

Article

Patch Similarity Convolutional Neural Network for Urban Flood Extent Mapping Using Bi-Temporal Satellite Multispectral Imagery

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

¹ Department of Geography, University of Wisconsin—Madison, Madison, WI 53706, USA; bo.peng@wisc.edu

² Department of Electrical and Computer Engineering, University of Wisconsin—Madison, Madison, WI 53706, USA

³ Department of Computer Sciences, University of Wisconsin—Madison, Madison, WI 53706, USA; zmeng32@wisc.edu

⁴ Department of Geomatics, University of Alaska Anchorage, Anchorage, AK 99508, USA; cwang12@alaska.edu

* Correspondence: qhuang46@wisc.edu

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

Received: 1 September 2019; Accepted: 20 October 2019; Published: 24 October 2019



Abstract: Urban flooding is a major natural disaster that poses a 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. Unfortunately, the near real-time production of accurate flood maps over impacted urban areas has not been well investigated due to three major issues. (1) Satellite imagery with high spatial resolution over urban areas usually has nonhomogeneous background due to different types of objects such as buildings, moving vehicles, and road networks. As such, classical machine learning approaches hardly can model the spatial relationship between sample pixels in the flooding area. (2) Handcrafted features associated with the data are usually required as input for conventional flood mapping models, which may not be able to fully utilize the underlying patterns of a large number of available data. (3) High-resolution optical imagery often has varied pixel digital numbers (DNs) for the same ground objects as a result of highly inconsistent illumination conditions during a flood. Accordingly, traditional methods of flood mapping have major limitations in generalization based on testing data. To address the aforementioned issues in urban flood mapping, we developed a patch similarity convolutional neural network (PSNet) using satellite multispectral surface reflectance imagery before and after flooding with a spatial resolution of 3 meters. We used spectral reflectance instead of raw pixel DNs so that the influence of inconsistent illumination caused by varied weather conditions at the time of data collection can be greatly reduced. Such consistent spectral reflectance data also enhance the generalization capability of the proposed model. 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 poor in either precision or recall. The study paves the way to fuse bi-temporal remote sensing images for near real-time precision damage mapping associated with other types of natural hazards (e.g., wildfires and earthquakes).

Keywords: flood mapping; patch similarity; convolutional neural network; bi-temporal multispectral; deep learning; natural hazards

1. Introduction

Natural hazards have been one of the major causes leading to the great risk of human lives and huge economic losses [1]. Flooding is one type of natural hazards, and has frequently visited coastal cities along with hurricanes, causing severe damages on city infrastructures such as transportation and communications systems, water and power lines, buildings, etc. [2–4]. Recently, improving the safety of human settlements and resilience of cities has become increasingly imminent. As such, the United Nations (UN) has proposed Sustainable Development Goal 11 (2015–2030) to decrease the number of impacted people by water-related disasters and the attributed financial losses [5]. As a result, near real-time urban flood extent mapping is necessary in response to emergency rescue and relief missions as well as reducing financial losses.

Remote sensing (satellite or aerial) imagery has been widely used for large-scale mapping of natural disasters such as flood extent mapping. Specifically, this include three types of image data. One is optical imagery with raw pixel digital numbers (DNs) which can be directly used for visual inspection, such as very high resolution (VHR) aerial imagery with abundant textures and colors [4,6–8]. A number of studies have demonstrated the effective application of VHR optical imagery for flood mapping. Using the VHR optical imagery collected by an unmanned aerial vehicle (UAV), Feng et al. [8] conducted urban flood mapping with a Random Forest (RF) classifier and the handcrafted spectral-texture feature. 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. [4] 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 [9], 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) [10], 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 imagery usually has nonhomogeneous background due to various types of objects in the scene such as buildings, moving vehicles, and road networks. Moreover, the high-resolution optical imagery often has different pixel DNs for the same ground objects (see Figure 1) due to highly inconsistent illumination conditions during a flood. As such, traditional flood mapping approaches may not generalize well on testing data.



Figure 1. VHR optical imagery with varied pixel DNs for the same objects (e.g., floodwaters in blue circles). Data are collected by National Oceanic and Atmospheric Administration (NOAA) [11] during the 2017 Hurricane Harvey Flood at Houston, Texas.

Furthermore, pixel-based classifiers hardly can model the spatial relationship between sample pixels in the flooding area due to heterogeneous image background. Therefore, traditional machine learning approaches such as RF, support vector machine (SVM), maximum likelihood (ML), and recent image segmentation models (e.g., FCN) may not perform well with VHR optical imagery.

