

Patch Similarity Learning for Urban Flood Extent Mapping

Bo Peng ^{1,2}, Zonglin Meng ³, Qunying Huang ^{1,*,†} and Caixia Wang ⁴

¹ Department of Geography, University of Wisconsin - Madison;

² Department of Electrical and Computer Engineering, University of Wisconsin - Madison;

³ Department of Computer Sciences, University of Wisconsin - Madison;

⁴ Department of Geomatics; University of Alaska Anchorage;

* Correspondence: qhuang46@wisc.edu;

† Current address: 550 N Park Street, Madison, WI 53706, USA

Version December 29, 2020 submitted to Remote Sens.

Abstract: Urban flooding is a major natural disaster that poses serious threat to the urban environment. It is highly demanded that the flood extent can be mapped in near real-time for disaster rescue and relief missions, reconstruction efforts, and financial loss evaluation. Many efforts have been taken to identify the flooding zones with remote sensing data and image processing techniques. However, urban flood mapping at high spatial resolution remains a major challenge due to three main reasons. (1) The very high resolution (VHR) optical remote sensing imagery over the urban area usually has heterogeneous background involving various ground objects (e.g., vehicles, buildings, roads, and trees). Therefore, traditional classification algorithms often fail to capture the underlying spatial correlation between neighboring pixels within the flood hazard area; (2) Traditional flood mapping methods with handcrafted features as input cannot fully leverage massive available data, which requires robust and scalable algorithms; and (3) Due to inconsistent weather conditions at different time of data acquisition, pixels of the same objects in VHR optical imagery could have very different pixel values, leading to the poor generalization capability of classical flood mapping methods. To address this challenge, this paper proposed a patch similarity convolutional neural network (PSNet) to map urban flood hazard zones using bi-temporal high resolution (3m) pre- and post-flooding multispectral surface reflectance satellite imagery. As spectral reflectance helps alleviate the impact of varying illuminations due to different data acquisition conditions, it further improves the generalization ability of PSNet. Experiments on the high resolution imagery before and after the urban flooding events (i.e., the 2017 Hurricane Harvey and the 2018 Hurricane Florence) showed that the developed PSNet can produce urban flood maps with consistently high precision, recall, F1 score and overall accuracy compared with baseline classification models including support vector machine, decision tree, random forest, and AdaBoost, which were often weak in either precision or recall. The research sheds light on bi-temporal image fusion for high precision image change detection, which in turn can be used for monitoring damages caused by other types of natural hazards (e.g., wildfires and earthquakes).

Keywords: Flood mapping; patch similarity; deep learning; flood extent estimation

1. Introduction

Natural hazard poses a major threat to people's life and living environment, especially for areas with high population density (e.g., urban regions). Flooding events are one of the most frequent natural disasters that have direct damage over man-made ground infrastructures, including roads and buildings [1–3]. The estimated global financial losses, with projected socio-economic change alone,

will increase close to \$52 billion by 2050 [4]. To improve the safety, resilience, and sustainability of the cities and human settlements, the United Nations (UN) has proposed Sustainable Development Goal 11 (2015-2030), targeting at decreasing the number of impacted people and economic losses caused by water-related disasters [5]. Therefore, an urgent need remains to map flooded urban areas in near real-time for an improved disaster response service (e.g., rescue and relief missions) and reducing economic loss.

Multisource remote sensing imagery has been widely used for flood mapping. In general, very high resolution (VHR) optical imagery could provide abundant color and texture information for better visual inspections [3,6–8]. Feng et al. [8] mapped flood inundation areas based on VHR aerial optical images acquired by a mini unmanned aerial vehicle (UAV) using a Random Forest (RF) classifier performed in the spectral-texture feature space. Xie et al. [7] considered digital elevation model (DEM) as the spatial dependency information when performing pixel-wise classification with hidden Markov tree (HMT) to identify unseen flood pixels such as pixels under trees. With a focus on flooded object detection, Doshi et al. [3] proposed a convolutional neural network (CNN) based object detection model to detect man made features (i.e., roads) in pre- and post-flooding VHR satellite imagery with Red (R), Green (G), and Blue (B) bands from DigitalGlobe, in which the flood mapping is actually flooded road detection. More recently, Gebrehiwot et al. [6] used image segmentation model, a fully convolutional network (FCN) [9], to classify each pixel into four classes including water, building, vegetation, and road. While the aforementioned studies could produce reasonable flood maps for urban areas, they required very accurate and time-consuming human annotation of training data for model training. Additionally, the VHR optical remote sensing imagery often has heterogeneous background involving various ground objects (e.g., vehicles, buildings, roads, and trees). Due to inconsistent illumination conditions at the time of VHR optical image acquisition, pixels of the same objects in VHR optical imagery could have very different pixel values, leading to the poor generalization capability of previous flood mapping models. For example, floodwaters in the same post-flooding image may have highly inconsistent pixel values as shown in Figure 1. Also such



Figure 1. Floodwaters with inconsistent pixel values on VHR optical imagery from NOAA: The pixels in red circles are all flooded but show different colors

heterogeneous background prohibitively prevents pixel-based classifiers, such as RF, support vector machine (SVM), maximum likelihood (ML), and recent image segmentation models (e.g., FCN), from capturing the underlying spatial correlation between neighboring pixels within the flood hazard area, and therefore from performing well.

Multispectral surface reflectance imagery often contains important spectral information for floodwater detection [1,10–13]. Li et al. [10] used remotely sensed multispectral data, Landsat Thematic Mapper/Enhanced Thematic Mapper Plus (TM/ETM+), to map inundation at a sub-pixel scale via discrete particle swarm optimization (DPSO). Malinowski et al. [12] used a decision tree (DT) algorithm with various combinations of input variables including spectral bands of the WorldView-2 image and spectral indices to analyze spatial patterns of localized flooding on a riverine floodplain. More recently, Wang et al. [13] added the spectral information, normalized difference water index (NDWI), into the traditional super-resolution flood inundation mapping (SRFIM) model to enhance

the model response to floodwaters. Most of the flood mapping studies based on multispectral surface reflectance imagery, however, explored homogeneous rural areas instead of heterogeneous urban areas, where a larger number of people would be in danger during flooding.

By virtue of the radar's active imaging property and its long-wavelength signal with penetration power, the space-borne synthetic aperture radar (SAR) is able to collect data over the flooded regions day and night regardless of weather conditions [1,2,14–16]. Giustarini et al. [15] introduced a Bayesian approach to generate probabilistic flood maps based on SAR data. Shen et al. [14] developed a near real time (NRT) system for flood mapping using SAR data, which involves classification based on statistics, morphological processing, multi-threshold-based compensation, and machine-learning correction. Li et al. [2] proposed an image patch classification model to map the flooded urban area with multi-temporal SAR imagery based on an active self-learning CNN framework, which addressed the issue of limited training data size. Although these studies based on SAR data made significant efforts to improve the accuracy of flood maps, the proposed models were usually complicated in terms of model architectures, and did not perform with very satisfying results in terms of overall accuracy, precision, recall, and F1 score. Moreover, for neural network based deep learning models, a large number of human annotated training samples were required.

Leveraging the advantages of different types of data, Rudner [1] fused multisource satellite imagery, including VHR, multispectral, and radar imagery, in a CNN model to detect flooded buildings in urban areas. As such, the spatial, spectral, and temporal information was integrated to improve the segmentation of flooded ground objects. However, the models discussed above required data from multi-modal sensors, some of which might be missing during flooding events.

