and drought monitoring and integration of remote sensing observations in hydrodynamic models for reducing prediction uncertainty.

Wolfgang Wagner (Senior Member, IEEE) was borne in Austria in 1969. He received the Dipl.-Ing. degree in physics and the Dr.techn. degree in remote sensing from TU Vienna, Vienna, Austria, in 1995 and 1999, respectively.

In support of his master's and Ph.D. studies, he received fellowships to carry out research at National Aeronautics and Space Administration (NASA), European Space Agency (ESA), and the European Commission (EC) Joint Research Centre. From 1999 to 2001, he was with DLR. In 2001, he was appointed as the Professor for remote sensing at TU Vienna. He is a Co-Founder of the Earth Observation Data Centre (EODC). His main research interest is to gain physical understanding of the mechanisms driving the interaction of electromagnetic waves with the land surface. He has developed models for retrieving soil moisture and other land surface variables from scatterometer, SAR, and full-waveform lidar observations.

Dr. Wagner is a member of the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)/ESA Science Advisory Group for Meteorological Operational Satellite-Second Generation (METOP-SG) Scatterometer (SCA) and the ESA's Mission Advisory Groups for Sentinel-1 and Hydro Global Navigation Satellite System (GNSS). He was a recipient of the International Society for Photogrammetry and Remote Sensing (ISPRS) Frederick J. Doyle Award for his scientific contributions in active remote sensing. From 2008 to 2012, he served as the ISPRS Commission VII President, from 2009 to 2011 as the Editor-in-Chief of the Open Access Journal *Remote Sensing*, and from 2016 to 2019 as the Chair of the Global Climate Observing System (GCOS)/World Climate Research Programme (WCRP) Terrestrial Observation Panel for Climate.

Marco Chini (Senior Member, IEEE) received the M.S. degree in electronic engineering from the Sapienza University of Rome, Rome, Italy, in 2003, and the Ph.D. degree in geophysics from the University of Bologna, Bologna, Italy, in 2008.

From 2003 to 2012, he collaborated with the Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome. From 2006 and 2008, he was a Visiting Researcher with the Department of Aerospace Engineering Science, University of Colorado, Boulder, CO, USA, where he was involved in very high-resolution optical images for urban monitoring. From 2008 to 2012, he was with the Remote Sensing Group, Istituto Nazionale di Geofisica e Vulcanologia, Rome, where his research topic was the development of Earth Observation (EO)-based innovative algorithms for detecting damages caused by earthquakes. Since 2013, he has been with the Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg, formerly the Centre de Recherche Public—Gabriel Lippmann. He has been involved in projects for mapping flooded areas, detecting damage caused by earthquakes, and identifying risk-/hazard-prone areas. His research interests include analysis of multitemporal data, classification, feature extraction, data fusion and segmentation using synthetic aperture radar and optical data, soil moisture retrieval, and SAR interferometry technique for geophysical applications.

Dr. Chini is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, an Editorial Advisor of Springer's publication program in the *Remote Sensing/Photogrammetry*, and an Associate Editor of the *European Journal of Remote Sensing*.