Another type of image data involves multispectral optical surface reflectance imagery which contains consistent and distinct spectral information associated with floodwaters [2,12–15]. Li et al. [12] performed the discrete particle swarm optimization (DPSO) for sub-pixel flood mapping using satellite multispectral reflectance imagery, the Landsat Thematic Mapper/Enhanced Thematic Mapper Plus (TM/ETM+) data. Malinowski et al. [14] 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. [15] 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.

The third type of image data widely used for natural disaster mapping is satellite synthetic aperture radar (SAR) imagery which can be acquired during the day or the night regardless of weather conditions due to radar's longwave active signals with penetration power for imaging [2,3,16–18]. Giustarini et al. [17] introduced a Bayesian approach to generate probabilistic flood maps based on SAR data. Shen et al. [16] 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. [3] 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 [2] proposed to fuse multimodal satellite data (i.e., VHR optical imagery with raw pixel DN_s, multispectral reflectance imagery, and SAR imagery) in a CNN model for flooded building detection 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 multimodal sensors, some of which might be missing during floods.

With regard to the mapping methods, most of literature focused on pixel-based dense classification approaches such as artificial neural network (ANN) [19], SVM [18], DT [14], RF [8], HMT [7], particle swarm optimization (PSO) [12], and deep CNN such as FCNs [6], U-Net [20], and Deeplab [21]. 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. [22] compared SVM, DT, and RF for Landsat image scene classification and showed that SVM performed with the highest overall accuracy. Heydari et al. [23] 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 [24], VGGNet [25], GoogLeNet [26], and ResNet [27]. 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 [28,29]. Sharma et al. [28] developed a patch-based CNN model tailored for medium resolution (pixel size = 30 m) multispectral Landsat-8 imagery for land cover mapping, which

outperformed pixel-based classifier in overall classification accuracy. Song et al. [29] 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. [3] 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. Furthermore, there has been an increasing number of available remote sensing data provided by private sectors (e.g., Planet Labs [30] and DigitalGlobe) and government agencies (e.g., NOAA). Such a huge volume of satellite or aerial imagery presents a significant challenge to traditional machine learning approaches in data analysis. Therefore, it is imperative to develop scalable and efficient algorithms for high throughput computing given such big data that near real-time flood mapping could be acquired.

To address the aforementioned challenges, we proposed a patch similarity convolutional neural network (PSNet) with two variants (i.e., PSNet-v1 and PSNet-v2) to map the flooding extent in urban areas using satellite multispectral surface reflectance imagery before and after flooding with a spatial resolution of 3 meters. It is worth noting that we used surface reflectance instead of raw pixel DNs to alleviate the impact of inconsistent illumination and different weather conditions at the time of data collection. As a result, corresponding ground objects from the bi-temporal (pre- and post-flooding) imagery would have consistent surface reflectance. Similar to the studies in [6–8,29], 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.

Main contributions of this work are summarized in the following:

- The proposed PSNet is a simplified two-branch CNN-based data fusion framework, performing urban flood extent mapping with pre- and post-flooding satellite multispectral surface reflectance imagery. Uni-temporal image patch classification with only post-flooding imagery was transformed into bi-temporal patch similarity estimation with both pre- and post-flooding data. Compared to uni- or bi-temporal SVM, DT, RF, and ADB, PSNet performed consistently better in F1 score and overall accuracy.
- This research demonstrated that multispectral surface reflectance data play a significant role in floodwater detection. Compared with raw pixel DNs, surface reflectance is more stable under varied inconsistent illumination conditions.
- The study paves the way to fuse bi-temporal remote sensing images for near real-time precision damage mapping associated with other types of natural hazards such as earthquakes, landslides, wildfires, etc.

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 [3,6–8]. For example, only land areas covered by visible floodwaters are considered as being flooded [6,8]. However, according to the National Flood Mapping Products from the Federal Emergency Management Agency (FEMA) [32], small areas covered by invisible floodwaters due to trees or surrounded by floodwaters are also treated as being flooded [3,7] (see Figure 2).



Figure 2. Flooding hazard areas. (a) aerial optical imagery from Texas Natural Resources Information Systems (TNRIS) before flooding; and (b) aerial optical imagery from NOAA after flooding.

For urban flood mapping with high spatial resolution imagery, this paper uses FEMA’s definition of flood hazard zones as previous works [3,7] considering expensive pixel-wise flood labeling.

This study proposed to map urban flooding areas using bi-temporal pre- and post-flooding satellite multispectral surface reflectance imagery. Given the bi-temporal co-registered satellite images I_1 (before flooding) and I_2 (after flooding), this work aims at developing 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).