With regard to the mapping methods, most of literature focused on pixel-based dense classification approaches such as artificial neural network (ANN) [17], SVM [16], DT [12], RF [8], HMT [7], particle swarm optimization (PSO) [10], and deep CNN such as FCNs [6], U-Net [18], and Deeplab [19]. While pixel based image segmentation approaches in the aforementioned studies could generate higher resolution flood maps, they depend on high resolution flooding masks for model training, which require intensive human annotation of training samples. The annotation process might be even more expensive for urban areas as they are more heterogeneous than rural areas. As such, these models might not be able to perform in near real time when flooding occurs in urban areas.

Some of the studies also investigated patch-based classification methods for land cover mapping, which have the potential for urban flood mapping. Traditional machine learning approaches have been widely used for image scene classification. Gong et al. [20] compared SVM, DT, and RF for Landsat image scene classification and showed that SVM performed with the highest overall accuracy. Heydari et al. [21] also reported the superior classification performance of SVM on 26 testing blocks of Landsat imagery in comparison with ANN and the ensemble of DT. More recently, CNN based deep learning approaches have shown promising performance for image classification, such as AlexNet [22], VGGNet [23], GoogLeNet [24], and ResNet [25]. Most of these neural network models are very deep in terms of the number of layers, which are not necessary for classification of small patches as demonstrated in [26,27]. Sharma et al. [26] developed a patch-based CNN model tailored for medium resolution (pixel size = 30m) multispectral Landsat-8 imagery for land cover mapping, which outperformed pixel-based classifier in overall classification accuracy. Song et al. [27] designed a light CNN (LCNN) model to map the land cover also using Landsat-8 imagery and achieved better results than pixel-based classifiers particularly at heterogeneous pixels, which are very common in urban areas. Additionally, traditional machine learning approaches (e.g., SVM and RF) were also tested and showed competitive results for patch-based classification compared with LCNN. It was also demonstrated that the patch-based approach has an advantage in large scale mapping in terms of computation time. Most recently, with a focus on urban flood mapping, Li et al. [2] proposed a patch-based active self-learning CNN framework to map the flooding areas in urban Houston with multi-temporal SAR imagery. However, there still exist great potentials to simplify the model architecture and improve the F1 score and overall accuracy. Additionally, patch-based approaches to flood mapping especially

over urban areas are still not well investigated. Moreover, considering the advantage of multispectral surface reflectance data, the extensive and quantitative study of patch-based urban flood mapping with multi-temporal multispectral surface reflectance imagery is still lacking. Finally, with the advancement of data acquisition technologies, current remotely sensed data are often in a huge volume, such as satellite images from DigitalGlobe [28] and Planet Labs [29], and aerial imagery collected by National Oceanic and Atmospheric Administration (NOAA [30]). With an increasing data volume, how to fully leverage the big data and to develop more robust and scalable algorithms for flood extent mapping still remains to be addressed.

Motivated by these key issues, this paper proposed a patch similarity convolutional neural network (PSNet) with two variants (i.e., PSNet-v1 and PSNet-v2) for precision urban flood mapping using bi-temporal (i.e., pre- and post-flooding) high resolution (i.e., 3m) multispectral satellite imagery. We used surface spectral reflectance imagery since spectral reflectance is more invariant with respect to data acquisition time and weather conditions. This means that corresponding ground objects from pre- and post-flooding imagery would have consistent spectral responses. As a result, the network would be more robust. Similar to the studies in [6–8,27], we conducted extensive experiments with PSNet and other baseline methods including SVM, DT, RF, and AdaBoost (ADB), using two datasets: 1) the 2017 Hurricane Harvey flood in Houston, Texas, and 2) the 2018 Hurricane Florence flood in Lumberton, North Carolina. We used default parameters in scikit-learn [31] for experiments with baseline methods as in [7]. Experiment results showed that the PSNet with bi-temporal data achieved superior performance in F1 score and overall accuracy compared with baseline methods (i.e., SVM, DT, RF, and ADB) with either uni- or bi-temporal data.

In summary, major contributions of this study include:

- The PSNet was developed with a simplified two-branch CNN-based data fusion framework, which incorporates high resolution bi-temporal multispectral surface reflectance imagery for flood mapping over dense urban residential, commercial, and industrial areas. We transformed uni-temporal flood mapping with only post-flooding data into bi-temporal image patch similarity evaluation with both pre- and post-flooding data. Compared to uni- or bi-temporal SVM, DT, RF, and ADB, PSNet performed consistently stronger in F1 score and overall accuracy.
- This study investigated the role of multispectral surface reflectance imagery in urban flood mapping. The use of spectral reflectance instead of raw pixel digital numbers plays an important role since spectral reflectance reduces the impact of irrelevant changes caused by data acquisition conditions (e.g., illumination).
- The research sheds light on other bi-temporal change detection problems for natural hazard damage evaluation (e.g., earthquake and wildfire damage detection). The proposed models can be easily applied to other types of disaster events without damage-specific model design for feature extraction.

2. Materials and Methods

2.1. Preliminaries

Flood extent mapping is a process to identify the land areas impacted by flooding. Various definitions of such flooding areas have been proposed [2,6–8]. For example, only land areas covered by visible floodwaters are considered as being flooded [6,8]. However, in some works, invisible flooding areas (e.g., hidden floodwaters under tree canopies) and small dry areas but surrounded by floodwaters as shown in Fig. 2 may also be considered as being flooded [7]. The latter definition corresponds to FEMA's National Flood Mapping Products [32], which define flood hazard zones as land areas that are either covered or surrounded by floodwaters [2,7]. For urban flood mapping with high spatial resolution imagery, this paper uses FEMA's definition of flood hazard zones as previous works [2,7] considering expensive pixel-wise flood labeling.



Figure 2. Flood hazard zones, roads and residential areas are flooded, (a) pre-flooding VHR optical imagery from Texas Natural Resources Information Systems (TNRIS), (b) post-flooding VHR optical imagery from NOAA

This study uses bi-temporal multispectral satellite imagery before and after flooding for urban flood mapping. Given a pair of co-registered pre- and post-flooding satellite imagery I_1 and I_2 , the goal of this work is to develop a binary classification model F which takes (I_1, I_2) as input and returns a binary flood hazard map O as output, $O = F(I_1, I_2)$, where each pixel in I_2 is classified as 1 (flood, FL) or 0 (non-flood, NF).

To incorporate bi-temporal imagery for flood mapping over heterogeneous urban areas, it is worth noting that I_1 and I_2 may not align well (i.e., corresponding pixels, $I_1(i, j)$ and $I_2(i, j)$, at the same geographical location, do not exactly refer to the same ground object even though I_1 and I_2 are co-registered). Major reasons include (1) trees grow differently in multi-temporal imagery acquired over different seasons, (2) moving objects (e.g., cars) are quite common over urban areas, and (3) ortho-rectification of I_1 and I_2 may not be perfect due to complex terrains and ground infrastructures (e.g., tall buildings). As a result, pixel-wise analysis of multi-temporal imagery may not perform well for urban flood mapping. Limited by strong heterogeneity over the urban area, this work conducted patch-wise flood mapping.

2.2. Datasets

We studied two flooding events caused by strong hurricanes over the urban areas. One is west Houston, Texas, which was flooded due to Hurricane Harvey in August 2017. The other is the city of Lumberton, North Carolina, which was flooded as a result of Hurricane Florence in September 2018.

