

Urban Flood Mapping With Bitemporal Multispectral Imagery Via a Self-Supervised Learning Framework

Bo Peng , Graduate Student Member, IEEE, Qunying Huang, Jamp Vongkusolkrit, Song Gao, Daniel B. Wright, Zheng N. Fang, and Yi Qiang

Abstract—Near realtime flood mapping in densely populated urban areas is critical for emergency response. The strong heterogeneity of urban areas poses a big challenge for accurate near realtime flood mapping. However, previous studies on automatic methods for urban flood mapping perform infeasible in near realtime or fail to generalize well to other floods, for several reasons. First, multitemporal pixel-wise flood mapping requires accurate image registration, hindering the efficiency of large-scale processing. Although automatic image registration has been investigated, precisely coregistered multitemporal image sequence requires time-consuming fine tuning. Additionally, the floods may lead to the loss of many corresponding image points across multitemporal images for accurate coregistration. Second, existing unsupervised methods generally rely on hand-crafted features for floodwater detection. Such features may not well represent the patterns of floodwaters in different areas due to inconsistent weather conditions, illumination, and floodwater spectra. This article proposes a self-supervised learning framework for patch-wise urban flood mapping using bitemporal multispectral satellite imagery. Patch-wise change vector analysis is used with patch features learned through a self-supervised autoencoder to produce patch-wise change maps showing potentially flood-affected areas. Postprocessing including spectral and spatial filtering is applied to these patch-wise change maps to remove non-flood related changes. Final flood maps and parameter sensitivities were evaluated using several performance metrics. Two flood events from areas with differing degrees of urbanization were considered: Hurricane Harvey flood (2017) in Houston, Texas, and Hurricane Florence flood (2018) in Lumberton, North Carolina. The proposed method shows strong performance for self-supervised urban flood mapping.

Index Terms—Flood mapping, multispectral (MS) imagery, self-supervised learning, urban.

Manuscript received June 18, 2020; revised August 31, 2020 and December 15, 2020; accepted December 16, 2020. Date of publication December 28, 2020; date of current version February 1, 2021. This work was supported in part by the National Science Foundation under Grant 1940091, in part by the Microsoft AI for Earth Grant, in part by the Vilas Associates Competition Award from University of Wisconsin-Madison (UW-Madison), and in part by the Trewartha Graduate Research Award from Department of Geography at UW-Madison. (Corresponding author: Qunying Huang.)

Bo Peng is with the Department of Geography, Department of Electrical and Computer Engineering, University of Wisconsin - Madison, Madison, WI 53706 USA (e-mail: bo.peng@wisc.edu).

Qunying Huang, Jamp Vongkusolkrit, and Song Gao are with the Department of Geography, University of Wisconsin - Madison, Madison, WI 53706 USA (e-mail: qhuang46@wisc.edu; vongkusolkrit@wisc.edu; song.gao@wisc.edu).

Daniel B. Wright is with the Department of Civil and Environmental Engineering, University of Wisconsin - Madison, Madison, WI 53706 USA (e-mail: danielb.wright@wisc.edu).

Zheng N. Fang is with the Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX 76019 USA (e-mail: nickfang@uta.edu).

Yi Qiang is with the School of Geosciences, University of South Florida, Tampa, FL 33620 USA (e-mail: qiangy@usf.edu).

Digital Object Identifier 10.1109/JSTARS.2020.3047677

I. INTRODUCTION

THROUGHOUT the history of human civilization, floods have brought catastrophe to human settlements, including huge losses of life and property. As the most frequent natural disaster, floods account for more than 75% of federally declared disasters in the U.S. [1]. Records of flood events globally show that the number of individuals affected by floods is growing at an alarming rate [2]. In response, the United Nations (UN) has set the goal to rapidly and accurately respond to upcoming floods for protecting vulnerable people and mitigating economic losses, as stated in the UN Sustainable Development Goal 11 (2015–2030) [3]. To help meet this goal, improved methods for realtime flood extent mapping over dense urban regions to support flood response efforts are needed.

Remote sensing (RS) data have played an important role in large-scale flood extent mapping. Optical multispectral (MS) or hyperspectral (HS) satellite imagery in particular has unique advantages for identifying flooded (FL) areas by virtue of the abundant spectral information associated with floodwater [4]–[9]. Wieland *et al.* [7] developed an operational processing chain for flood extent mapping with Landsat and Sentinel-2 images. Wang *et al.* [6] used MS imagery from Landsat 8 to explore the role of normalized difference water index in super-resolution flood inundation mapping. Li *et al.* [4] proposed to use discrete particle swarm optimization for subpixel flood mapping on Landsat images. Recently, Du *et al.* [10] and Tong *et al.* [11] proposed improved particle swarm optimization methods for endmember extraction, which have great potential for subpixel flood mapping. These aforementioned flood mapping studies, however, have focused on rural areas with relatively homogeneous image backgrounds. Meanwhile, flood extent mapping is insufficiently investigated in urban areas due to heterogeneous land cover and land use, low spatial resolution of MS imagery, and lack of flood extent ground truth datasets [12], [13].

Satellite sensors are capable of scanning the entire earth surface with a high revisit frequency, abundant spectral bands, and high spatial resolution, which enable high-resolution mapping over heterogeneous urban areas. There have been a growing number of satellite constellations in orbit, including PlanetLab's [14] PlanetScope, RapidEye, and SkySat and Maxar/DigitalGlobe's [15] QuickBird, WorldView, IKONOS, and GeoEye. The raw pixel digital numbers in multitemporal optical imagery are usually inconsistent for the same land cover types due to the changing weather and illumination conditions [16]. Such inconsistency is a barrier to robust

floodwater detection from multitemporal RS images using data-driven machine learning models. Geometric and radiometric corrections are needed before surface reflectance imagery products can be useful for consistent spectral-based identification of FL areas [12], [13].

Regarding the mapping strategy, both pixel-wise [4], [5], [16]–[20] and patch-wise [12], [13], [21] models have been proposed for flood mapping. Pixel-wise flood mapping methods assign each pixel of the input imagery a label [e.g., FL or nonflooded (NF)], whereas patch-wise methods predict the class of an entire patch cropped from the imagery. Both pixel- and patch-wise mapping have been explored for flood extent mapping, since each has unique advantages in specific scenarios. Patch-wise mapping has been widely used for flood mapping over heterogeneous urban areas [12], [13], [21], for several reasons. First, patch-wise mapping helps mitigate the impact of errors from multitemporal image registration. It is challenging to have large-scale coregistered multitemporal high spatial resolution images in near realtime due to radiometric and geometric distortions [22]–[24]. Intensive labor work is required to fine-tune the corresponding image points for accurate coregistration. Such preprocessing is time-consuming for large-scale flood mapping, precluding application in near realtime. Moreover, since floodwaters may have covered a large part of the study area, many corresponding image points across the multitemporal images may not be available for image registration. Second, pixel-wise mapping requires pixel-wise human annotations for model training and validation, which is more time-consuming and labor-intensive than patch-wise mapping. Even if un-/self-supervised methods are used, human annotated validation data are still required for model evaluation over a small study area, which is expensive due to the heterogeneous image background over urban areas. For example, to map FL areas on an image of size 100×100 pixels (px), pixel-wise labeling requires 10^4 labels, whereas patch-wise labeling requires only 10^2 labels if the image is divided into nonoverlapping patches of size 10×10 px. In addition, it is more difficult to accurately label a pixel than a patch for heterogeneous urban areas.

It is worth noting that, for urban flood mapping with high spatial resolution imagery, the U.S. Federal Emergency Management Agency (FEMA) defined the National Flood Mapping Products [25] with FL areas that are either submerged or surrounded by floodwaters. As such, not only floodwater pixels but also their neighboring nonfloodwater pixels (e.g., building pixels surrounded by floodwaters) are included in FL areas. When combined with self-supervised learning, patch-wise mapping is able to address the aforementioned issues associated with pixel-wise mapping and to produce large-scale flood maps in near realtime. The patch-wise flood maps are consistent with FEMA's Flood Mapping Products, since non-floodwater pixels near floodwater pixels within the same patch are included in FL areas.

To our best knowledge, few studies have investigated self-/un-supervised patch-based methods using optical MS imagery for flood mapping over heterogeneous urban areas. To overcome the above limitations, this study proposes a self-supervised patch-based urban flood mapping method using bitemporal

pre- and postflood MS satellite imagery with four spectral bands including blue (B), green (G), red (R), and near infrared (NIR). Self-supervised learning is a special type of unsupervised learning [26], [27]. Self-supervised models are trained with automatically generated labels, which require no manual annotation for training. In this study, patch features were learned in a self-supervised manner using an autoencoder model [28]. We performed bitemporal patch change vector analysis (Patch-CVA) followed by spectral and spatial filtering to map urban floods in near realtime. The method was evaluated for two hurricane-induced flood events in the United States: Hurricane Harvey (2017) in Houston, Texas, and Hurricane Florence (2018) in Lumberton, North Carolina. Results show that the method provides strong performance and robust generalizability.

Major contributions of this study include the following.

- 1) The proposed method obviates massive human-annotated training data through a self-supervised learning framework. This enables the application of the method in large-scale and realtime, which offers the potential for its deployment in operational workflows at the front lines of emergency humanitarian assistance and disaster relief.
- 2) The flood maps generated by the proposed method for the two study areas demonstrate the method's robustness and generalizability in emergency response for upcoming floods.
- 3) The proposed method is resilient to nonflood related changes between the bitemporal data by leveraging spectral and spatial filtering, which effectively removed non-flood related changes for accurate flood mapping.
- 4) Although this framework is proposed to map urban floods in near realtime, it also paves the way for damage mapping in response to other types of disasters such as wildfires, earthquakes, etc.

The structure of this article is organized as follows. Related work is discussed in Section II. The study areas and datasets are described in Section III. The proposed method is presented in Section IV. Experimental results are summarized in Section V. Discussion about the results is given in Section VI. Finally, Section VII concludes this article.

II. RELATED WORK

Recent studies have shown promising results of supervised patch-based land cover mapping and its potential in flood mapping. Li *et al.* [21] developed an active self-learning convolutional neural network (CNN) to classify the synthetic-aperture-radar (SAR) image patches into three classes (i.e., NF, FL with buildings, and FL without buildings). Peng *et al.* [12], [13] designed a Siamese CNN model to evaluate the patch similarity for identification of FL MS image patches. Song *et al.* [29] and Sharma *et al.* [30] proposed CNN-based models to map land cover with superior performance compared with pixel-based methods, especially in heterogeneous urban areas.

The aforementioned CNN based models demonstrate the power of data-driven supervised deep learning and computer vision in flood mapping by leveraging an increasing volume of RS data with massive human labels. However, such a

time-consuming manual labeling process poses further challenges for near realtime flood mapping and emergency response in a large scale. Additionally, these models often fail to generalize well when applied to other floods or locations [12], [16], [27].

To address the challenges associated with supervised methods, unsupervised pixel-wise flood mapping with bitemporal pre- and postflood imagery has been explored with as few human annotations as possible through change detection [24], [31]–[34] followed by automatic thresholding techniques such as the minimum error thresholding proposed by Kittler and Illingworth (KI) [35], [36] and Otsu’s method [37], [38]. In 2009–2010, the IEEE Geoscience and Remote Sensing Society launched a Data Fusion Contest [34] for flood mapping based on multitemporal change detection, which involves both supervised and unsupervised flood mapping on optical MS and SAR data. In the category of unsupervised flood mapping with optical MS data, the winning algorithm [34] used the NIR band of pre- and postflood imagery with an unsupervised clustering algorithm, leveraging the high absorption of water in the NIR band. Byun *et al.* [24] proposed an unsupervised change detection approach to pixel-wise flood mapping based on bitemporal MS image fusion with detection of spectral distortion. Schlaffer *et al.* [32] conducted harmonic analysis of multitemporal SAR imagery to identify FL pixels, which showed strong deviations from the corresponding NF pixels. Giustarini *et al.* [33] developed a hybrid framework integrating SAR backscatter thresholding, FL region growing, and change detection for flood mapping using bitemporal SAR imagery. Du *et al.* [39] recently proposed the unsupervised deep slow feature analysis for unsupervised pixel-wise change detection based on bitemporal MS imagery, which can be further applied in flood mapping.