While incorporating bi-temporal imagery for flood mapping over heterogeneous urban areas, it is worth noting that I_1 and I_2 may not align well. Corresponding pixels, $I_1(i, j)$ and $I_2(i, j)$, at the same geographical location, may not exactly refer to the same ground object even though I_1 and I_2 are co-registered. This may be due to three major reasons: (1) trees grow and therefore display 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-based analysis of multi-temporal imagery may not perform well for urban flood mapping. To overcome the limitation of high heterogeneity over the urban area, this work conducted patch-based flood mapping.

2.2. Datasets

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

The data used in this study were satellite imagery from the Planet Lab [30] with a spatial resolution of 3 m, and four spectral bands including blue (B), green (G), red (R), and near infrared (NIR) (see Table 1). All imagery were orthorectified and radiometrically calibrated into surface spectral reflectance

such that the data are more independent from weather conditions. In addition, the bi-temporal pre- and post-flooding imagery were co-registered for similarity analysis.

Table 1. Flood imagery data characteristics.

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

The Harvey pre- and post-flooding satellite multispectral images over west Houston, Texas, were collected on 31 July 2017 and 31 August 2017, respectively (Table 1). The bi-temporal images were split into non-overlapping patches of spatial size 14×14 . Since the spatial resolution of the satellite images (i.e., pixel size) is 3 m, all image patches cover a ground area of $42 \text{ m} \times 42 \text{ m}$ ($42 = 14 \times 3$). The patch size was set to a similar value used in a most recent study on urban flood mapping [3], in which patches of size $40 \text{ m} \times 40 \text{ m}$ were cropped from the original SAR imagery. Therefore, patch-based classification in this study for mapping flooded urban areas can be compared qualitatively with the existing work [3]. Regarding the annotation for the classes of post-flooding satellite multispectral patches, two classes are defined including flooded (FL) patches with floodwaters and non-flooded (NF) patches without floodwaters, and image patches without visible floodwaters were annotated as NF [3]. During the annotation, aerial VHR optical images with a spatial resolution of 0.3 m collected by NOAA on 31 August 2017 were used as reference. Specifically, we used the aerial VHR optical image with the same ground coverage as the Harvey pre- and post-flooding satellite multispectral images. Non-overlapping patches of size 140×140 were cropped from the VHR image such that each VHR patch corresponds to a pair of pre- and post-flooding satellite multispectral patches. As each VHR patch covers the same ground area (i.e., $140 \times 0.3 = 42 \text{ m}$) as the satellite multispectral patch, the satellite multispectral patches were labeled by visual inspection of the corresponding VHR patches. Therefore, we obtained 28,908 annotated patches, of which 8517 are in class FL and 20,391 are in class NF. Figure 3 shows the Harvey pre- and post-flooding satellite multispectral surface reflectance images with ground truth over the whole study area. For model training and evaluation, we randomly sampled 5000 pairs of patches from the bi-temporal pre- and post-flooding dataset for training and validation, whereas the rest of samples (23,908) were used for testing.

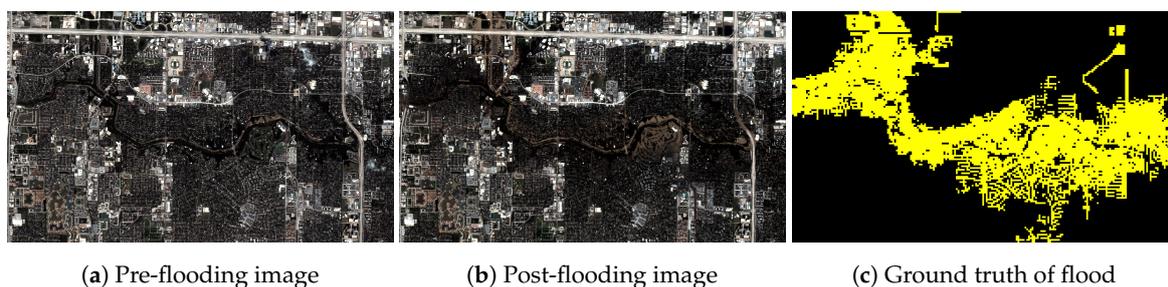


Figure 3. Harvey: Optical view of pre- and post-flooding satellite multispectral surface reflectance images with ground truth of flooded patches (FL) in yellow, and non-flooded patches (NF) in black.

The Florence pre- and post-flooding satellite images with corresponding ground truth of flood map (Figure 4) over the Lumberton city were acquired on 30 August 2018 and 18 September 2018, respectively (Table 1). Similar to the data pre-processing for Hurricane Harvey, 33,600 annotated patches were obtained, of which 5003 are in class FL and 28,597 are in class NF. We randomly sampled 5000 samples for model training and validation, and kept the remaining 28,600 for testing.

For both Harvey and Florence data with 5000 samples for training and validation, 4000 samples were used for training while the rest 1000 samples were fixed for validation and model selection.