The data used in this work are satellite imagery from the Planet Lab [29] with spatial resolution of 3 meters, and 4 spectral bands including blue (B), green (G), red (R), and near infrared (NIR) (see Table 1). All imagery have been orthorectified and radiometrically calibrated into surface spectral reflectance such that the data are more invariant with respect to weather conditions. In addition, the bi-temporal pre- and post-flooding imagery were co-registered for similarity analysis.

Event	Scene & Date	Band	Height, Width (px)	Spatial Resolution (m)	Product
Harvey	Pre, Jul. 31, 2017	B, G, R, NIR	(1848, 3066)	3	Reflectance
	Post, Aug. 31, 2017				
Florence	Pre, Aug. 30, 2018		(2240, 2940)		
	Post, Sept. 18, 2018				

Table 1. Flood imagery data characteristics

The Harvey pre- and post-flooding satellite images over west Houston, Texas, were collected on July 31, 2017 and August 31, 2017, respectively (Table 1). The bi-temporal images were split into non-overlapping patches of spatial size 14×14 . As a result, each patch corresponds to the ground spatial area of $42m \times 42m$, where $42 = 14 \times 3$. We set the patch size approximately equal to the one in a most recent study on urban flood mapping [2], in which the patch size was $40m \times 40m$. As such, the patch-wise classification results for flood mapping over urban Houston area can be compared

qualitatively with the ones in [2]. To label the class of each pair of patches, we use VHR (pixel size=0.3m) aerial imagery acquired by NOAA on August 31, 2017 as reference. More specifically, the VHR image covers the same study area as the pre- and post-flooding multispectral images. Similarly, we cut the VHR image into small patches of size 140×140 such that each VHR patch covers the same spatial area (i.e., $140 \times 0.3 = 42m$) as the multispectral patch. Classes considered in this study are: flooded (FL) patches with floodwaters and non-flooded (NF) patches without floodwaters. It is worth noting that image patches without visible floodwaters were not annotated as FL [2]. Therefore, we obtained 28,908 annotated patches, of which 8,517 are in class FL and 20,391 are in class NF. The pre- and post-flooding images with the labeled ground truth over the whole study area are shown in Figure 3. For model training and evaluation, we randomly sampled 5,000 pairs of patches from the bi-temporal pre- and post-flooding dataset for training and validation, and the rest 23,908 for testing.

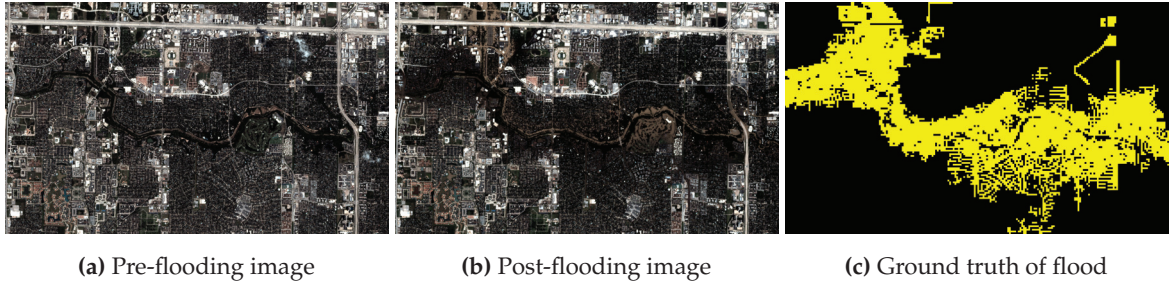


Figure 3. Harvey: optical view of pre- and post-flooding multispectral images with ground truth of flooded patches (FL) highlighted in yellow, and non-flooded patches (NF) in black.

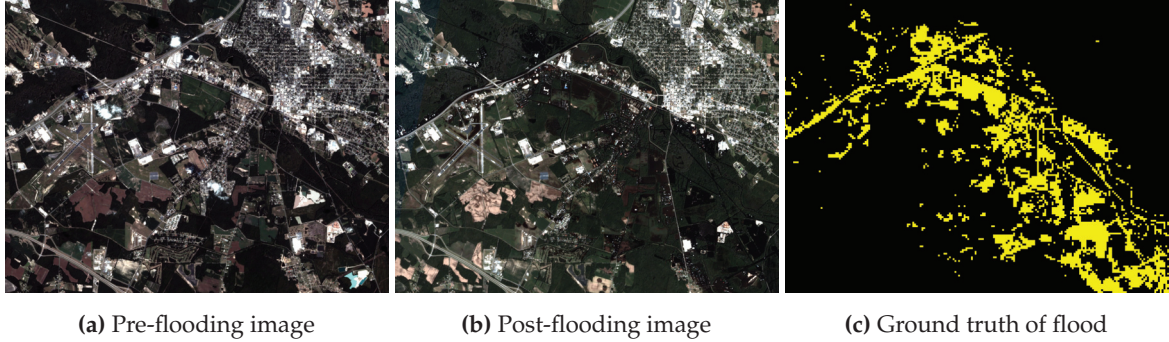


Figure 4. Florence: optical view of pre- and post-flooding multispectral images with ground truth of flooded patches (FL) highlighted in yellow, and non-flooded patches (NF) in black.

The Florence pre- and post-flooding satellite images with corresponding ground truth of flood map (Figure 4) over the Lumberton city were acquired on August 30, 2018 and September 18, 2018, respectively (Table 1). Similar to the data pre-processing for Hurricane Harvey, a total of 33,600 annotated patches were obtained, in which 5,003 are in class FL and 28,597 are in class NF. We randomly sampled 5,000 samples for model training and validation, and keep the remaining 28,600 for testing.

For both Harvey and Florence data with 5,000 samples for training and validation, 4,000 samples were used for training while the rest 1,000 samples were fixed for validation and model selection.

2.3. Methods

2.3.1. Patch Similarity Evaluation

The bi-temporal satellite images, (I_1, I_2) , were divided into non-overlapping image patches, $P_1(m, n)$ and $P_2(m, n)$, of the same size. Each pair of patches cover the same geographic area. Therefore, instead of classifying each pixel pair, $I_1(i, j)$ and $I_2(i, j)$, we predict the class of each patch pair, $P_1(m, n)$ and $P_2(m, n)$, to be either FL or NF. In this study, we evaluate the flooding probability of each patch

pair, $P_1(m, n)$ and $P_2(m, n)$, based on their similarity. Note that we assume that the major dissimilarity between $P_1(m, n)$ and $P_2(m, n)$ is resulted from flooding since the pre- and post-flooding images were collected recently before and shortly after the flooding event, respectively. Accordingly, the patch similarity is negatively correlated with the probability that the patch pair under test is flooded. The less similar of $P_1(m, n)$ and $P_2(m, n)$, the more likely of being flooded.

This work proposed the PSNet to learn the nonlinear mapping from the pre- and post-flooding patch pairs, $P_1(m, n)$ and $P_2(m, n)$, to the output class, FL or NF. Two variants (PSNet-v1 and PSNet-v2) of the network architecture are shown in Figure 5a and 5b, respectively, in which the convolutional operation block (Conv block) is shown in Figure 6. The PSNet in this work basically consists of two modules, *Encoding* and *Decision*.