The previous works discussed above showed good results of flood mapping in an unsupervised manner without intensive human labeling of training data. Unfortunately, those unsupervised methods focused on pixel-wise mapping, which are not directly applicable to patch-wise mapping over heterogeneous urban areas.

To sum up, patch-wise flood mapping over heterogeneous urban areas using a self-supervised deep learning approach is still missing in previous works to the best of the authors’ knowledge, which is the main objective of this study.

III. STUDY AREA AND DATASETS

This work investigates self-supervised mapping of urban floods in two different cities in the United States (U.S.), including the 2017 Hurricane Harvey flood in Houston, Texas (see Fig. 1), and the 2018 Hurricane Florence flood in Lumberton, North Carolina (see Fig. 2). We choose these two study areas because both floods involve dense residential, industrial, and commercial areas. Experiments based on these two study areas will validate the effectiveness of the proposed method in flood mapping over heterogeneous urban areas.

For each flood event, the data used in this study were bitemporal pre- and postflood MS imagery from PlanetLab [14] covering the same geographic area. All images are surface reflectance products with four spectral bands (i.e., B, G, R, and NIR). Spatial

TABLE I
SPECIFICATIONS FOR HARVEY AND FLORENCE MS IMAGES

Specifications	Harvey	Florence
Hurricane Landfall Date	August 25, 2017	September 14, 2018
Image Date (Before)	July 31, 2017	August 31, 2018
Image Date (After)	August 31, 2017	September 18, 2018
Spectral Band	B, G, R, NIR	
Height, Width (px)	(1850, 3070)	(2240, 2940)
Spatial Resolution (m)	3	
Pre-processing	Surface Reflectance	
Total # of Patches	56,795	65,856
Ratio of FL	0.1777	0.1916

resolution of both datasets over the two study areas are 3 m of ground sampling distance.

This work performed patch-wise flood extent mapping, where nonoverlapping corresponding pre- and postflood patches were cropped from the before and after bitemporal MS images, respectively. Each patch is of size 10×10 px and thus covers ground area of 30×30 m. The goal of this study is to classify the postflood patches into FL or NF.

Although the proposed method is self-supervised (i.e., without human annotated training labels), ground truth data are needed for quantitative evaluation of experimental results. This ground truth data were generated by manually labeling the datasets for the two study areas. All post-flood image patches were labeled by visual inspection of the very high resolution (VHR) imagery with spatial resolution of 0.3 m from National Oceanic and Atmospheric Administration (NOAA) collected on the same day as those from PlanetLab. We cropped patches from NOAA’s VHR imagery, with each patch covering the same geographical area (i.e., 30×30 m) as the colocated PlanetLab MS patch. Each NOAA VHR patch thus contains $100 \text{ px} \times 100 \text{ px}$ ($100 = 30/0.3$), which was labeled by three expert annotators based on the FL area within the patch. The ground truth label of each postflood patch was determined by the majority vote out of all three annotations. Patches with a tiny negligible FL area or with floodwaters under tree canopy were intentionally labeled as NF [12], [21]. 56 795 labeled patches (10 094 FL and 46 701 NF) were collected for the Harvey dataset and 65 856 labeled patches (12 619 FL and 53 237 NF) were collected for the Florence dataset. More detailed specifications of the datasets are summarized in Table I.

IV. METHODOLOGY

A. Problem Formulation

Given a pair of pre- and postflood MS images ($\mathbf{I}^{\text{pre}}, \mathbf{I}^{\text{post}}$), we cropped M rows by N columns nonoverlapping patch pairs $(\mathbf{I}_{i,j}^{\text{pre}}, \mathbf{I}_{i,j}^{\text{post}})$, where $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$ are patch coordinates. Each patch is with 10×10 px and four bands. This study proposed a fully self-supervised framework $\mathcal{F}$ for generating the flood map $\mathbf{P}$ by

$$\begin{aligned} \mathbf{P} &= \mathcal{F}(\mathbf{I}^{\text{pre}}, \mathbf{I}^{\text{post}}) \\ &= \mathcal{F}_4(\mathcal{F}_3(\mathcal{F}_2(\mathcal{F}_1(\mathbf{I}^{\text{pre}}, \mathbf{I}^{\text{post}})))) \end{aligned} \quad (1)$$

where $(\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{F}_4)$ denote the four interlocking modules in the framework $\mathcal{F}$, including 1) $\mathcal{F}_1$: *patch encoding* for patch

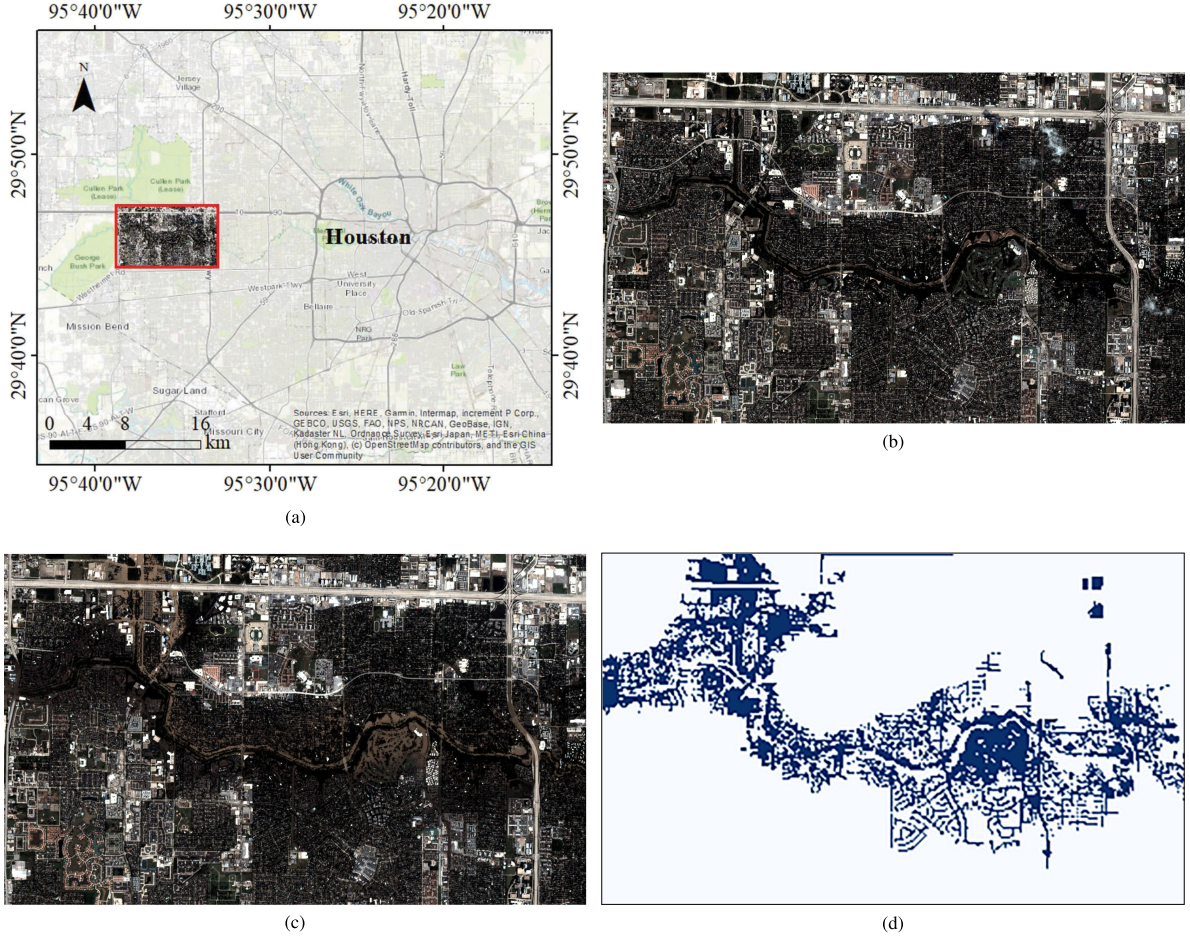


Fig. 1. Harvey data. (a) Study area of Harvey flood. (b) Satellite image before flood. (c) Satellite image after flood. (d) Manually classified ground truth of patch-wise flood extent map.

feature extraction, 2) $\mathcal{F}_2$: *Patch-CVA* for *initial* flood mapping, 3) $\mathcal{F}_3$: *spectral filtering* to remove NF changes for *intermediate* flood mapping, and 4) $\mathcal{F}_4$: *spatial filtering* to remove noise for *final* flood mapping. Fig. 3(a) illustrates the workflow of the proposed framework.

Since the pre and post-flood images were captured with a time interval of less than one month, we assume that the land cover changes over the study areas mainly resulted from floods. Accordingly, patch similarity estimation based on patch feature vectors was used for patch change detection and *initial* flood mapping. The more similar $(\mathbf{I}_{i,j}^{\text{pre}}, \mathbf{I}_{i,j}^{\text{post}})$, the less likely that $\mathbf{I}_{i,j}^{\text{post}}$ was indeed FL.

Furthermore, *initial* flood maps based on patch change detection may lead to unexpected false alarms due to noise and isolated NF changes. Using the spectral signature of floodwaters and the spatial topology of FL areas, we conducted further spectral filtering to remove NF changes and spatial filtering to remove minor isolated or noisy changes for accurate FL patch detection.

B. Patch Encoding

The *Patch Encoding* module learned the multidimensional features of both pre- and postflood patches $(\mathbf{I}_{i,j}^{\text{pre}}, \mathbf{I}_{i,j}^{\text{post}})$, for patch

change detection. Specifically, each pair of $(\mathbf{I}_{i,j}^{\text{pre}}, \mathbf{I}_{i,j}^{\text{post}})$ were fed into a pretrained Siamese patch *Encoder* to extract the representative features $(\mathbf{c}_{i,j}^{\text{pre}}, \mathbf{c}_{i,j}^{\text{post}})$. Patch change detection was then conducted in the *Patch-CVA* module based on the encoded patch features $(\mathbf{c}_{i,j}^{\text{pre}}, \mathbf{c}_{i,j}^{\text{post}})$.

To enable self-supervised learning of patch features without human annotated labels, we developed an autoencoder model, which encodes the high-dimensional input into low-dimensional features and then decodes the features for reconstructing the input. As such, the autoencoder model was trained with patches as both the input and the target. The *Encoder* of the pretrained autoencoder was then used for encoding the patches. The network architecture of the autoencoder developed in this study is shown in Fig. 3(b) with hyperparameters listed in Table II, which includes *Encoder* and *Decoder* subnetworks. The *Encoder* consists of several layers of convolutional nets (Conv), while the *Decoder* is composed of a stack of transposed convolutional nets (ConvTrans).