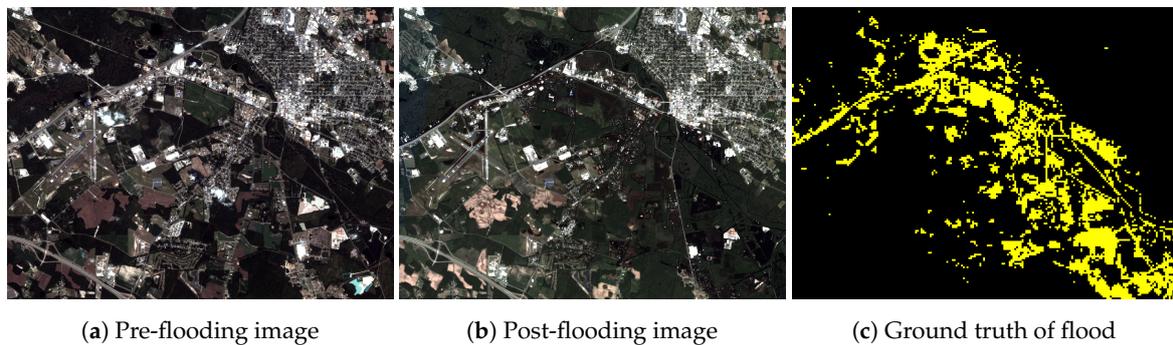


Figure 4. Florence: Optical view of pre- and post-flooding satellite multispectral surface reflectance images with ground truth of flooded patches (FL) in yellow, and non-flooded patches (NF) in black.

2.3. Methods

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 covers 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 intermediately 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 are $P_1(m, n)$ and $P_2(m, n)$, the more likely they are 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,b, respectively, in which the convolutional block (ConvBlock) is shown in Figure 6. The PSNet in this work basically consists of two modules, *Encoding* and *Decision*.

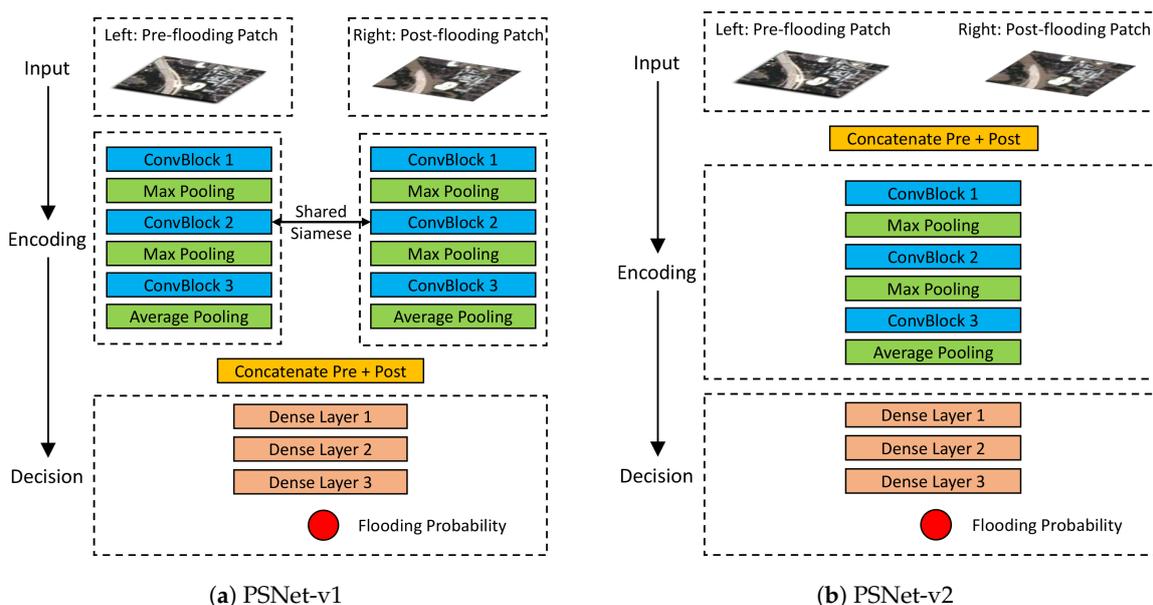


Figure 5. The patch similarity convolutional neural network (PSNet).

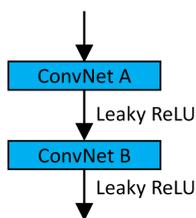


Figure 6. The ConvBlock with two layers of convolutional networks (ConvNet).

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 to learn 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 ConvBlocks (Figure 6). Feature representations of pre- and post-flooding patches from the left and right paths would then join after the *Encoding* module through concatenation along the channel dimension. Different from PSNet-v1, the other variant PSNet-v2 first concatenates the pre- and post-flooding patches and then feeds 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 architecture of PSNet-v1 are listed in Table 2. PSNet-v2 was developed with similar set of hyperparameters used for PSNet-v1.