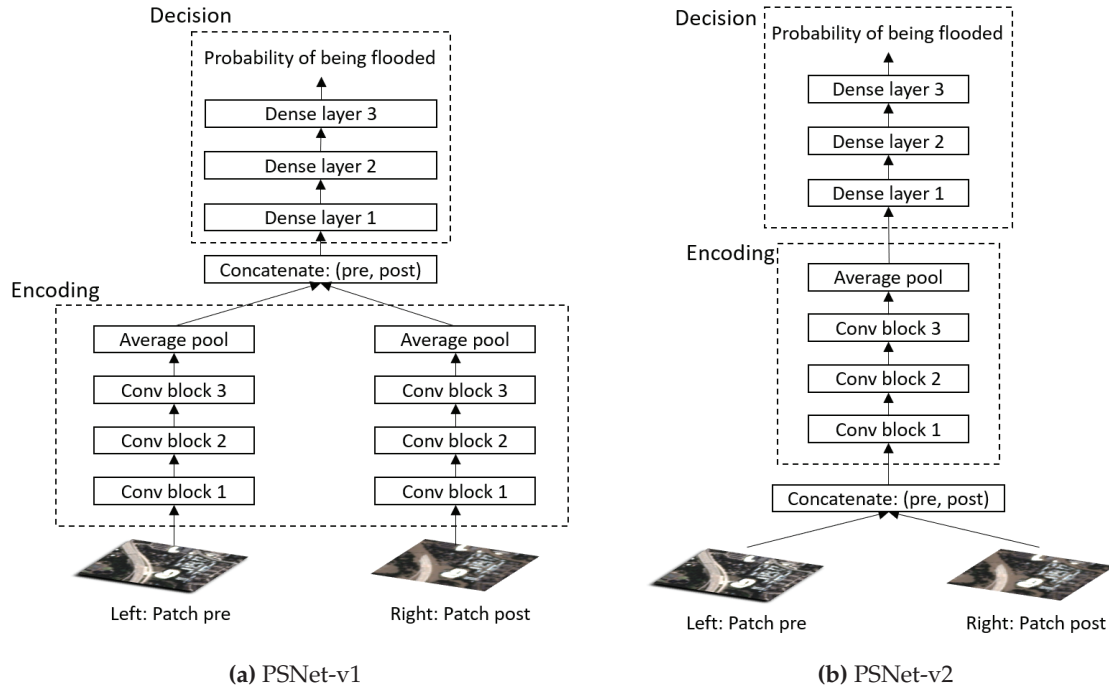


Figure 5. The patch similarity convolutional neural network (PSNet)

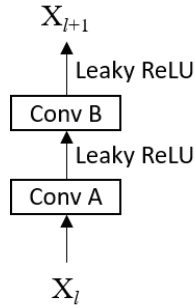


Figure 6. Convolution operation block (Conv block).

The *Encoding* module learns the feature representations from the input pre- and post-flooding patches, respectively. More specifically, in PSNet-v1, the *Encoding* module has a Siamese sub-network architectures on the left and right paths for learning the feature representations from the pre- and post-flooding patches. To perform similarity analysis in the *Decision* module, the left and right sub-networks share the weights [33], which in turn alleviates the computing load. The sub-network in the *Encoding* module contains a stack of convolutional operation blocks which is shown in Figure 6.

Layer	Parameters
Input patch	$C@14 \times 14$, C : $n_channels$
Conv block 1	A: $96, 3 \times 3$, stride=1, pad=1, LeakyReLU (0.1) B: $96, 3 \times 3$, stride=1, pad=1, LeakyReLU (0.1)
Pooling	Max Pooling (2×2)
Conv block 2	A: $192, 3 \times 3$, stride=1, pad=1, LeakyReLU (0.1) B: $192, 3 \times 3$, stride=1, pad=1, LeakyReLU (0.1)
Pooling	Max Pooling (2×2)
Conv block 3	A: $192, 3 \times 3$, stride=1, pad=0, LeakyReLU (0.1) B: $192, 1 \times 1$, stride=1, pad=0, LeakyReLU (0.1)
Pooling	Adaptive average pooling, (1×1)
Concatenation	Feature vector concatenation, pre + post
Dense layer 1	Fully connect, $384 \rightarrow 384$, LeakyReLU (0.1)
Dense layer 2	Fully connect, $384 \rightarrow 192$, LeakyReLU (0.1)
Dense layer 3	Fully connect, $192 \rightarrow 1$, Sigmoid

Table 2. Hyperparameters of the PSNet

Feature representations of pre- and post-flooding patches from the left and right paths would then join on top of the *Encoding* module through concatenation along the channel dimension. Different from PSNet-v1, the other variant PSNet-v2 would first concatenate the pre- and post-flooding patches and then feed the patch stack into the *Encoding* module for joint feature learning.

The *Decision* module evaluates the similarity between the feature representations learned from the pre- and post-flooding patches through the *Encoding* module. It performs binary classification (i.e., FL or NF) by taking as input the joint feature representations, and following a set of dense layers.

Detailed settings and hyperparameters of the PSNet are listed in Table 2.

2.4. Evaluation Metrics

For all experiments, we evaluated the overall accuracy (OA), precision, recall, and F1 score [34–36], which are defined as Equation 1.

$$\begin{aligned}
 OA &= \frac{TP + TN}{TP + FP + TN + FN} \\
 Precision &= \frac{TP}{TP + FP} \\
 Recall &= \frac{TP}{TP + FN} \\
 F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}
 \end{aligned} \tag{1}$$

where TP, FP, TN, FN denote the number of true positives, false positives, true negatives, and false negatives. For comparative analysis, we performed patch classification with baseline algorithms including: support vector machine (SVM), decision trees (DT), random forest (RF), and AdaBoost (AdB). We tested all baselines with uni-temporal data (i.e., post-flooding patches) and bi-temporal data (i.e., pre- and post-flooding patches).

2.5. Model Training and Testing

For training supervised PSNet, we take as input the pre- and post-flooding patch pairs and as target the corresponding true labels (FL or NF). The Adam optimizer [37] is applied with batched patch pairs to minimize the weighted binary cross entropy loss, $L(x, y)$, defined as Equation 2.

$$\begin{aligned}
 L(x, y) &= \frac{1}{N} \sum_{i=1}^N l_i \\
 l_i &= -w_i [y_i \log x_i + (1 - y_i) \log (1 - x_i)]
 \end{aligned} \tag{2}$$

where x is the output of the network (i.e., the probability of being flooded), y is the target label, N is the number of patch pairs in a batch, l_i is the weighted cross entropy loss for the i^{th} patch pair with associated weight w_i . We assigned different weights for the class FL and NF due to high class imbalance of the training data. The sample weight is defined as the complementary of its occurrence frequency in the training set. More specifically, with regard to the training set including $p\%$ FL and $(1 - p)\%$ NF samples, we set the weights of FL and NF samples as $(1 - p)\%$ and $p\%$, respectively.

All models were trained with batched samples for 200 epochs. We initialized the learning rate to be $1e-4$ and divide it by 10 when observing no further decrease of validation loss. Weight decay of $1e-5$ and momentum parameters $(\beta_1, \beta_2) = (0.9, 0.999)$ for the Adam optimizer were used during training.

Considering limited size of the training data, data augmentation was used to enhance the model generalization capability, including random horizontal and vertical flip, rotation of degrees in $[0^\circ, 90^\circ, 180^\circ, 270^\circ]$, and normalization of pixel reflectance into the range of $[0, 1]$.

Before the training process, good weight initialization is important for networks with multiple paths to avoid partial node activation [18]. In this study, the weights were initialized by random sampling from the Gaussian distribution, $\mathcal{N} \sim (0, V/2)$, where V is the number of associated parameters for each node. More specifically, for a $k \times k$ convolutional kernel with C channels in the previous layer, $V = k^2 C$.