We trained the Siamese autoencoder for the pre- and postflood patches with shared weights for further patch change detection. 80% pairs of pre- and postflood patches were used for training, and the rest 20% for validation. We took batches of bitemporal patch pairs $(\mathbf{I}_{i,j}^{\text{pre}}, \mathbf{I}_{i,j}^{\text{post}})$ as both inputs and targets of the autoencoder, using the Adam optimizer [40] along with the L1 loss,

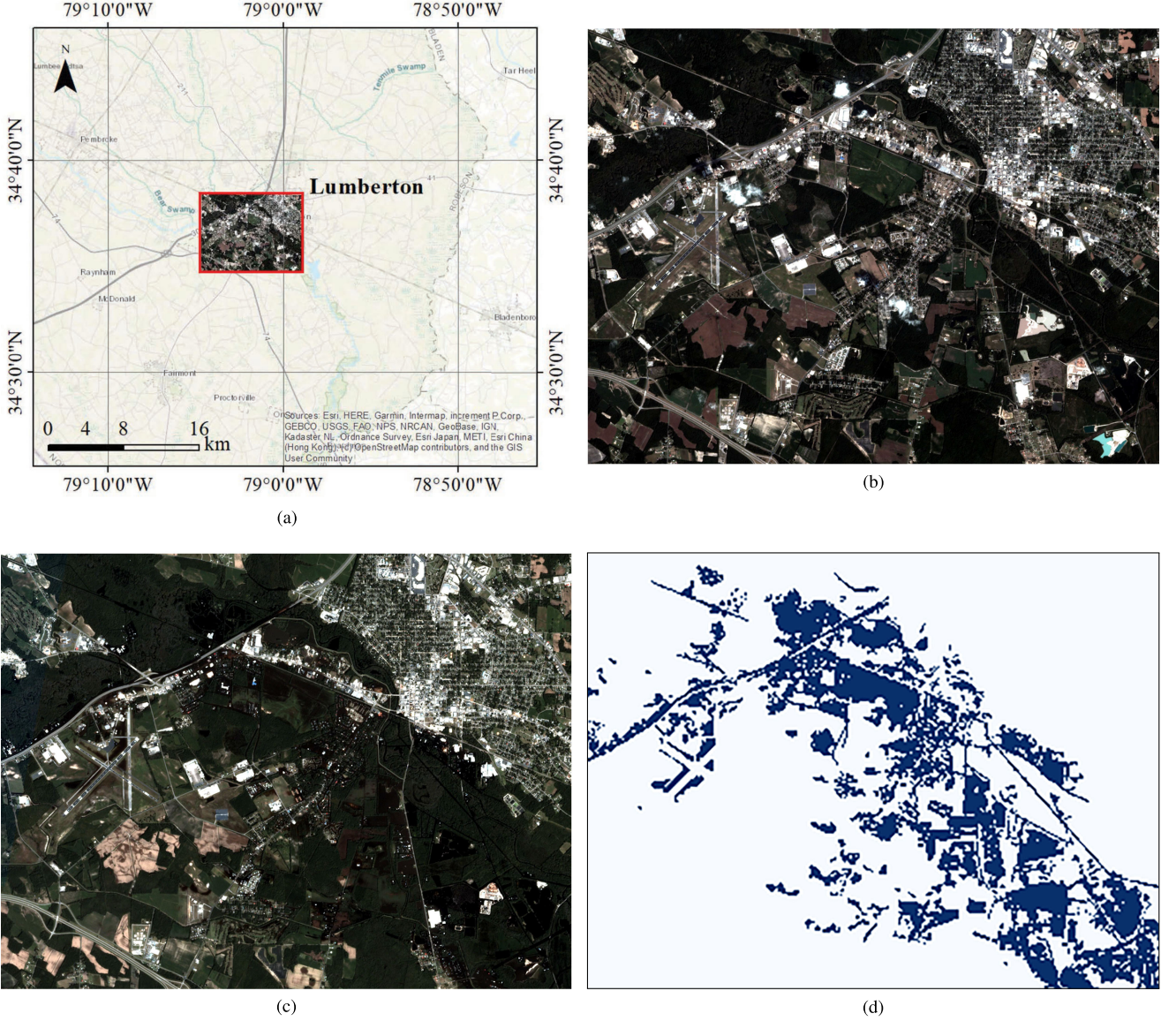


Fig. 2. Florence data. (a) Study area of Florence flood. (b) Satellite image before flood. (c) Satellite image after flood. (d) Manually classified ground truth of patch-wise flood extent map.

$\mathcal{L}(x, y)$, defined as follows:

$$\mathcal{L}(x, y) = \frac{1}{B} \sum_{i=1}^B |x_i - y_i| \quad (2)$$

where x and y denote the output and target of the model, and B is the number of patches in each batch. In the autoencoder model, the target is the input itself. It should be noted that the class imbalance (i.e., the relative frequencies of FL and NF patches) of the datasets (see Table I) does not undermine the performance of patch change detection for FL patch detection. This is a major advantage of self-supervised patch change detection for flood mapping compared to supervised models trained with highly imbalanced datasets, in which class weights typically must be set during training [12], [21].

We start with the learning rate of 0.01 and reduce it by a factor of 10 if the validation loss does not decrease for 10

consecutive epochs. Default momentum parameters $(\beta_1, \beta_2) = (0.9, 0.999)$ of the Adam optimizer were used. Weight decay was set to $1e-5$. For better model generalization, common data augmentations were used, including random rotation with degrees in $[0^\circ, 90^\circ, 180^\circ, 270^\circ]$, random vertical and horizontal flipping, and normalization to the range of $[0, 1]$. After training with 150 epochs, the pretrained *Encoder* was then used to encode both pre- and postflood patches for further patch change detection.

C. Patch-CVA

The *Patch-CVA* module followed the widely used technique of CVA [41]–[44] to select potential FL patches. We computed the magnitude and direction of change between pairs of pre- and postflood patch features $(c_{i,j}^{\text{pre}}, c_{i,j}^{\text{post}})$. The magnitude component of the change is defined as the Euclidean distance (ED) between

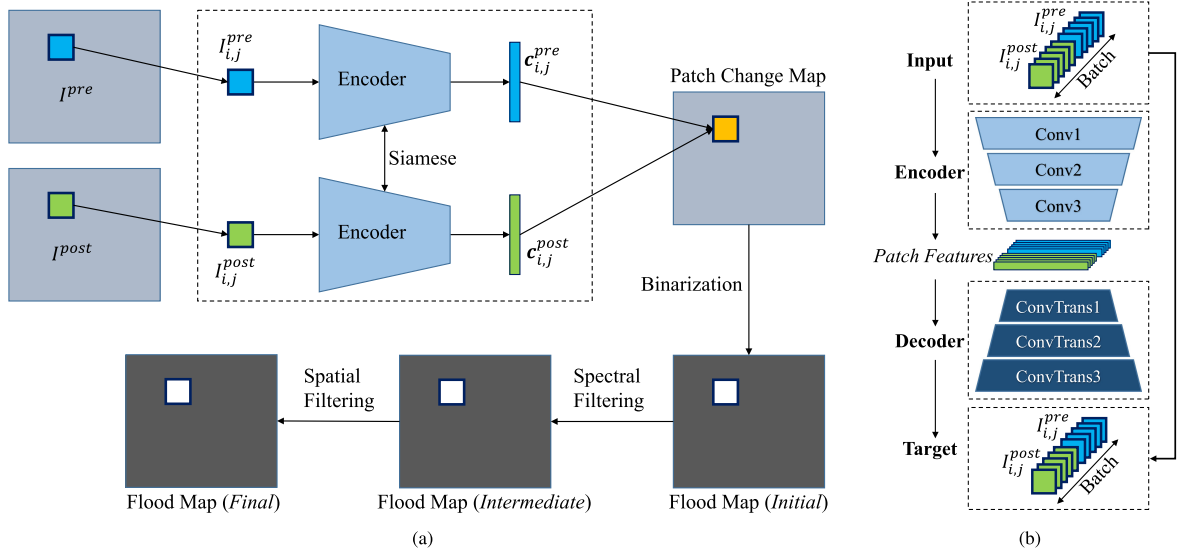


Fig. 3. (a) Self-supervised framework for patch-wise urban flood extent mapping. (b) Autoencoder architecture for training the patch Encoder.

TABLE II
AUTOENCODER HYPERPARAMETERS

Module	Operation	Parameters
Input	Image Patches	Size: [Batch, 4, 10, 10]
Encoder	Conv1	Convolution (out: 64, kernel: 3) Batch Normalization LeakyReLU (0.01) Max-pool (2)
	Conv2	Convolution (out: 128, kernel: 3) Batch Normalization LeakyReLU (0.01) Max-pool (2)
	Conv3	Convolution (out: 64, kernel: 1) Sigmoid
Decoder	ConvTrans1	ConvTransposed (out: 128, kernel: 1) Batch Normalization LeakyReLU (0.01) Upsample (scale: 2, bilinear)
	ConvTrans2	ConvTransposed (out: 64, kernel: 3) Batch Normalization LeakyReLU (0.01) Upsample (scale: 2, bilinear)
	ConvTrans3	ConvTransposed (out: 4, kernel: 3) Sigmoid
Target	Same as Input	Size: [Batch, 4, 10, 10]

the $c_{i,j}^{pre}$ and $c_{i,j}^{post}$ as follows:

$$ED(c_{i,j}^{pre}, c_{i,j}^{post}) = \|c_{i,j}^{pre} - c_{i,j}^{post}\|_2. \quad (3)$$

Suppose the patch feature vector consists of K elements, traditional direction change was defined as $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_K)$, where α_k is computed by the following:

$$\cos(\alpha_k) = \frac{c_{i,j}^{pre}(k) - c_{i,j}^{post}(k)}{ED(c_{i,j}^{pre}, c_{i,j}^{post})} \quad (4)$$

where $c_{i,j}^{pre}(k)$ and $c_{i,j}^{post}(k)$ denote the k th elements in $c_{i,j}^{pre}$ and $c_{i,j}^{post}$, respectively.

However, the above direction change in traditional CVA is not often used [43] because the direction component in traditional CVA often leads to a quantity α with the same dimension as the input change vectors and even more complex high-dimensional change maps for further change detection. Hence, we used the cosine similarity score (5) to measure the angle between the pre- and postflood patch feature vectors, which indicated the direction component of the change.

$$\cos\langle c_{i,j}^{pre}, c_{i,j}^{post} \rangle = \frac{(c_{i,j}^{pre})^\top (c_{i,j}^{post})}{\|c_{i,j}^{pre}\|_2 \|c_{i,j}^{post}\|_2} \quad (5)$$

where $\cos\langle c_{i,j}^{pre}, c_{i,j}^{post} \rangle$ denotes the cosine score of the angle between $c_{i,j}^{pre}$ and $c_{i,j}^{post}$.

Following the maps of magnitude and direction changes, binarization through automatic thresholding was used to select initial potential FL patches. We applied the technique proposed by Rosin [45], [46] to select the threshold, where the *corner* of the histogram was selected as the threshold. Rosin [45], [46] assumed that the histogram of the change map of either magnitude or direction is a unimodal distribution with one dominant population with respect to the secondary population. Specifically, the selected point on the histogram corresponding to the threshold is the most distant from the line between the peak and the end of the histogram (see Fig. 4).

Based on the above thresholding technique, we binarized both magnitude and direction change maps to obtain the *initial* flood maps.

D. Spectral Filtering

The *initial* flood maps may contain multiple types of changes not related to floodwaters—shadows, human built-up areas, errors in radiometric or geometric corrections, etc. Since all FL

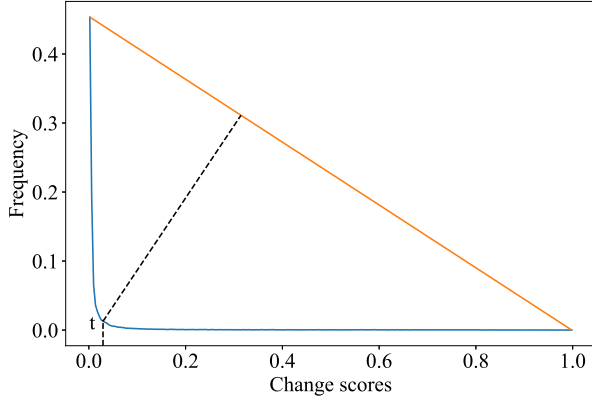


Fig. 4. Rosin's method [45], [46] for selection of the threshold t based on the histogram of the magnitude or direction change map.