Table 2. The hyperparameter values used for the architecture of PSNet-v1.

Module	Operation	Parameter
Input	Image patch	Size: 14×14 , # of Channels: 4
Encoding	ConvBlock 1	ConvNet A (out: 96, kernel: 3×3) + LeakyReLU (0.1) ConvNet B (out: 96, kernel: 3×3) + LeakyReLU (0.1)
	Max Pooling	Kernel: 2×2
	ConvBlock 2	ConvNet A (out: 192, kernel: 3×3) + LeakyReLU (0.1) ConvNet B (out: 192, kernel: 3×3) + LeakyReLU (0.1)
	Max Pooling	Kernel: 2×2
	ConvBlock 3	ConvNet A (out: 192, kernel: 3×3 , pad: 0) + LeakyReLU (0.1) ConvNet B (out: 192, kernel: 1×1 , pad: 0) + LeakyReLU (0.1)
	Average Pooling	Output size: 1×1
	Concatenation	Pre- and post- feature vector concatenation
Decision	Dense layer 1	Fully connection (in: 384, out: 384) + LeakyReLU (0.1)
	Dense layer 2	Fully connection (in: 384, out: 192) + LeakyReLU (0.1)
	Dense layer 3	Fully connection (in: 192, out: 1) + Sigmoid

2.4. Evaluation Metrics

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

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

where TP, FP, TN, FN denote the number of true positives, false positives, true negatives, and false negatives, respectively. For comparative analysis, we performed patch classification with baseline algorithms: 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, and 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 [20]. 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, $V = k^2C$.

To investigate how the training set size may influence the classification performance, we trained all models with different sizes of training set. Specifically, we randomly sampled various numbers of training samples from the original training subset of size 4000 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 @ 4.00 GHz \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, Figure 7a,b 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. Figure 7c,d 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 Figure 7c,d, PSNet-v2 performed with significantly higher overall accuracy and F1 score than SVM, decision tree, random forest, and AdaBoost did.

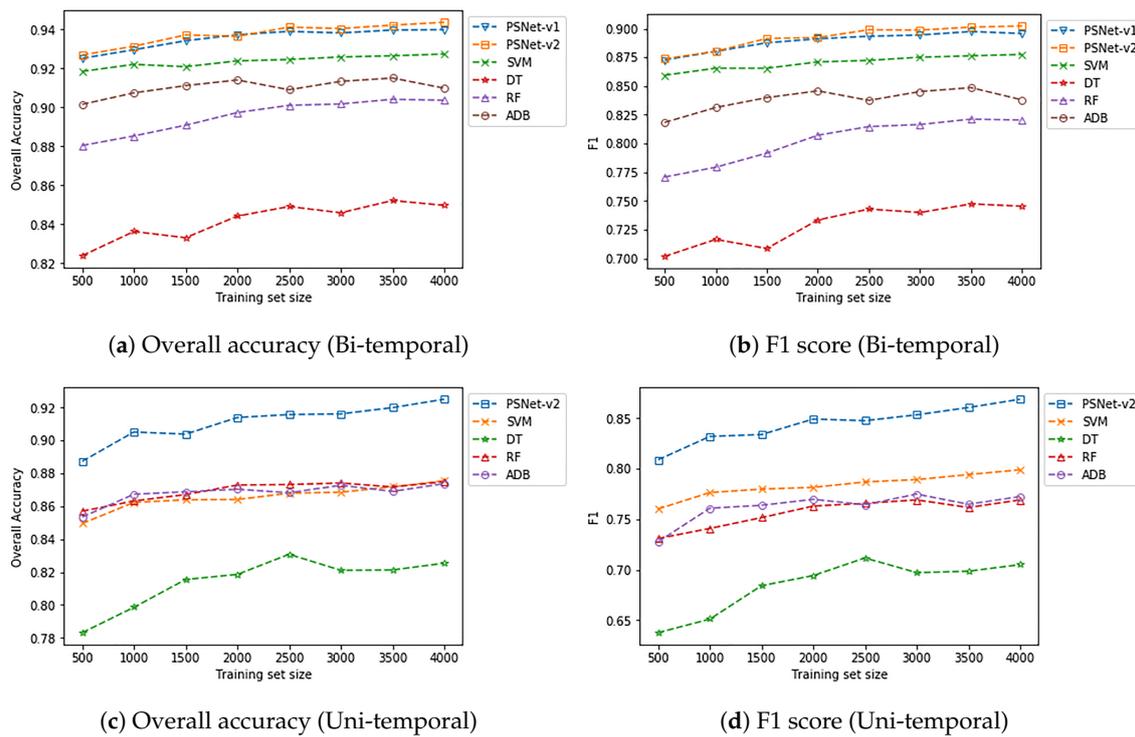


Figure 7. Classification performance on Harvey testing data in terms of OA and F1.