To investigate how the training set size may influence the classification performance, we trained all models with different sizes of training set. To be more specific, we randomly sampled various numbers of training samples from the original training subset of size 4,000 and trained multiple PSNet models. Fixed validation and testing subsets were used for model selection and performance evaluation. In this work, we selected trained models with highest validation F1 scores for testing.

All experiments of PSNet were conducted with Pytorch [38] on a Dell workstation with 16 GiB Intel(R) Xeon(R) W-2125 CPU @ 4.00GHz $\times$ 8, 8 GiB Quadro P4000 GPU, and 64-bit Ubuntu 18.04.2 LTS.

3. Results

3.1. Hurricane Harvey Flood

With bi-temporal pre- and post-flooding data, Figures 7a and 7b illustrate classification performance in terms of overall accuracy and F1 score with respect to varying training set sizes. It shows that the PSNet-v1 and PSNet-v2 performed comparatively and outperformed traditional SVM, decision tree, random forest, and AdaBoost with consistently higher overall accuracy and F1 score. In addition, as the size of training set increases, all models tend to generalize better on testing data as demonstrated by increasing overall accuracy and F1 score.

With only uni-temporal post-flooding data, we also compared all models except for PSNet-v1 since PSNet-v1 requires both pre- and post-flooding patches as input of the Siamese sub-networks in the *Encoding* module. Figures 7c and 7d show the learning curves of PSNet and other baseline algorithms, illustrating how overall accuracy and F1 score would change with different training data size. As demonstrated in Figures 7c and 7d, PSNet-v2 performed with significantly higher overall accuracy and F1 score than SVM, decision tree, random forest, and AdaBoost did.

Take one training set of size 1,500 as an example, Table 3 summarized the detailed numerical classification performance with uni- and bi-temporal data in terms of all evaluation metrics with best result highlighted in bold. It is worth noting that the ensemble methods (e.g., random forest and AdaBoost) are likely to produce higher precision but lower recall, which resulted in poor F1 scores and overall accuracy. Unlike other models with good performance on only one metric (e.g., random forest, strong in precision but weak in recall), PSNet could produce consistently good results across all metrics.

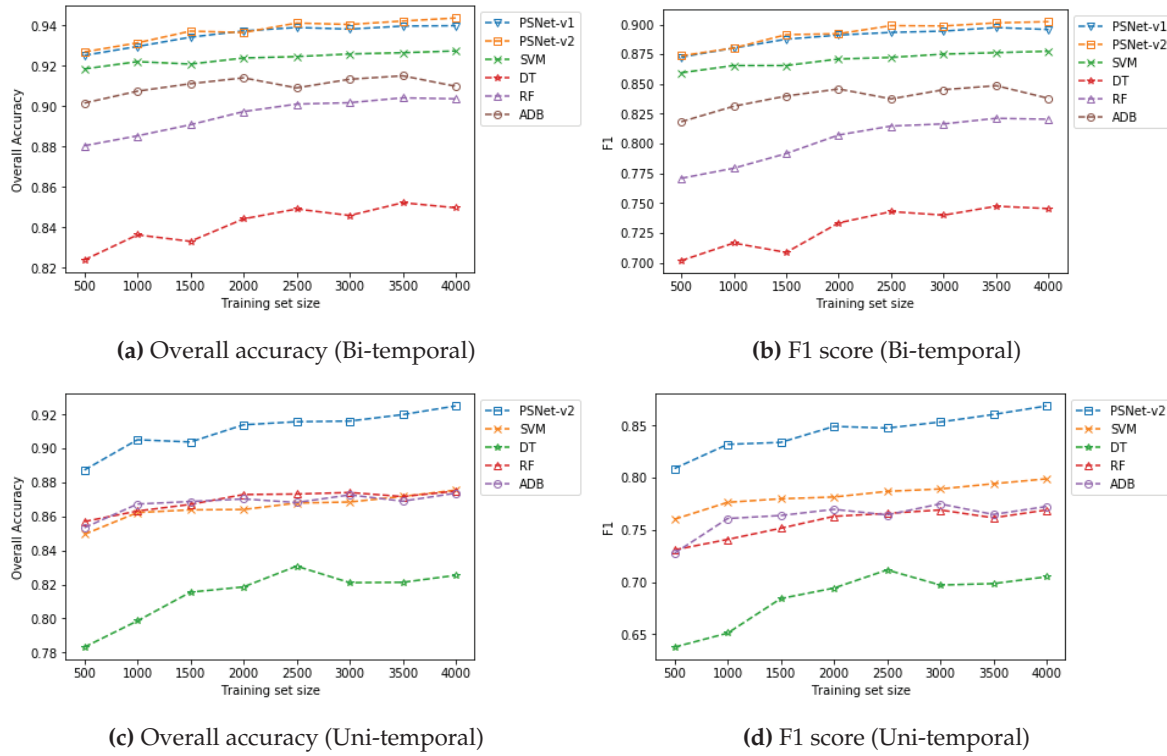


Figure 7. Classification performance on Harvey testing data in terms of OA and F1.

We also observed that, due to limited size and high class imbalance of training data, it remains a challenge for the uni-temporal classifiers to learn the abstract feature representations of the input patch, as reflected by their poorer performance compared with the corresponding bi-temporal classifiers. However, leveraging the bi-temporal information, patch similarity is an important *a priori* for binary classification. Therefore, we do not need to learn the high level abstract features through very deep architectures, which usually require a large number of training data.

For visual inspection, we showed the classification maps of the entire image scene produced by the model trained with 1,500 pairs of pre- and post-flooding patches for PSNet-v1 and PSNet-v2 in Figure 8. Patches in yellow represents the true positives of FL, indicating the correct predictions of flooded patches, and patches in red represent the false alarms of FL, which mean non-flooded patches were detected as being flooded. Patches in green show false negatives of FL, i.e., flooded patches were classified as being non-flooded. Qualitatively compared with ground truth shown in Figure 3c, the proposed PSNet proved to be effective with only 1,500 training samples, as demonstrated by very few false alarms (Red) and false negatives (Green).

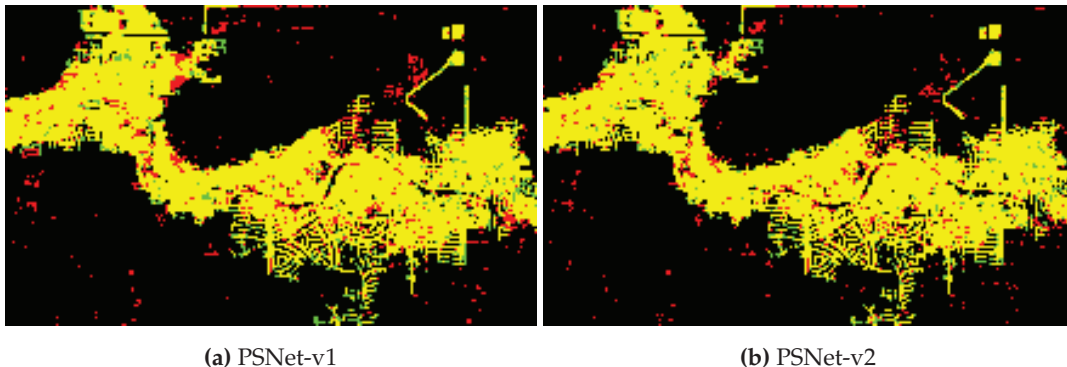


Figure 8. Classification results of Harvey data by PSNet, with patches in yellow for true positives of FL, red for false alarms of FL, and green for false negatives of FL.