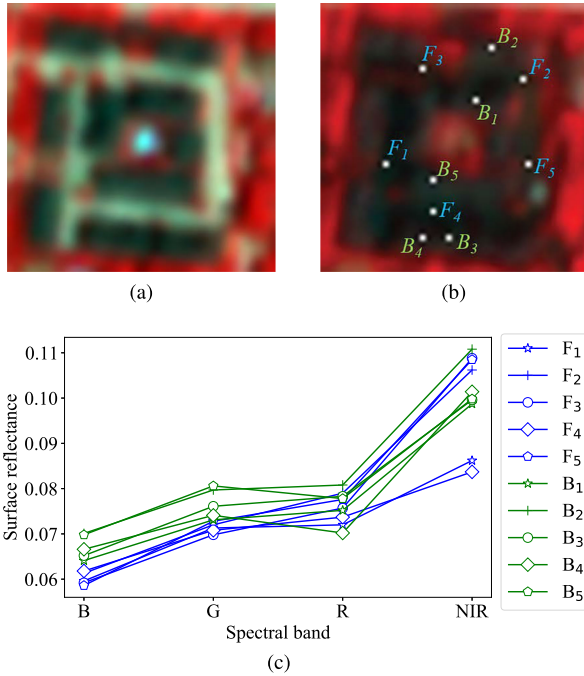


Fig. 5. Buildings and floodwaters in dense neighborhoods show similar spectral features. (a) False-color image of a neighborhood before flood. (b) Corresponding image after flood. (c) Spectral plots of five building samples ($B_1, B_2, \dots, B_5$) and five floodwater samples ($F_1, F_2, \dots, F_5$).

postflood patches contain floodwater pixels, spectral features of the *initial* flood maps were investigated to remove nonflood related changes.

Floodwater typically has lower surface reflectance than that of other major land cover types such as built-up areas and vegetation. We performed a pixel-wise binary unsupervised clustering for all pixels in potential post-flood FL patches based on the *initial* flood maps. In this study, k -means [47] clustering on raw pixel spectra was used for both Harvey and Florence datasets. It should be noted that the spectra of some buildings in dense neighborhoods are very similar to those of floodwaters (see Fig. 5), leading to misclassification of some building pixels. By leveraging Microsoft building footprints [48], we removed

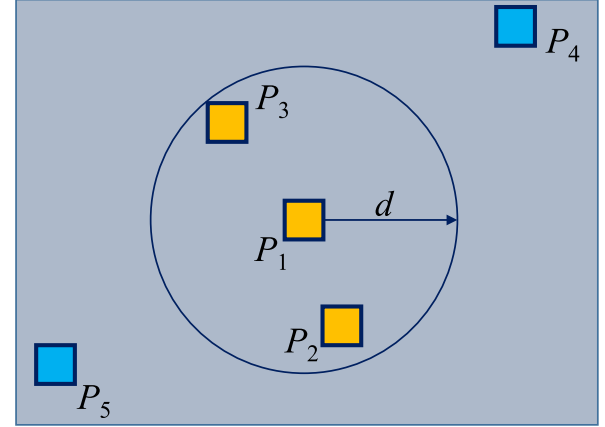


Fig. 6. Patch connectivity: patches (P_1, P_2, P_3) marked in yellow are connected since they are within a distance of d patches, whereas patches (P_4, P_5) marked in blue are not connected to any patches in the map since they are far away from other patches.

building pixels classified as FL and thus obtained *intermediate* flood maps.

E. Spatial Filtering

After *Spectral Filtering*, there can still exist false alarms in the *intermediate* flood map due to small isolated areas previously identified as FL patches due to noise, shadow, and other errors. Often, these areas are isolated and far from major FL areas. The spatial topology of flood movement means that that FL patches tend to be close to each other. Therefore, we assume that FL patches are connected in major FL areas. FL patches need not be strictly adjacent to each other in heterogeneous urban areas, however, since patches with floodwaters under trees are defined as NF [12], [21], leading to disconnection of some FL patches. Therefore, we defined in this study that FL patches were connected if they were within a predefined maximum distance of d patches as illustrated in Fig. 6.

After identification of major FL areas through connectivity analysis, small isolated areas were removed if they contained less than a predefined minimum number of a patches in the *intermediate* flood maps as they exhibited false alarms. Finally, the *final* flood maps were generated based on both *direction* and *magnitude* change, respectively.

F. Performance Evaluation

This work investigated the problem of FL patch detection. With ground truth labels, we computed the number of true positives (TPs), false positives (FPS), true negatives (TNS), and false negatives (FNS), respectively. Therefore, we quantitatively evaluated the performance of the proposed framework using precision (Pr), recall (Re), $F1$ score, and overall accuracy (OA), defined as

$$Pr = \frac{TPs}{TPs + FPS}$$

$$Re = \frac{TPs}{TPs + FNS}$$

TABLE III
TIME CONSUMPTION ON TRAINING THE AUTOENCODERS AND ENCODING ALL IMAGE PATCHES WITH THE PRETRAINED ENCODERS FOR HARVEY AND FLORENCE DATASETS

Dataset	Training (mins)	Encoding (mins)
Harvey	40.97	0.33
Florence	46.76	0.37

$$F1 = \frac{2}{1/Pr + 1/Re}$$

$$OA = \frac{TPs + TNs}{TPs + FPs + TNs + FNs} \quad (6)$$

Additionally, as the proposed framework generated the flood maps, we also evaluated the accuracy based on the intersection over union (IoU), also known as the Jaccard Index [49]. Given the ground truth flood map $\mathbf{G}$ and the predicted map $\mathbf{P}$, IoU is defined as

$$IoU(\mathbf{G}, \mathbf{P}) = \frac{\text{Area of Intersection of } (\mathbf{G}, \mathbf{P})}{\text{Area of Union of } (\mathbf{G}, \mathbf{P})} = \frac{|\mathbf{G} \cap \mathbf{P}|}{|\mathbf{G} \cup \mathbf{P}|} \quad (7)$$

V. EXPERIMENTAL RESULTS

This section describes the application of the proposed method for patch-wise urban flood mapping over two study areas in the U.S., including 1) the 2017 Hurricane Harvey Flood at Houston, Texas and 2) the 2018 Hurricane Florence Flood at Lumberton, North Carolina. The *Patch Encoding* models in Section IV-B were trained on a server with 64-b Ubuntu 16.04.6 LTS and a 24 GiB Titan RTX GPU based on PyTorch [50]. Table III lists the time consumption on training the autoencoders for 150 epochs and encoding all the pre- and postflood image patches with the pretrained encoders for both Harvey and Florence datasets, respectively.

All other experiments were conducted on a Dell workstation with an Intel(R) Xeon(R) W-2125 CPU @ 4.00 GHz $\times$ 8, 16 GiB RAM, and 64-b Window 10.

A. Flood Event 1: Harvey

1) *Results of Patch-CVA*: Using the pretrained encoder described in Section IV-B, we encoded both pre- and postflood patches into low-dimensional feature vectors. Fig. 7 shows the patch-wise change maps generated based on the changes of direction [see Fig. 7(a)] and magnitude [see Fig. 7(c)], respectively. Both change maps were normalized into the range of [0, 1].

By leveraging Rosin's thresholding technique discussed in Section IV-C, we selected the thresholds of both direction and magnitude change maps based on their corresponding histograms [Fig. 7(b) and 7(d)].

We thus obtained two different binary classification maps [see Fig. 8(a) and 9(a)] associated with the *direction* and *magnitude* change maps, respectively. Compared with ground truth labels in Fig. 1(d), we highlighted *TPs* in blue, *FPs* in red, and *FNs* in yellow. Corresponding to Figs. 8(a) and 9(a), we evaluated the

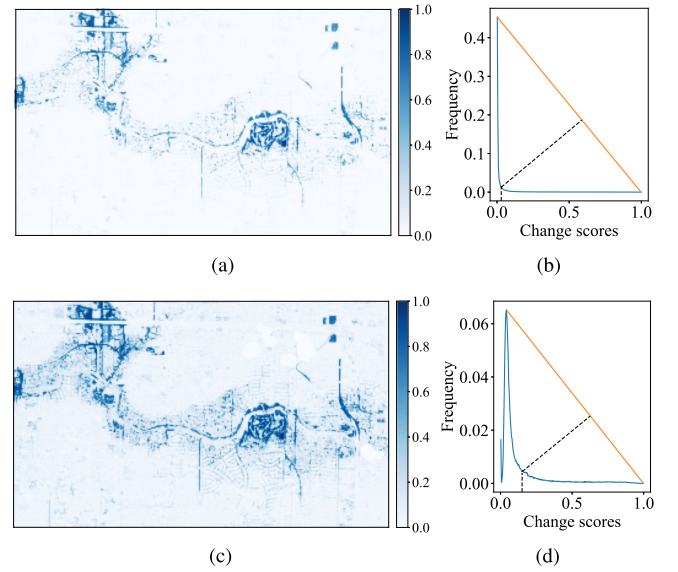


Fig. 7. Harvey change maps based on encoded pre- and postflood patch features: (a) *Direction* change map and (b) histogram with Rosin's thresholding corresponding to *direction* change; (c) *magnitude* change map and (d) histogram with Rosin's thresholding corresponding to *magnitude* change.

TABLE IV
HARVEY: QUANTITATIVE RESULTS OF THE PATCH-WISE FLOOD MAPS AT DIFFERENT PROCESSING STAGES BASED ON THE CHANGE OF *DIRECTION* AND *MAGNITUDE* (DIR: *DIRECTION*; MAG: *MAGNITUDE*; INT: *INITIAL*; IMD: *INTERMEDIATE*; FNL: *FINAL*). THE BEST RESULTS FOR EACH EVALUATION METRIC ARE MARKED IN BOLD FACE.

Type	Stage	Pr	Re	$F1$	IoU	OA
DIR	(a) INT	0.8939	0.9175	0.9055	0.8274	0.9659
	(b) IMD	0.9205	0.9133	0.9169	0.8465	0.9705
	(c) FNL	0.9428	0.9130	0.9276	0.8651	0.9747
MAG	(a) INT	0.8579	0.9313	0.8931	0.8068	0.9603
	(b) IME	0.8874	0.9278	0.9071	0.8301	0.9662
	(c) FNL	0.9107	0.9276	0.9190	0.8502	0.9709

TABLE V
EVALUATION USING DIFFERENT THRESHOLDING METHODS ON THE *DIRECTION* CHANGE MAP. ALL BOLD FACE VALUES ARE THE HIGHEST VALUES.

Thresholding	Pr	Re	$F1$	IoU	OA
Rosin [46]	0.8939	0.9175	0.9055	0.8274	0.9659
Otsu [38]	0.9969	0.2897	0.4489	0.2894	0.8735
KI [35]	0.6154	0.9934	0.7599	0.6128	0.8884

Pr , Re , $F1$, IoU , and OA of the *initial* flood maps as listed in Table IV, where the best results were marked in bold.

We also tested binarization of the initial *direction* change map [see Fig. 7(a)] using the thresholding methods proposed by Otsu [38] and Kittler and Illingworth (KI) [35]. Table V summarizes the results associated with each thresholding method. Rosin's method outperformed Otsu's and KI's methods in terms of $F1$, IoU , and OA . Otsu's method produced high Pr with very low Re , whereas KI's method performed conversely with low Pr but high Re . Only Rosin's method was able to consistently perform well in all metrics.