Taking one training set of size 1500 as an example, Table 3 summarizes 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, which was strong in precision but weak in recall), PSNet could produce consistently good results across all metrics.

Table 3. Classification performance comparison with 1500 uni- and bi-temporal Harvey training samples.

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

We also observed that, the high class imbalance and small size of training data pose a grand 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 neural networks with a large number of training data required.

For visual inspection, we show the classification maps of the entire image scene produced by the model trained with 1500 pairs of pre- and post-flooding patches for PSNet-v1 and PSNet-v2 in Figure 8. True positives of FL representing correct predictions of flooded patches are displayed in yellow color. Red patches indicate that non-flooded patches were detected as being flooded, i.e., false alarms of FL. Green patches are actually flooded patches but misclassified as NF (i.e., false negatives of FL). Qualitatively compared with the true flooding mask in Figure 3c, the proposed PSNet proved to be effective with only 1500 training samples, as demonstrated by very few false alarms (Red) and false negatives (Green).

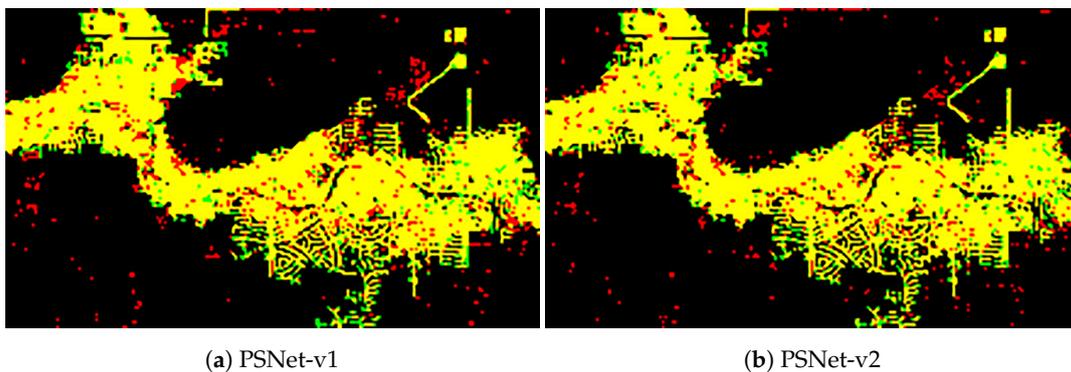


Figure 8. Classification results of Harvey data by PSNet, with patches in yellow for true positives of FL, red for false alarms of FL, and green for false negatives of FL.

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.