Models	Temporal	Precision	Recall	F1	OA
PSNet-v1	pre+post	0.8665	0.9152	0.8876	0.9341
	post	–	–	–	–
PSNet-v2	pre+post	0.8809	0.9073	0.8914	0.9371
	post	0.8272	0.8489	0.8338	0.9038
SVM	pre+post	0.8628	0.8682	0.8655	0.9208
	post	0.7429	0.8207	0.7798	0.8639
DT	pre+post	0.7269	0.6912	0.7086	0.8331
	post	0.6875	0.6811	0.6843	0.8155
RF	pre+post	0.9000	0.7066	0.7916	0.8908
	post	0.8328	0.6848	0.7516	0.8671
ADB	pre+post	0.8909	0.7944	0.8399	0.9111
	post	0.8103	0.7224	0.7638	0.8688

Table 3. Classification performance comparison with 1,500 uni- and bi-temporal Harvey training samples.

3.2. Hurricane Florence Flood

Figure 9 shows the change of overall accuracy and F1 score with respect to the training data size using uni- and bi-temporal data generated during the Hurricane Florence Flood. It is obvious

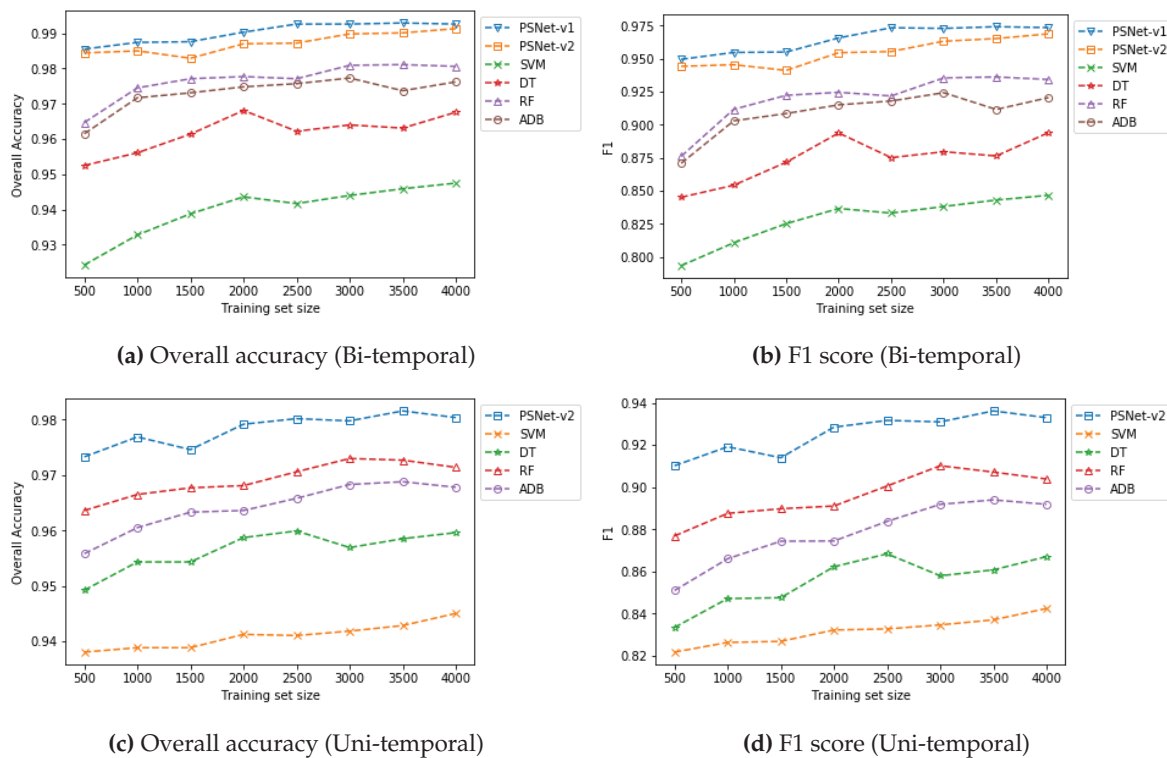


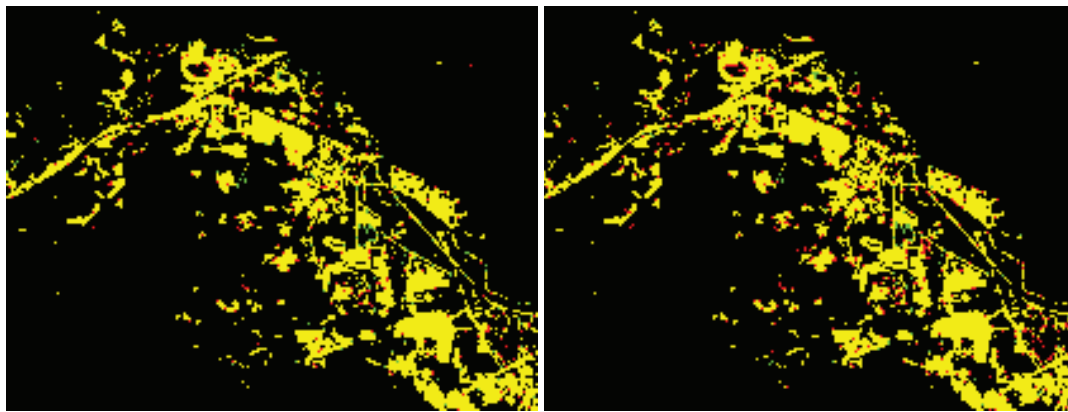
Figure 9. Classification performance on Florence testing data in terms of OA and F1.

that, with both uni- and bi-temporal data, PSNet performed consistently better than SVM, decision tree, random forest, and AdaBoost in terms of F1 score and overall accuracy. Table 4 summarizes the evaluation results by the model trained with 1,500 uni- and bi-temporal samples in terms of precision, recall, F1, and overall accuracy. PSNet-v1 with bi-temporal pre- and post-flooding data achieved very high performance with 0.9551 F1 score and 0.9876 overall accuracy.

Models	Temporal	Precision	Recall	F1	OA
PSNet-v1	pre+post	0.9476	0.9684	0.9551	0.9876
	post	–	–	–	–
PSNet-v2	pre+post	0.9116	0.9792	0.9412	0.9829
	post	0.8625	0.9808	0.9139	0.9746
SVM	pre+post	0.7187	0.9686	0.8251	0.9388
	post	0.7156	0.9791	0.8268	0.9388
DT	pre+post	0.8637	0.8797	0.8716	0.9614
	post	0.8428	0.8523	0.8475	0.9543
RF	pre+post	0.9343	0.9107	0.9223	0.9771
	post	0.9076	0.8725	0.8897	0.9677
ADB	pre+post	0.9210	0.8964	0.9085	0.9731
	post	0.8923	0.8570	0.8743	0.9633

Table 4. Classification performance comparison with 1,500 uni- and bi-temporal Florence training samples.

Figure 10 displays classification maps of the entire image scene for visual interpretation. With only a few false positives (Red) and false negatives (Green), PSNet could produce highly accurate flood maps over the urban area.



(a) PSNet-v1

(b) PSNet-v2

Figure 10. Classification results of Florence data by PSNet, with patches in yellow for true positives of FL, red for false alarms of FL, and green for false negatives of FL.

4. Discussions

Unlike pixel based classification for flood mapping [6–8,10,39,40], this study investigated image patch based flood mapping similar to the study in [2]. Major motivations include: 1) reducing the impact of heterogeneous image background over urban area, which is challenging for pixel-wise classification; and 2) accelerating human annotation of training samples since pixel-wise labeling would be much more time-consuming and labor-intensive.