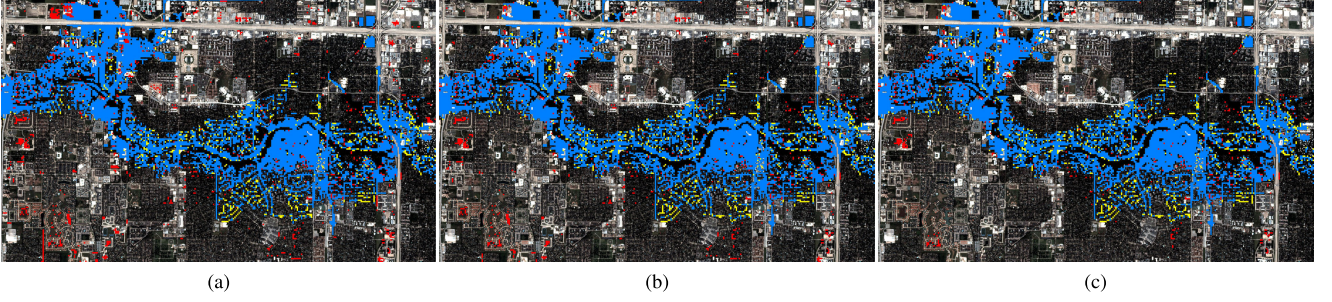


Fig. 8. Harvey: patch *direction* change based flood mapping with *TPs* in blue, *FPs* in red, and *FNs* in yellow. (a) *initial* flood map. (b) *Intermediate* flood map after spectral filtering. (c) *Final* flood map after spatial filtering.

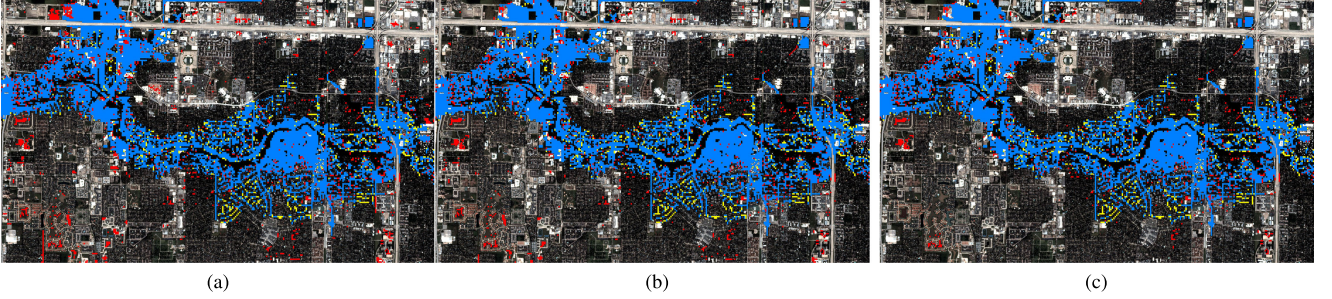


Fig. 9. Harvey: patch *magnitude* change based flood mapping with *TPs* in blue, *FPs* in red, and *FNs* in yellow. (a) *Initial* flood map. (b) *Intermediate* flood map after spectral filtering. (c) *Final* flood map after spatial filtering.

2) *Results of Spectral Filtering*: We clustered pixels located in patches detected in the above *initial* binary change maps. After removing building pixels with building footprints, some of initial changed patches were further classified as NF patches, in which no floodwater pixels were found based on the results of the two-class *K*-Means clustering. Figs. 8(b) and 9(b) showed the refined flood maps associated with *direction* and *magnitude* changes, where patches were marked in different colors indicating *TPs*, *FPs*, and *FNs*.

To check how spectral filtering improves the accuracy of patch-wise flood mapping, we also evaluated the associated *Pr*, *Re*, *F1*, *IoU*, and *OA* as listed in Table IV.

3) *Results of Spatial Filtering*: Results in Figs. 8(b) and 9(b) from Section V-A2 after spectral filtering show that there are still small isolated regions of false positive patches marked in red. Constrained by the spatial topology of floodwaters, FL patches should be close to each other within a FL region. We also assumed that a FL region should surpass some minimum area. In this experiment, we tested multiple options of both the maximum distance *d* for FL patch connectivity and the minimum area *a* for a FL region. Figs. 8(c) and 9(c) show the final patch-wise flood maps after spatial filtering with *d* = 5 and *a* = 20. Fig. 10 demonstrates the impact of the hyperparameter *a* and *d* on the *final* patch-wise flood map in terms of the *F1* score and *IoU* based on the *direction* change. The corresponding quantitative results are summarized in Table IV for both the *direction* and *magnitude* change maps at different processing stages.

B. Flood Event 2: Florence

1) *Results of Patch-CVA*: We trained a new patch encoder using the same model architecture [see Fig. 3(b)] with Florence

TABLE VI

FLORENCE: QUANTITATIVE RESULTS OF THE PATCH-WISE FLOOD MAPS AT DIFFERENT PROCESSING STAGES BASED ON THE CHANGE OF *DIRECTION* AND *MAGNITUDE* (DIR: DIRECTION; MAG: MAGNITUDE; INT: INITIAL; IMD: INTERMEDIATE; FNL: FINAL). THE BEST RESULTS FOR EACH EVALUATION METRIC ARE MARKED IN BOLD FACE.

Type	Stage	<i>Pr</i>	<i>Re</i>	<i>F1</i>	<i>IoU</i>	<i>OA</i>
DIR	(a) INT	0.7355	0.9293	0.8211	0.6965	0.9224
	(b) IMD	0.8249	0.9265	0.8728	0.7743	0.9483
	(c) FNL	0.8352	0.9251	0.8779	0.7823	0.9507
MAG	(a) INT	0.7360	0.9354	0.8238	0.7004	0.9233
	(b) IME	0.8252	0.9326	0.8756	0.7788	0.9492
	(c) FNL	0.8355	0.9311	0.8807	0.7868	0.9517

data, which enables better patch feature extraction for the new study area. Following the same processing chain for Harvey data, we obtained two patch-wise flood maps corresponding to the *direction* [Fig. 11(a)] and *magnitude* [see Fig. 11(c)] change maps, respectively. To obtain the *initial* flood maps, thresholds were selected based on the histograms of the *direction* [Fig. 11(b)] and the *magnitude* [see Fig. 11(d)] change maps using Rosin's method [45], [46]. Figs. 12(a) and 13(a) showed the *initial* flood maps. Corresponding quantitative evaluation is listed in Table VI.

2) *Results of Spectral Filtering*: The two-class unsupervised *K*-Means clustering was performed with all pixels from the potential FL patches detected in Figs. 12(a) and 13(a), respectively. Figs. 12(b) and 13(b) show the refined patch-wise flood maps after removing nonflood-related patches, which were originally classified as FL in the *initial* flood maps [see Figs. 12(a) and 13(a)] by Patch-CVA. Table VI summarizes the evaluation results associated with the refined *intermediate* flood maps.

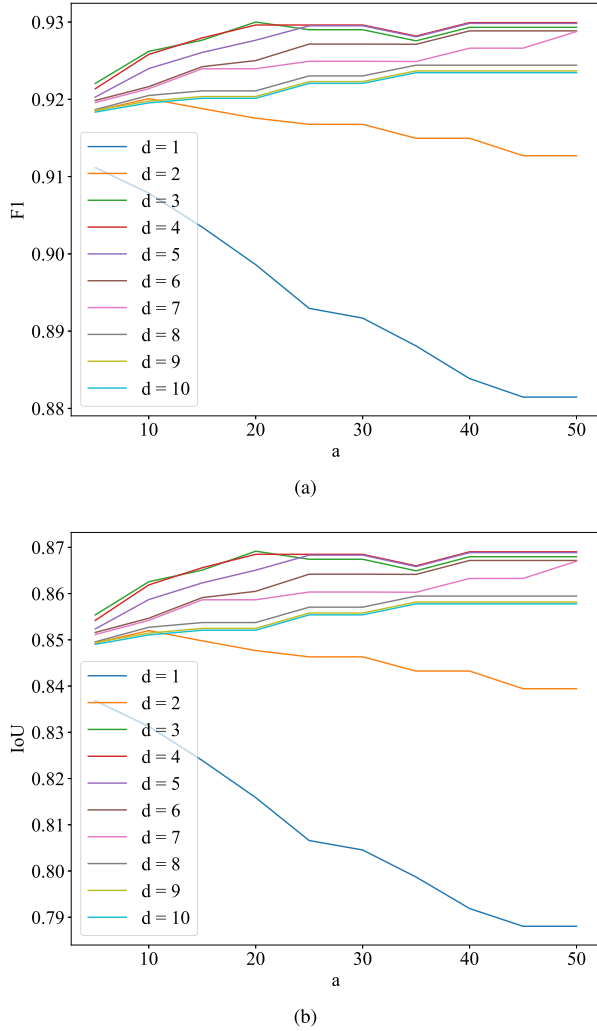


Fig. 10. Harvey: impact of d and a on the (a) $F1$ and (b) IoU of final flood maps.

3) *Results of Spatial Filtering*: After removing NF changed patches with spectral filtering, spatial filtering based on spatial topology of floodwaters was applied to further remove false positive FL patches in the *intermediate* flood maps. We used the same hyperparameters as those in Harvey experiment. Figs. 12(c) and 13(c) demonstrated the results of the *final* patch-wise flood maps after spatial filtering.

Associated with Figs. 12(c) and 13(c) were the evaluation of Pr , Re , $F1$, IoU , and OA summarized in Table VI for both the *direction* [see Fig. 12(c)] and *magnitude* [see Fig. 13(c)] change maps.

We also experimented with different (d, a) to show their impact on the accuracy of the final patch-wise flood extent maps. Fig. 14 plotted the change of $F1$ and IoU with respect to various (d, a) .

VI. DISCUSSION

A. Patch Feature Learning

The proposed method starts with patch feature extraction by a patch encoder trained with the input data without manual

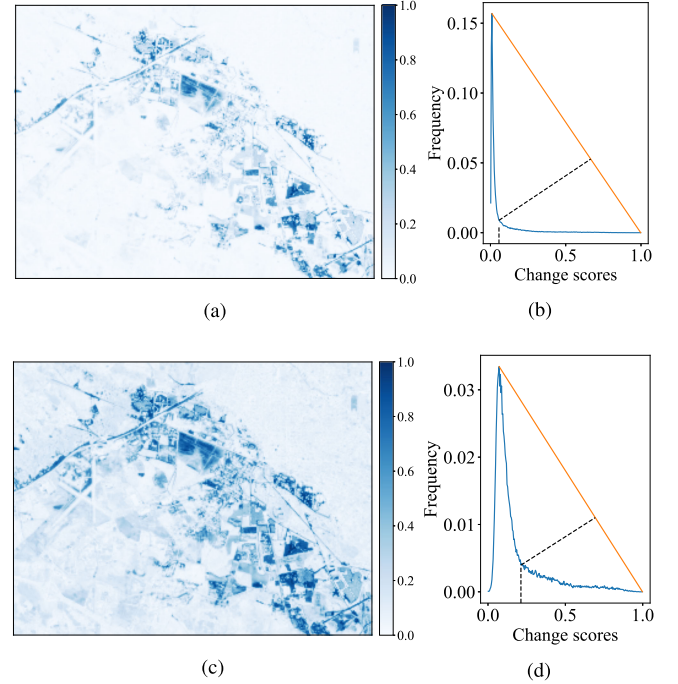


Fig. 11. Florence change maps based on encoded pre- and postflood patch features. (a) *direction* change map and (b) histogram with Rosin's thresholding corresponding to *direction* change; (c) *magnitude* change map and (d) histogram with Rosin's thresholding corresponding to *magnitude* change.