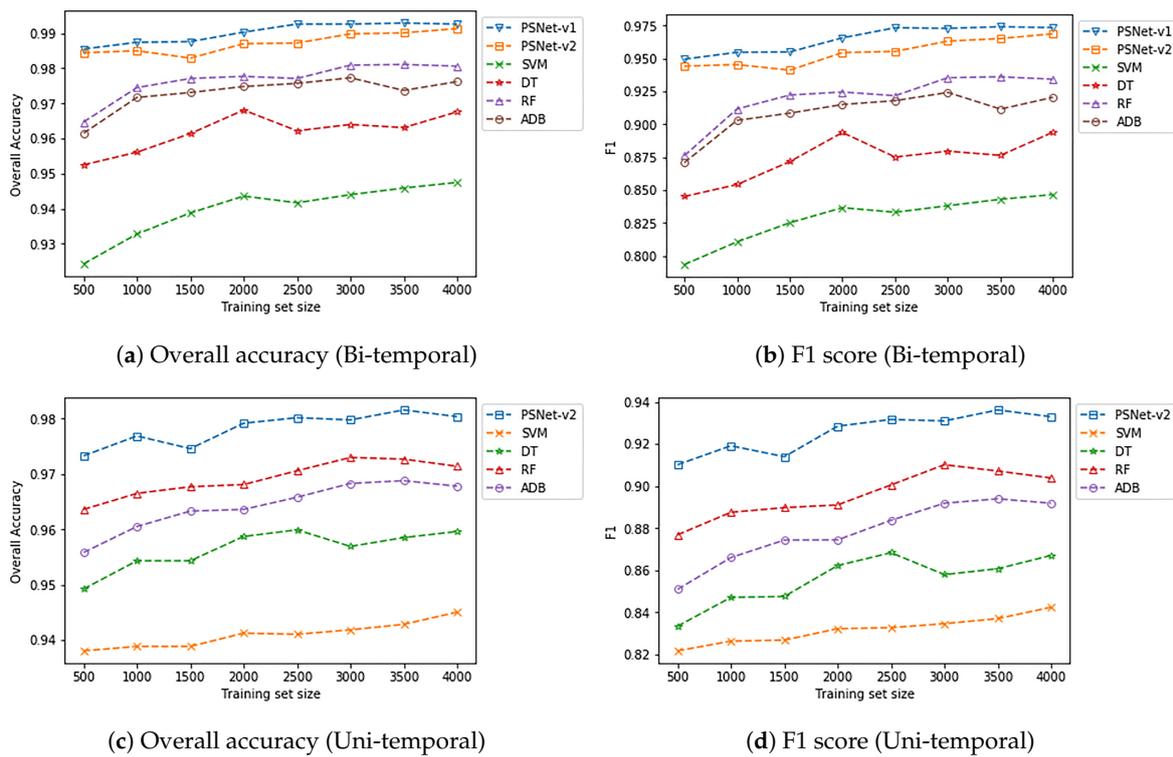


Figure 9. Classification performance on Florence testing data in terms of OA and F1.

It is obvious 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 1500 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.

Table 4. Classification performance comparison with 1500 uni- and bi-temporal Florence training samples.

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

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.

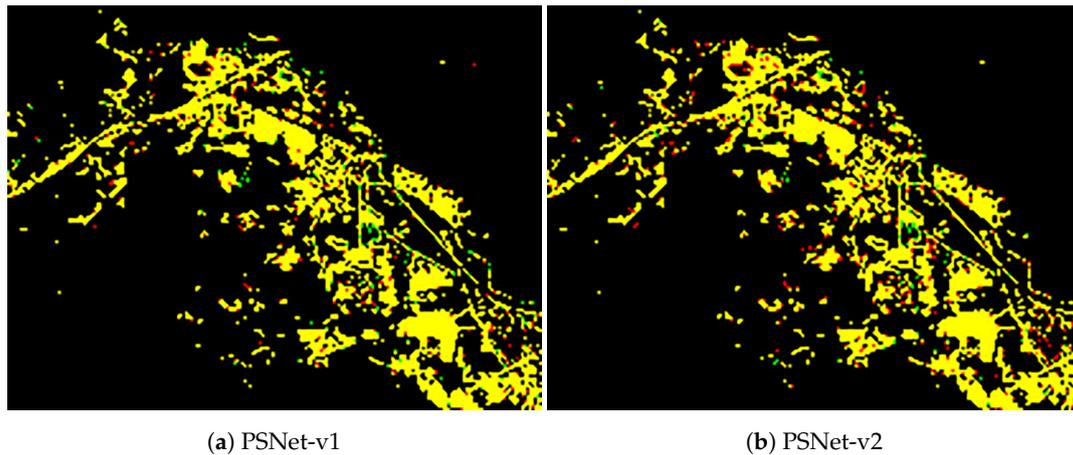


Figure 10. Classification results of Florence data by PSNet, with patches in yellow for true positives of FL, red for false alarms of FL, and green for false negatives of FL.

4. Discussions

Unlike pixel-based classification for flood mapping [6–8,12,39,40], this study investigated image patch based flood mapping similar to the study in [3]. Major motivations include: (1) reducing the impact of heterogeneous image background over urban area, which is challenging for pixel-based 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 [3,6–8,29] 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 (Figures 7 and 9 and Tables 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 [28,29]. With regard to other CNN-based patch classification models discussed in [28,29], 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, regarding patch-based urban flood mapping, this study followed the experimental settings of a recent research for urban flood mapping with SAR data [3], 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 $42 \text{ m} \times 42 \text{ m}$, which is close to the area (i.e., $40 \text{ m} \times 40 \text{ m}$) covered by patches used in [3]. 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 [3], only patches that were severely flooded were labeled as being flooded. In other words, there are fewer patches in [3] 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 poor response. Therefore, the results in this study cannot be directly and quantitatively compared with those in [3].

For qualitative comparison regarding the Harvey flooding event, as reported in [3], the developed active self-learning CNN model detected flooded 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.

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

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

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 flooded 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 image background over urban areas, patches of class FL usually have various patterns which are difficult to be learned by the classification algorithms with a very limited number of training samples. As shown in Figures 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, the proposed PSNet was able to perform with, approximately, a F1 score of 0.87 and an 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., 1500), we could identify the flooded 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 [3] and optical imagery with raw pixel DNAs [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 min to train the PSNet with 500 samples and 1 min 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, and ADB) is even less than 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. In this study, three research assistants could label 500 training samples in less than 20 min, 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 flooding 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 1500 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 flooding area may contain clouds, which is 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 [2].