Similar to the studies in [2,6–8,27] for comparative analysis, we performed patch-based classification with traditional machine learning models as baselines, including SVM, DT, RF, and ADB. The experiment results of the two urban flood events (i.e., the 2017 Hurricane Harvey flood and the 2018 Hurricane Florence flood) demonstrate the superior performance of the proposed PSNet over all baseline algorithms (Figure 7 and 9, Table 3 and 4). With regard to patch-based classification models, the PSNet developed in this study leveraged an efficient two-branch data fusion framework specifically for urban flood mapping. It is worth noting that the *Encoding* module can be developed with different variants of the patch-based CNN architecture used in this study. As a result, the specific architecture of the *Encoding* module along with its hyperparameters used in this study can be considered as a representative of patch-based CNN encoding for the input patches. This work

did not experiment with image segmentation models (e.g., FCNs, U-Net, and Deeplab) since image segmentation works for pixel-based, instead of patch-based, dense classification. In addition, we did not compare with deep image classification models, such as AlexNet, VGGNet, GoogLeNet, and ResNet, since classification of small patches does not require such deep architectures [26,27]. With regard to other CNN-based patch classification models discussed in [26,27], direct comparison is not valid due to different input dimensions and image resolutions which require major modification of the *Encoding* module architectures and tuning of hyperparameters.

More specifically, with regard to patch-based urban flood mapping, this study followed the experiment settings of a recent research for urban flood mapping with SAR data [2], in which the study area (i.e., west Houston) is smaller than the one investigated in this study. For reference, we used patches of size 14×14 to cover the ground area of $42m \times 42m$, which is close to the area (i.e., $40m \times 40m$) covered by patches used in [2]. We did not experiment with the exact same size of patches due to the constraint of different spatial resolutions of images used in the two studies. It should be noted that we labeled all patches with floodwaters as being flooded, whereas, in [2], only patches that were severely flooded were labeled as being flooded. In other words, there are less patches in [2] labeled as being flooded than that of this study. For patches that were partially covered by floodwaters but not heavily flooded, the classification model would have very weak response. Therefore, the results in this study cannot be directly and quantitatively compared with those in [2]. For qualitative comparison regarding the Harvey flooding event, as reported in [2], the developed active self-learning CNN model detected flood patches with precision of 0.684, recall of 0.824, F1 score of 0.746, and overall accuracy of 0.928 when using model trained with 600 pre- and post-flooding SAR patches. However, this study achieved the performance with precision of 0.848, recall of 0.906, F1 score of 0.873, and overall accuracy of 0.925 with model (PSNet-v1) trained with 500 bi-temporal multispectral patches (Table 5). In addition, the PSNet was designed with simple architectures for easy implementation. More importantly, it shows that only a small number (e.g., 500) of training samples are needed for training a competitive model (PSNet) that generalizes well on the testing data, and thus contributes to quick mapping of the flooding area.

Models	Temporal	Precision	Recall	F1	OA
PSNet-v1	pre+post	0.848	0.906	0.873	0.925
PSNet-v2	pre+post	0.867	0.887	0.874	0.927

Table 5. Classification performance comparison with 500 bi-temporal Harvey training samples.

With experiments on both uni- and bi-temporal data, the results show that bi-temporal pre- and post-flooding data contribute significantly to boosting the performance of PSNet for patch similarity analysis and thus for flood patch identification. Patch similarity learning has proved to be effective in patch-based matching of stereo images [33,41–43]. Due to the heterogeneity of the satellite imagery background over urban area, patches of class FL usually have various patterns which are difficult to be learned by the classification algorithms with very limited number of training samples. As shown in Figure 7 and 9, patch similarity evaluation based PSNet with bi-temporal data consistently outperformed those floodwater pattern recognition based models with uni-temporal data. It is worth noting that, with only 500 training samples available, the proposed PSNet was able to perform with, approximately, F1 score of 0.87 and overall accuracy of 0.93 on Harvey testing data. Similar high performance can also be observed in the experiment for the Florence data.

We investigated the important role of spectral reflectance in urban flood mapping. As spectral reflectance has been recognized as the signature of ground objects [44], it would be more invariant with respect to illumination conditions. Therefore, with only a small number of human annotated samples (e.g., 1,500), we could identify the flood image patches with around 0.8914 F1 score and 0.9371 overall accuracy for Harvey testing data (Table 3) and 0.9551 F1 score and 0.9876 overall accuracy for Florence testing data (Table 4), which are consistently better than the results produced by the baseline algorithms. Compared with studies using SAR imagery [2] and optical imagery with raw pixel digital

numbers [8], spectral reflectance data in this study play an important role in helping PSNet achieve superior performance in urban flood mapping with merely a small number of training samples (e.g., 500) as demonstrated by the learning curves in Figures 7 and 9.

It is worth noting that PSNet achieved higher F1 score and overall accuracy on the Florence data (Table 4) than that on the Harvey data (Table 3). It is mainly because the Harvey data covering the west Houston area are more heterogeneous than the Florence data covering the Lumberton city. More specifically, the west Houston area includes dense residential, industrial, and commercial regions, where various ground objects result in more heterogeneous image background. As a result, it would be relatively easier for the PSNet trained with the Florence training data to achieve better performance on the Florence testing data.

With regard to the processing time on model training and testing for creating the flood maps, it took about 6 minutes to train the PSNet with 500 samples and 1 minute to create the flood map of the study area (e.g., west urban Houston) on the Dell workstation used in this work. The running time associated with traditional approaches (e.g., SVM, DT, RF, ADB) is even less than that corresponding to the PSNet. As such, the time consumption on PSNet training and testing can be ignored for near real-time urban flood mapping. It should be noted that the major time-consuming process is human annotation of training samples. As in this study, a total of 3 research assistants could label 500 training samples in less than 20 minutes, which can also be ignored for near real-time urban flood mapping.

To sum up, the major strength of the proposed PSNet with bi-temporal data is to map the urban flood area with high overall accuracy and F1 score as demonstrated by the quantitative results in Figures 7 and 9. More detailed evaluation results over all metrics corresponding to 1,500 training samples can be found in Tables 3 and 4. One major limitation of this study in practice is that part of the satellite imagery covering the flood area may contain clouds, which are the major challenge for multispectral image analysis. In this case, further work could be dedicated to fusing both multispectral imagery and SAR imagery for joint urban flood mapping by virtue of the penetration power of the SAR signals [1].

5. Conclusions

This paper addressed the challenge of urban flood mapping via patch similarity learning instead of pixel-wise classification, since patch based analysis reduced the impact of heterogeneous image background over urban area and contributed to efficient annotation of training samples. We proposed the patch similarity network (PSNet) with two variants (PSNet-v1 and PSNet-v2) to evaluate the similarity between bi-temporal pre- and post-flooding patches cropped from the surface spectral reflectance imagery, and thus to determine whether the post-flooding patch under test is flooded. Results showed that both PSNet-v1 and PSNet-v2 developed in this study achieved superior performance with approximately 89% F1 score and 93% overall accuracy on the 2017 Hurricane Harvey flood testing data, and 95% F1 score and 98% overall accuracy on the 2018 Hurricane Florence flood testing data with only 1,500 training samples available. Extensive experiments with PSNet and other baseline algorithms demonstrated high performance of PSNet. Moreover, it is not required to design handcrafted floodwater related features for PSNet, which further improves the generalization capability of PSNet. While multispectral reflectance imagery used in this study maybe influenced by severe weather conditions (e.g., heavy clouds), they are effective and accurate in urban flood mapping.