TABLE VII
HARVEY: EVALUATION OF THE INITIAL MAGNITUDE CHANGE MAPS BY ESTIMATING PATCH CHANGES WITH DIFFERENT METHODS: (1) AGGREGATION OF RAW PIXEL FEATURE CHANGE, (2) COMPUTING CHANGE OF PATCH FEATURES EXTRACTED BY PCA, AND (3) COMPUTING CHANGE OF PATCH FEATURES LEARNED BY THE AUTOENCODER EMPLOYED IN THE PROPOSED FRAMEWORK. THE BEST RESULTS FOR EACH EVALUATION METRIC ARE MARKED IN BOLD FACE.

Change Map	Pr	Re	$F1$	IoU	OA
Raw pixel	0.8524	0.7717	0.8100	0.6807	0.9356
PCA	0.3745	0.1149	0.1759	0.0964	0.8084
Autoencoder	0.8579	0.9313	0.8931	0.8068	0.9603

feature engineering. Such a self-supervised learning framework enables automatic learning of patch features and generalized well to both study areas because the model was trained with local data. This feature is critical for near realtime automated flood mapping since traditional hand-crafted features often fail to generalize well to new datasets associated with heterogeneous image background, which is common in urban areas. Without patch feature extraction, patch-wise flood mapping cannot be implemented through direct pixel-wise processing.

To demonstrate the power of patch features learned by the proposed method, we tested 1) the raw pixel feature and 2) the patch feature extracted by principal component analysis (PCA) for patch-wise flood mapping. First, we computed the pixel-wise *magnitude* change between the pre- and postflood images (I^{pre} , I^{post}) and obtained the patch-wise *magnitude* change map via aggregation of pixel-wise *magnitude* change across each patch. Second, PCA was used to extract features from each pair of pre- and postflood patches and we kept the first two principal components for further patch-wise flood mapping. Table VII

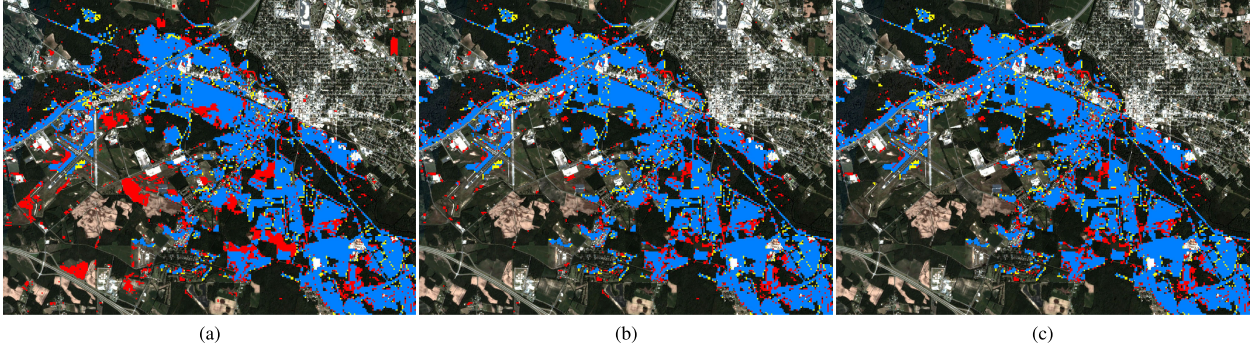


Fig. 12. Florence: patch *direction* change based flood mapping with *TPs* in blue, *FPs* in red, and *FNs* in yellow. (a) *Initial* flood map. (b) *Intermediate* flood map after spectral filtering. (c) *Final* flood map after spatial filtering.

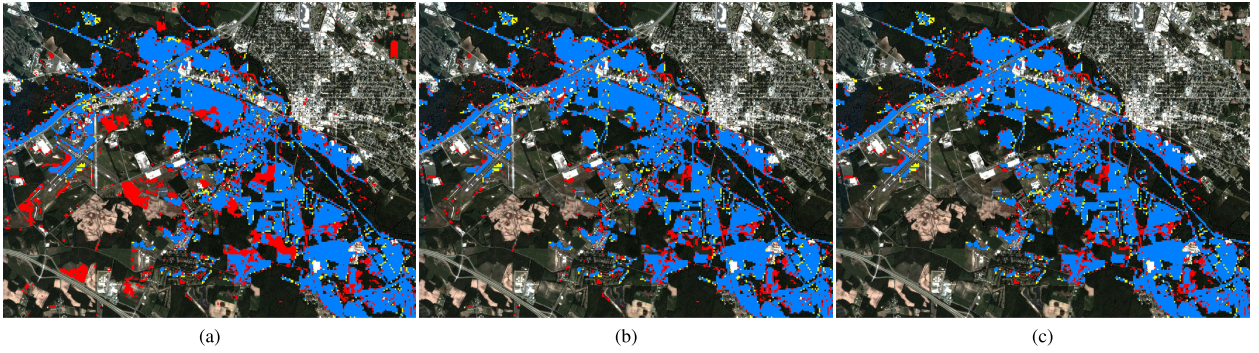


Fig. 13. Florence: patch *magnitude* change based flood mapping with *TPs* in blue, *FPs* in red, and *FNs* in yellow. (a) *Initial* flood map. (b) *Intermediate* flood map after spectral filtering. (c) *Final* flood map after spatial filtering.

summarizes the results of patch-wise *initial* flood mapping for Hurricane Harvey based on the raw pixel feature and the PCA extracted feature without further spectral and spatial filtering, compared to the results generated by the proposed method (i.e., Patch-CVA with *magnitude* change). Rosin’s method was used for thresholding both change maps.

The results show that flood mapping based on the raw pixel feature and the PCA extracted feature fails to capture some of the weak changes resulting from flooding as demonstrated by a lower *Re* than that corresponding to patch feature based processing. It is worth noting that the autoencoder employed in the proposed framework significantly outperforms PCA in terms of patch feature learning for patch-wise flood mapping.

Similarly, we performed flood mapping based on raw pixel feature change for Hurricane Florence. Fig. 15(a) shows that direct pixel-wise change estimation leads to a large number of false positives due to noninformative changes such as varying illumination, inaccurate radiometric correction, and poor image registration. The corresponding histogram [see Fig. 15(b)] also exhibits completely different patterns than those associated with patch-wise change maps, leading to the failure of Rosin’s, Otsu’s, or KI’s thresholding techniques.

B. Comparison of Direction and Magnitude Change

As shown in the experimental results for both Harvey and Florence floods, there is no significant difference between *direction* and *magnitude* change-based patch-wise flood mapping, as

demonstrated by 1) quantitative evaluation of *Pr*, *Re*, *F1*, *IoU*, and *OA* in Table IV for Harvey and Table VI for Florence, and 2) by qualitative visual inspection of the *final* patch-wise flood maps in Fig. 9(c) for Harvey and Fig. 13(c) for Florence.

However, we observed in Figs. 7 and 11 that the *direction* change map showed more abrupt change around the selected threshold while the *magnitude* change map changed more gradually, which was also reflected by the respective histograms. The abrupt change around the threshold indicated that the patch *direction* change-based processing encouraged separation of FL from NF patches with higher contrast compared to *magnitude*-based change estimation.

Furthermore, as *direction* change is more invariant to changes in illumination compared to *magnitude* change, *direction*-based change detection has the potential to suppress minor noninformative changes due to varying illumination, as demonstrated in Figs. 7(a) and 11(a), in which salient changes were highlighted, whereas subtle changes were suppressed.

C. Impact of Spectral and Spatial Filtering

To demonstrate how spectral and spatial filtering have refined the patch-wise flood maps, Figs. 8 and 12 present the patch *direction* change-based flood maps sequentially for visual inspection of Harvey and Florence floods with corresponding quantitative evaluation illustrated in Fig. 16, where *Initial*, *Intermediate*, *Final* correspond to the results of the *initial* flood maps based on Patch-CVA [see Fig. 8(a)], the *intermediate* flood maps after

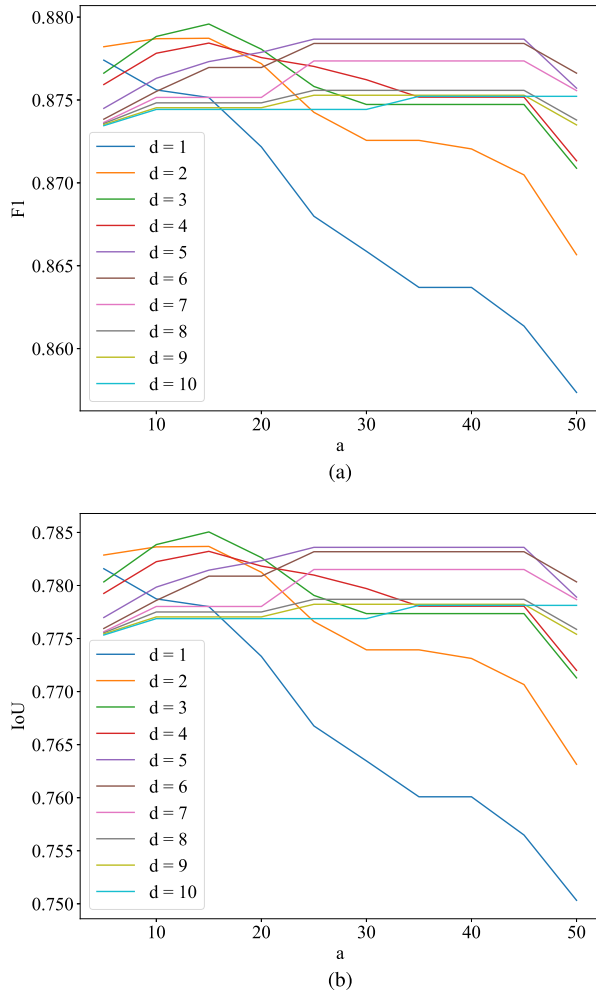


Fig. 14. Florence: impact d and a on the (a) $F1$ and (b) IoU of final flood extent maps.

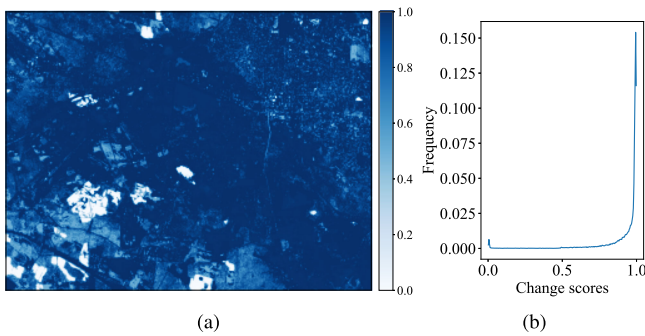


Fig. 15. Florence: (a) *initial* change map based on aggregation of pixel-wise magnitude change across each patch and (b) the corresponding histogram.

spectral filtering [see Fig. 8(b)], and the *final* flood maps after spatial filtering [see Fig. 8(c)].

It is obvious that the sequential processing through spectral and spatial filtering improve accuracy in terms of $F1$ and IoU based on the *initial* patch-wise flood maps. Spectral filtering helped remove nonfloodwater related changes. With a two-class K -Means clustering of all pixels from the *initial* changed

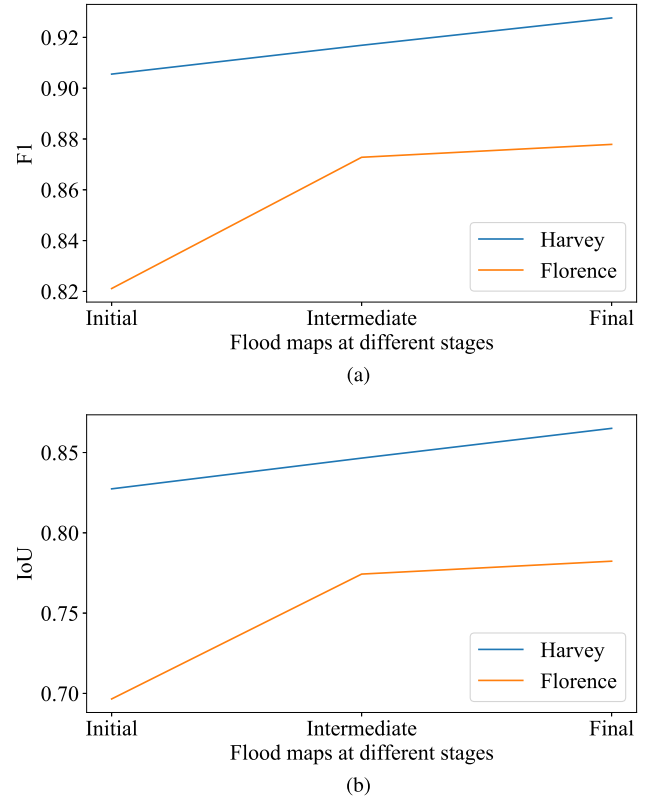


Fig. 16. Impact of spectral and spatial filtering on the (a) $F1$ and (b) IoU of flood extent maps.

patches, most NF patches were removed. Spatial filtering considered noise and incorrect classification of a few false positive patches from previous steps. In particular, a few small isolated changed patches remained after Patch-CVA and spectral filtering. Those small regions were further removed by leveraging the spatial topology of the FL regions. In this study, FL patches were constrained to be close to each other with distance less than 5 patches in a major FL region with area of at least 20 patches. *Final* flood maps for both Harvey and Florence floods highlight the robust performance of the proposed method.

Regarding the spatial filtering hyperparameters, (d, a) , Fig. 10 shows the impact of different (d, a) on the accuracy of the *final* flood maps in terms of $F1$ and IoU based on the *direction* change. We noted that the performance dropped significantly if we constrained the maximum distance of FL patch connectivity strictly with $d = 1$. That is, FL patches were connected only if they were strictly adjacent to their neighborhoods. There are several contributing explanations. Some FL patches under tree canopy are labeled as NF patches [12], [21]; such conditions are common in urban areas. As a result, ground truth flood maps may contain FL patches that are disconnected with the neighboring FL regions. In addition, false negative patches could result from thresholding of the *initial* flood maps, spectral filtering for the intermediate flood maps, and image noise. As such, we assumed that any two FL patches were connected if they were within a maximum distance of d patches. The other hyperparameter a also influences the performance of spatial filtering for final flood

TABLE VIII

EVALUATION OF $F1$ AND IoU FOR FLOOD MAPS AFTER THE SEQUENTIAL (SPECTRAL FILTERING + SPATIAL FILTERING) AND AFTER ONLY SPATIAL FILTERING BASED ON THE *INITIAL DIRECTION* CHANGE-BASED FLOOD MAPS PRODUCED BY PATCH-CVA

Flood	Metric	Initial	Intermediate	Final
Harvey	$F1$	0.9055	0.9169 (spectral)	0.9276 (spatial)
			0.9205 (only spatial)	
	IoU	0.8274	0.8465 (spectral)	0.8651 (spatial)
			0.8528 (only spatial)	
Florence	$F1$	0.8211	0.8728 (spectral)	0.8779 (spatial)
			0.8267 (only spatial)	
	IoU	0.6965	0.7743 (spectral)	0.7823 (spatial)
			0.7047 (only spatial)	

mapping. We observed that, in ground truth flood maps, some flooding occurred in localized FL areas relatively far from other FL patches. Though the localized FL regions are isolated relative to the major FL regions, however, they should have some area greater than a predefined empirical value a . As some small FL patches are attributable noise or other errors (e.g., small area of the wet road surface), and not removed before spatial filtering, such an assumption contributed to further refinement of the *final* flood maps.

In this study, spectral filtering and spatial filtering are two interlocking sequential modules for producing the *final* flood maps. Hence, spectral filtering is a prerequisite for spatial filtering to ensure the robustness and effectiveness of spatial filtering. To further prove the effectiveness of spatial filtering, we conducted spatial filtering alone based on the *initial direction* change-based flood maps without spectral filtering. Table VIII shows that the $F1$ and IoU of flood maps after only spatial filtering are higher than those of the *initial* flood maps, demonstrating the effectiveness of spatial filtering without spectral filtering. However, comparing the results with only spatial filtering to those with the sequential (spectral filtering + spatial filtering), we observe that spectral filtering plays a key role as the prerequisite for spatial filtering. Spatial filtering is used based on the spatial topology of floods (i.e., small isolated and changed areas are considered as NF). Without spectral filtering, there remain some large NF but changed areas connected with major FL areas, resulting in less effective spatial filtering to filter out such false positive changes. With spectral filtering performed before spatial filtering, most of the large NF but changed areas are removed. Then the remaining NF but changed areas are smaller and isolated from major FL areas such that spatial filtering performs better for producing the *final* flood maps. Therefore, it is important to perform spectral filtering and spatial filtering sequentially to ensure the accuracy of the *final* flood maps.

D. Model Performance Across Different Study Areas

Fig. 16 shows that the model performs well for urban flood mapping, with $F1$ above 0.87 and IoU above 0.78, highlighting

the model's capability in near realtime processing for upcoming unseen floods.

The evaluations of $F1$ and IoU associated with the Florence flood were lower than those with the Harvey flood. These differences were due mainly to different degrees of NF changes between the bitemporal data corresponding to the Harvey and Florence floods over different study areas, respectively. As demonstrated in Figs. 12(a) and 13(a) for Florence, the *initial* flood maps exhibit higher rates of FPs (i.e., NF changes marked in red) than those corresponding to the *initial* flood maps [Figs. 8(a) and 9(a)] for Harvey.

With spectral and spatial filtering, the majority of the FPs were removed as demonstrated by the increasing Pr in Tables IV and VI. The Florence *final* flood maps contain a higher rate of FPs compared to the Harvey *final* flood maps. Potential causes include: 1) many FPs in the Florence *final* flood maps consist of pixels with spectra similar to those of floodwaters such as patches that are wet but not FL, leading to misclassification in spectral filtering; and 2) quite a few FPs in the Florence *final* flood maps are mixed and connected with TPs , which were not removed by spatial filtering. As a result, there remain higher rates of FPs in both *direction* and *magnitude* change-based *final* flood maps in the Florence experiment.

As the main goal of this study is to map urban floods in near realtime, time consumption on training and testing the models plays an important role. Deep learning based models are often time-consuming due to the training of deep neural networks with a large volume of data. Table III shows that it took 40.97 mins and 46.76 min for training the autoencoders corresponding to the Harvey and Florence floods, respectively. Additionally, it took less than 1 min to encode all pre- and postflood image patches into patch features. Such features were then used for further Patch-CVA, spectral filtering, and spatial filtering, which also took negligible time to produce the *final* flood maps. Hence, the total time consumption of the proposed self-supervised learning framework was less than 1 h for mapping floods at two different urban areas with bitemporal satellite MS images, enabling near realtime processing for emergency response.

E. Comparison of Different Thresholding Methods

Automatic thresholding of the *initial direction* or *magnitude* change maps is critical for identifying *initial* FL patches for further spectral and spatial filtering. As demonstrated by Figs. 7 and 11, the histograms exhibit unimodal distributions for both *direction* and *magnitude* change maps associated with Harvey and Florence data. Table V shows that Rosin's *corner*-based thresholding method achieved superior performance without specific assumption of the foreground or background sample distribution. Note that the foreground samples consisted of changed postflood patches, while the background samples were unchanged ones.

It is worth noting that Otsu's thresholding method tends to set a larger threshold than the one by Rosin's method, as demonstrated by the low recall Re but high precision Pr of the *initial* flood maps. One assumption of Otsu's method is that the histogram of the image is a bimodal distribution. Additionally, Otsu's method

achieves good performance when the valley between the two peaks of the histogram is deep and sharp [51]. As shown in Figs. 7 and 11, however, the histograms of both *direction* and *magnitude* change maps for Harvey and Florence data exhibit unimodal distributions with only one major peak.

In contrast, KI's minimum error thresholding method picked up a smaller threshold compared to Rosin's threshold, leading to a low precision *Pr* but a high recall *Re*. The minimum error thresholding method assumes that foreground and background samples are both normally distributed with distinct mean and standard deviations [35]. This assumption was not true in either *direction* or *magnitude* change maps. Neither FL nor NF patches exhibit a normal distribution in the histogram of either *direction* or *magnitude* change maps.

VII. CONCLUSION

We propose a fully automated patch-wise urban flood extent mapping method via a self-supervised learning framework. The before- and after-flood patch features were automatically learned through a self-supervised autoencoder. Patch-CVA was performed based on patch features learned from the pretrained encoder of the autoencoder model, which generated the patch-wise *direction* and *magnitude* change maps. Potential FL patches were extracted through robust binarization of the corresponding change maps, where the binarization thresholds were picked at the change intensity corresponding to the corners of the unimodal change map histograms. Since noise and other errors can cause false alarms in flood mapping, spectral, and spatial filtering were performed on the *initial* patch-wise flood maps by leveraging the spectral signatures and spatial topology of floodwaters. Our results show that the proposed method achieves good performance for both Harvey and Florence floods in terms of *F1*, *IoU*, and *OA*. For example, we obtained the *final* flood maps with 0.9276 *F1* and 0.8651 *IoU* for Harvey flood, and 0.8779 *F1* and 0.7823 *IoU* for Florence flood based on respective *direction* change maps. The self-supervised learning framework enables patch feature learning without a large number of human-annotated training data. The pretrained patch encoder extracts informative features from both pre- and postflood patches, where no hand-crafted feature engineering is required. The majority of the desired changes were detected through Patch-CVA based on the corresponding patch features as demonstrated by the *initial* flood maps. Spectral and spatial filtering further boosts the performance by removing nonflood related changes and noise.

Regarding the future work on near realtime flood mapping in practice, optical MS imagery may not be available due to the impact of clouds. It is likely worth testing the proposed method using SAR data or a fusion of SAR and optical data, to take advantage of the nighttime and all-weather data acquisition capability of SAR. Moreover, when only limited number of human annotated training data are available, we would like to try semisupervised learning, transfer learning, and active learning for further improvement of flood mapping accuracy over heterogeneous urban areas.

ACKNOWLEDGMENT

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, Microsoft, and UW-Madison.