In the future, we would experiment with data for other types of disaster events (e.g., California wildfires in 2018) to test the model generalization ability. Moreover, as multispectral imagery might be cloudy for some flooding events, resulting in insufficient data. The fusion of SAR and multispectral imagery may help reduce the impact clouds, which contributes to near real-time urban flood mapping.

Author Contributions: Conceptualization, Bo Peng, Qunying Huang and Caixia Wang; Data curation, Bo Peng and Zonglin Meng; Formal analysis, Bo Peng, Qunying Huang and Caixia Wang; Funding acquisition, Qunying Huang; Investigation, Bo Peng, Zonglin Meng, Qunying Huang and Caixia Wang; Methodology, Bo Peng and Qunying Huang; Project administration, Bo Peng and Qunying Huang; Resources, Qunying Huang; Software, Bo

Peng and Zonglin Meng; Supervision, Qunying Huang; Validation, Bo Peng; Visualization, Bo Peng; Writing – original draft, Bo Peng; Writing – review & editing, Bo Peng, Qunying Huang and Caixia Wang.

Funding: This research is funded by the National Science Foundation with project ID: 1940091.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rudner, T.G.J.; Rußwurm, M.; Fil, J.; Pelich, R.; Bischke, B.; Kopackova, V.; Bilinski, P. Multi³Net: Segmenting Flooded Buildings via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery. 2019.
2. Li, Y.; Martinis, S.; Wieland, M. Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *152*, 178–191. doi:10.1016/j.isprsjprs.2019.04.014.
3. Doshi, J.; Basu, S.; Pang, G. From Satellite Imagery to Disaster Insights. 32nd Conference on Neural Information Processing Systems Workshop, 2018.
4. Hallegatte, S.; Green, C.; Nicholls, R.J.; Corfee-Morlot, J. Future flood losses in major coastal cities. *Nature climate change* **2013**, *3*, 802.
5. United Nations. The Sustainable Development Goals Report, 2018.
6. Gebrehiwot, A.; Hashemi-Beni, L.; Thompson, G.; Kordjamshidi, P.; Langan, T. Deep Convolutional Neural Network for Flood Extent Mapping Using Unmanned Aerial Vehicles Data. *Sensors* **2019**, *19*, 1486. doi:10.3390/s19071486.
7. Xie, M.; Jiang, Z.; Sainju, A.M. Geographical hidden markov tree for flood extent mapping. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2545–2554.
8. Feng, Q.; Liu, J.; Gong, J. Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier A Case of Yuyao, China. *Water* **2015**, *7*, 1437–1455. doi:10.3390/w7041437.
9. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
10. Li, L.; Chen, Y.; Yu, X.; Liu, R.; Huang, C. Sub-pixel flood inundation mapping from multispectral remotely sensed images based on discrete particle swarm optimization. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *101*, 10–21. doi:10.1016/j.isprsjprs.2014.11.006.
11. Cian, F.; Marconcini, M.; Ceccato, P. Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data. *Remote Sensing of Environment* **2018**, *209*, 712–730. doi:10.1016/j.rse.2018.03.006.
12. Malinowski, R.; Groom, G.; Schwanghart, W.; Heckrath, G. Detection and delineation of localized flooding from WorldView-2 multispectral data. *Remote sensing* **2015**, *7*, 14853–14875.
13. Wang, P.; Zhang, G.; Leung, H. Improving super-resolution flood inundation mapping for multispectral remote sensing image by supplying more spectral information. *IEEE Geoscience and Remote Sensing Letters* **2018**.
14. Shen, X.; Anagnostou, E.N.; Allen, G.H.; Brakenridge, G.R.; Kettner, A.J. Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar. *Remote Sensing of Environment* **2019**, *221*, 302–315. doi:10.1016/j.rse.2018.11.008.
15. Giustarini, L.; Hostache, R.; Kavetski, D.; Chini, M.; Corato, G.; Schlaffer, S.; Matgen, P. Probabilistic flood mapping using synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing* **2016**, *54*, 6958–6969.
16. Insom, P.; Cao, C.; Boonsrimuang, P.; Liu, D.; Saokarn, A.; Yomwan, P.; Xu, Y. A support vector machine-based particle filter method for improved flooding classification. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 1943–1947.
17. Skakun, S. A neural network approach to flood mapping using satellite imagery. *Computing and Informatics* **2012**, *29*, 1013–1024.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; Navab, N.; Hornegger, J.; Wells, W.M.; Frangi, A.F., Eds.; Springer International Publishing: Cham, 2015; pp. 234–241.

19. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs **2018**. *40*, 834–848.
20. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; others. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* **2013**, *34*, 2607–2654.
21. Heydari, S.S.; Mountrakis, G. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sensing of Environment* **2018**, *204*, 648–658.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural. *Neural Information Processing Systems*, 2014, pp. 1–9.
23. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
24. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
26. Sharma, A.; Liu, X.; Yang, X.; Shi, D. A patch-based convolutional neural network for remote sensing image classification **2017**. *95*, 19–28.
27. Song, H.; Kim, Y.; Kim, Y. A Patch-Based Light Convolutional Neural Network for Land-Cover Mapping Using Landsat-8 Images. *Remote Sensing* **2019**, *11*, 114.
28. DigitalGlobe. Open Data Program.
29. Planet Team. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA., 2018.
30. National Oceanic and Atmospheric Administration. Data and Imagery from NOAA National Geodetic Survey.
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
32. FEMA. Federal Emergency Management Agency Flood Mapping Products.
33. Zbontar, J.; LeCun, Y. Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research* **2016**, *17*, 2.
34. Manning, C.D.; Manning, C.D.; Schütze, H. *Foundations of statistical natural language processing*; MIT press, 1999.
35. Raghavan, V.; Bollmann, P.; Jung, G.S. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems (TOIS)* **1989**, *7*, 205–229.
36. Manning, C.; Raghavan, P.; Schütze, H. Introduction to information retrieval. *Natural Language Engineering* **2010**, *16*, 100–103.
37. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization, 2014. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
38. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. *NIPS-W*, 2017.
39. Benoudjit, A.; Guida, R. A Novel Fully Automated Mapping of the Flood Extent on SAR Images Using a Supervised Classifier. *Remote Sensing* **2019**, *11*, 779. doi:10.3390/rs11070779.
40. Sublime, J.; Kalinicheva, E. Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami. *Remote Sensing* **2019**, *11*, 1123. doi:10.3390/rs11091123.
41. Zagoruyko, S.; Komodakis, N. Deep compare: A study on using convolutional neural networks to compare image patches. *Computer Vision and Image Understanding* **2017**, *164*, 38–55. doi:10.1016/J.CVIU.2017.10.007.
42. Zagoruyko, S.; Komodakis, N. Learning to Compare Image Patches via Convolutional Neural Networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4353–4361.

- 534 43. Han, X.; Leung, T.; Jia, Y.; Sukthankar, R.; Berg, A.C. MatchNet: Unifying Feature and Metric Learning for
535 Patch-Based Matching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*,
536 2015, pp. 3279—3286.
- 537 44. Chang, C.I. *Hyperspectral imaging: techniques for spectral detection and classification*; Vol. 1, Springer Science &
538 Business Media, 2003.

539 © 2020 by the authors. Submitted to *Remote Sens.* for possible open access publication
540 under the terms and conditions of the Creative Commons Attribution (CC BY) license
541 (<http://creativecommons.org/licenses/by/4.0/>).