REFERENCES

- [1] USGS, "Flood Inundation Mapping (FIM) Program," 2019. [Online]. Available: https://www.usgs.gov/mission-areas/water-resources/science/flood-inundation-mapping-fim-program?qt-science_center_objects=0#qt-science_center_objects
- [2] P. J. Pilon, "Guidelines for reducing flood losses," United Nations Int. Strategy Disaster Reduction, Tech. Rep., 2002. [Online]. Available: <https://www.undrr.org/publication/guidelines-reducing-flood-losses>
- [3] United Nations, "Sustainable Development Goals Report," 2018. [Online]. Available: <https://www.un.org/sustainabledevelopment/cities>
- [4] L. Li, Y. Chen, X. Yu, R. Liu, and C. Huang, "Sub-pixel flood inundation mapping from multispectral remotely sensed images based on discrete particle swarm optimization," *ISPRS J. Photogrammetry Remote Sens.*, vol. 101, pp. 10–21, Mar. 2015.
- [5] R. Malinowski, G. Groom, W. Schwanghart, and G. Heckrath, "Detection and delineation of localized flooding from WorldView-2 multispectral data," *Remote Sens.*, vol. 7, no. 11, pp. 14 853–14 875, 2015.
- [6] P. Wang, G. Zhang, and H. Leung, "Improving super-resolution flood inundation mapping for multispectral remote sensing image by supplying more spectral information," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 771–775, May 2019.
- [7] M. Wieland and S. Martinis, "A modular processing chain for automated flood monitoring from multi-spectral satellite data," *Remote Sens.*, vol. 11, no. 19, Oct. 2019, Art. no. 2330.
- [8] L. Zhang *et al.*, "Fast real-time causal line-wise progressive hyperspectral anomaly detection via cholesky decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4614–4629, Oct. 2017.
- [9] B. Peng, L. Zhang, T. Wu, and H. Zhang, "Fast real-time target detection via target-oriented band selection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2016, pp. 5868–5871.
- [10] B. Du, Q. Wei, and R. Liu, "An improved quantum-behaved particle swarm optimization for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6003–6017, Aug. 2019.
- [11] L. Tong, B. Du, R. Liu, and L. Zhang, "An improved multiobjective discrete particle swarm optimization for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7872–7882, Oct. 2019.
- [12] B. Peng, Z. Meng, Q. Huang, and C. Wang, "Patch similarity convolutional neural network for urban flood extent mapping using bi-temporal satellite multispectral imagery," *Remote Sens.*, vol. 11, no. 21, Oct. 2019, Art. no. 2492.
- [13] B. Peng, X. Liu, Z. Meng, and Q. Huang, "Urban flood mapping with residual patch similarity learning," in *Proc. 3rd ACM SIGSPATIAL Int. Workshop AI Geographic Knowl. Discov. - GeoAI*. New York, NY, USA: ACM Press, 2019, pp. 40–47.
- [14] Planet Team, "Planet application program interface: In *Space for Life on Earth*. San Francisco, CA," 2018.
- [15] DigitalGlobe, "Open Data Program," [Online]. Available: <https://www.digitalglobe.com/ecosystem/open-data>
- [16] M. Xie, Z. Jiang, and A. M. Sainju, "Geographical hidden markov tree for flood extent mapping," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 2545–2554.
- [17] S. Skakun, "A neural network approach to flood mapping using satellite imagery," *Comput. Informat.*, vol. 29, no. 6, pp. 1013–1024, 2012.
- [18] P. Insom *et al.*, "A support vector machine-based particle filter method for improved flooding classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1943–1947, Sep. 2015.
- [19] Q. Feng, J. Liu, and J. Gong, "Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A Case of Yuyao, China," *Water*, vol. 7, no. 12, pp. 1437–1455, Mar. 2015.
- [20] A. Gebrehiwot, L. Hashemi-Beni, G. Thompson, P. Kordjamshidi, and T. Langan, "Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data," *Sensors*, vol. 19, no. 7, Mar. 2019, Art. no. 1486.

- [21] 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. Photogrammetry Remote Sens.*, vol. 152, pp. 178–191, Jun. 2019.
- [22] T. Liu, L. Yang, and D. D. Lunga, "Towards misregistration-tolerant change detection using deep learning techniques with object-based image analysis," in *Proc. 27th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. - SIGSPATIAL '19*. New York, NY, USA: ACM Press, 2019, pp. 420–423.
- [23] L. Bruzzone and R. Cossu, "An adaptive approach to reducing registration noise effects in unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2455–2465, Nov. 2003.
- [24] Y. Byun, Y. Han, and T. Chae, "Image fusion-based change detection for flood extent extraction using bi-temporal very high-resolution satellite images," *Remote Sens.*, vol. 7, no. 8, pp. 10 347–10 363, Aug. 2015.
- [25] FEMA, "Federal Emergency Management Agency Flood Mapping Products," [Online]. Available: <https://www.fema.gov/flood-mapping-products>
- [26] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: [10.1109/TPAMI.2020.2992393](https://doi.org/10.1109/TPAMI.2020.2992393).
- [27] A. Kolesnikov, X. Zhai, and L. Beyer, "Revisiting self-supervised visual representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2019, pp. 1920–1929.
- [28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [29] H. Song, Y. Kim, and Y. Kim, "A patch-based light convolutional neural network for land-cover mapping using Landsat-8 images," *Remote Sensing*, vol. 11, no. 2, Art. no. 114, 2019.
- [30] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Netw.*, vol. 95, pp. 19–28, 2017.
- [31] A. S. Dhakal, T. Amada, M. Aniya, and R. R. Sharma, "Detection of areas associated with flood and erosion caused by a heavy rainfall using multitemporal Landsat TM data," *Photogrammetric Eng. Remote Sens.*, vol. 68, no. 3, pp. 233–239, 2002.
- [32] S. Schlaffer, P. Matgen, M. Hollaus, and W. Wagner, "Flood detection from multi-temporal SAR data using harmonic analysis and change detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 38, pp. 15–24, Jun. 2015.
- [33] L. Giustarini *et al.*, "A change detection approach to flood mapping in urban areas using TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2417–2430, Apr. 2013.
- [34] N. Longbotham *et al.*, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009-2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.
- [35] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, 1986.
- [36] G. Moser and S. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, Oct. 2006.
- [37] F. Greifeneder, W. Wagner, D. Sabel, and V. Naeimi, "Suitability of SAR imagery for automatic flood mapping in the lower mekong basin," *Int. J. Remote Sens.*, vol. 35, no. 8, pp. 2857–2874, Apr. 2014.
- [38] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [39] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
- [40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, May 2015.
- [41] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with landsat," in *LARS Symposia*, 1980, p. 385.
- [42] E. F. Lambin and A. H. Strahlers, "Change-vector analysis in multi-temporal space: A tool to detect and categorize land-cover change processes using high temporal-resolution satellite data," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 231–244, May 1994.
- [43] O. A. Carvalho Júnior, R. F. Guimarães, A. R. Gillespie, N. C. Silva, and R. A. T. Gomes, "A new approach to change vector analysis using distance and similarity measures," *Remote Sens.*, vol. 3, no. 11, pp. 2473–2493, Nov. 2011.
- [44] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.
- [45] P. L. Rosin, "Thresholding for change detection," *Comput. Vis. Image Understanding*, vol. 86, no. 2, pp. 79–95, May 2002.
- [46] P. L. Rosin, "Unimodal thresholding," *Pattern Recognit.*, vol. 34, no. 11, pp. 2083–2096, 2001.
- [47] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Unsupervised rapid flood mapping using Sentinel-1 GRD SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3290–3299, Jun. 2018.
- [48] Microsoft Bing Maps Team, "Microsoft Building Footprints," 2018. [Online]. Available: <https://www.microsoft.com/en-us/maps/building-footprints>
- [49] M. Levandowsky and D. Winter, "Distance between sets," *Nature*, vol. 234, no. 5323, pp. 34–35, Nov. 1971.
- [50] P. Adam *et al.*, "Automatic differentiation in Pytorch," in *Proc. Neural Inf. Process. Syst.*, 2017.
- [51] J. Kittler and J. Illingworth, "On threshold selection using clustering criteria," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 5, pp. 652–655, Sep/Oct. 1985.



Bo Peng (Graduate Student Member) received the B.E. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, the M.S. degree in cartography and geographic information system from the University of Chinese Academy of Sciences, Beijing, China, in 2017, and the M.S. degree in electrical engineering from the University of Wisconsin-Madison, Madison, WI, USA, in 2019. He is currently working toward the Ph.D. degree in geography (remote sensing track) with a doctoral minor in electrical engineering (machine learning track)

from the University of Wisconsin-Madison.

His research interests include span GeoAI, earth observation, computer vision, machine learning, and disaster resilience.



Qunying Huang received the B.S. degree in survey and mapping engineering from Central South University, Changsha, China, in 2004, the M.S. degree in cartography and geographical information science from Peking University, Beijing, China, in 2007, and the Ph.D. degree in earth systems and geoinformation sciences from George Mason University, Fairfax, VA, USA, in 2011.

She is currently an Associate Professor with the Department of Geography, University of Wisconsin-Madison, Madison, WI, USA. Her fields of expertise include spatial computing, spatial data mining, and spatial data analytics. Her research bridges the gap between Computer and Information Science (CIScience) and GIScience by generating new computational algorithms and methods to make sense of complex big spatial data sets obtained from both the physical (e.g., remote sensing) and social (e.g., social media) sensing networks. The problem domains of her research are related to natural hazards and human mobility.



Jamp Vongkusolkiet received the B.S. degree in environmental sciences, in 2019, from University of Wisconsin-Madison, Madison, WI, USA, where she is currently working toward the M.S. degree in geographic information science.

She is interested in applying geospatial analysis and spatial modeling to environmental issues and natural disasters. She is also interested in using image processing and analysis techniques on satellite imagery for environmental analyses.



Song Gao received the B.S. degree (Hons.) from the School of Geography, Beijing Normal University, Beijing, China, in 2009, the M.S. degree in cartography and GIS from Peking University, Beijing, China, in 2012, and the Ph.D. degree in geography from the University of California-Santa Barbara, Santa Barbara, CA, USA, in 2017.

He is currently an Assistant Professor in GIScience with the University of Wisconsin-Madison, Madison, WI, USA. His main research interests include place-based GIS, GeoAI, geospatial big data, and social

sensing.



Daniel B. Wright received the bachelor's and master's degrees in civil and environmental engineering from the University of Michigan, Ann Arbor, MI, USA, in 2005 and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 2013, where he studied urban rainfall and flooding.

He worked at the World Bank from 2013 to 2014, focusing on flood and landslide risk reduction in Latin America and the Caribbean before becoming a NASA Postdoctoral Program fellow at Goddard Space Flight Center from 2014 to 2016. He joined the Civil and

Environmental Engineering Department, UW-Madison as an Assistant Professor in 2016. His research and teaching focuses on floods and how they are influenced by meteorology, urbanization, and climate change.

Dr. Wright is a member of NASA's Precipitation Measurement Mission Science Team and received a National Science Foundation CAREER Award in Hydrologic Sciences, in 2018. He is a Founder and Co-Chair of the Infrastructure Working Group within the Wisconsin Initiative on Climate Change Impacts, as well as a member of its Science Advisory Board.



Zheng N. Fang received the Ph.D. degree in civil and environmental engineering from Rice University, Houston, TX, USA, in 2008.

He is currently an Associate Professor with the Civil Engineering Department, University of Texas at Arlington, Arlington, TX, USA. As a Texas licensed PE, he has been actively working on surface water and groundwater problems for over 17 years including floodplain studies, hydrologic/hydraulic modeling, water treatment, hydrodynamic simulation, storm water management modeling, and water quality assessment for a number of watersheds and areas in Texas, Florida, Georgia, Connecticut, California, and Louisiana. He has been conducting quality research in the areas of inland flood prediction and hurricane-induced storm surge, urban hydraulic prediction tool-floodplain map library (FPML), advanced radar-based flood warning system, dynamic moving storm (DMS) generator and stochastic storm transposition (SST), climate forecast-aided drought decision support, integrated sensing and prediction of urban water, evaluation of urbanization and infiltration, and unmanned aerial vehicles (UAVs) enabled data collection system for hydrology and hydraulics.

Dr. Fang is a contributing member to the state emergency operations during Hurricane Harvey (2017) for the Texas Division of Emergency Management (TDEM), the U.S. Army Corps of Engineers (USACE), the Texas Medical Center (TMC), Rice University, and the U.T. MD Anderson Cancer Center.



Yi Qiang received the B.Sc. degree in geographic information systems from Beijing Normal University, Beijing, China, in 2006, and the M.Sc. degree in geographical information science from University of Edinburgh, Edinburgh, U.K., in 2007, and the Ph.D. degree in geography from Ghent University, Ghent, Belgium, in 2012.

He is currently an Assistant Professor with the School of Geosciences, University of South Florida, Tampa, FL, USA. His research focuses on GIScience, spatial data science, and geovisualization. He leads

multiple research projects in space-time modeling, GeoAI, and data-driven methods for disaster resilience assessment.