5. Conclusions

This study investigated near real-time flood mapping over urban areas by leveraging patch similarity estimation instead of pixel-based classification to mitigate the impact of heterogeneous image background over urban areas, and to achieve an efficient annotation of training samples. Specifically, this work proposes the patch similarity convolutional neural network (PSNet) with two variants (PSNet-v1 and PSNet-v2) to estimate the similarity between pre- and post-flooding satellite multispectral surface reflectance image patches, and then to determine whether the post-flooding patch under test is flooded. The results show 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 1500 training samples. Comparison between PSNet and other baseline algorithms demonstrated the high performance of PSNet. Moreover, PSNet does not require the design of handcrafted floodwater related features, which further improves its generalization capability. While multispectral reflectance imagery used in this study might be influenced by severe weather conditions (e.g., heavy clouds), they are effective and accurate in urban flood mapping.

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

Author Contributions: Conceptualization, B.P., Q.H. and C.W.; Data curation, B.P. and Z.M.; Formal analysis, B.P., Q.H. and C.W.; Funding acquisition, Q.H.; Investigation, B.P., Z.M., Q.H. and C.W.; Methodology, B.P. and Q.H.; Project administration, B.P. and Q.H.; Resources, Q.H.; Software, B.P. and Z.M.; Supervision, Q.H.; Validation, B.P.; Visualization, B.P.; Writing—original draft, B.P.; and Writing—review and editing, B.P., Q.H. and C.W.

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

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

References

1. Hallegatte, S.; Green, C.; Nicholls, R.J.; Corfee-Morlot, J. Future flood losses in major coastal cities. *Nat. Clim. Chang.* **2013**, *3*, 802. [[CrossRef](#)]
2. Rudner, T.G.J.; Rufswurm, M.; Fil, J.; Pelich, R.; Bischke, B.; Kopackova, V.; Bilinski, P. Multi³Net: Segmenting Flooded Buildings via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.
3. 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 J. Photogramm. Remote Sens.* **2019**, *152*, 178–191. [[CrossRef](#)]
4. Doshi, J.; Basu, S.; Pang, G. From Satellite Imagery to Disaster Insights. In Proceedings of the 32nd Conference on Neural Information Processing Systems Workshop, Montréal, QC, Canada, 3–8 December 2018.

5. United Nations. *The Sustainable Development Goals Report*; United Nations: New York, NY, USA, 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. [[CrossRef](#)] [[PubMed](#)]
7. Xie, M.; Jiang, Z.; Sainju, A.M. Geographical hidden markov tree for flood extent mapping. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 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. [[CrossRef](#)]
9. DigitalGlobe. *Open Data Program*; DigitalGlobe: Westminster, CO, USA, 2017.
10. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
11. National Oceanic and Atmospheric Administration. *Data and Imagery from NOAA National Geodetic Survey*; National Oceanic and Atmospheric Administration: Silver Spring, MD, USA, 2017.
12. 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 J. Photogramm. Remote Sens.* **2015**, *101*, 10–21. [[CrossRef](#)]
13. Cian, F.; Marconcini, M.; Ceccato, P. Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data. *Remote Sens. Environ.* **2018**, *209*, 712–730. [[CrossRef](#)]
14. Malinowski, R.; Groom, G.; Schwanghart, W.; Heckrath, G. Detection and delineation of localized flooding from WorldView-2 multispectral data. *Remote Sens.* **2015**, *7*, 14853–14875. [[CrossRef](#)]
15. Wang, P.; Zhang, G.; Leung, H. Improving super-resolution flood inundation mapping for multispectral remote sensing image by supplying more spectral information. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 771–775. [[CrossRef](#)]
16. 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 Sens. Environ.* **2019**, *221*, 302–315. [[CrossRef](#)]
17. Giustarini, L.; Hostache, R.; Kavetski, D.; Chini, M.; Corato, G.; Schlaffer, S.; Matgen, P. Probabilistic flood mapping using synthetic aperture radar data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6958–6969. [[CrossRef](#)]
18. 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 Geosci. Remote Sens. Lett.* **2015**, *12*, 1943–1947. [[CrossRef](#)]
19. Skakun, S. A neural network approach to flood mapping using satellite imagery. *Comput. Inform.* **2012**, *29*, 1013–1024.
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
21. 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. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
22. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [[CrossRef](#)]
23. 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 Sens. Environ.* **2018**, *204*, 648–658. [[CrossRef](#)]
24. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 1–9.
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

26. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
28. Sharma, A.; Liu, X.; Yang, X.; Shi, D. A patch-based convolutional neural network for remote sensing image classification. *Neural Netw.* **2017**, *95*, 19–28. [[CrossRef](#)]
29. Song, H.; Kim, Y.; Kim, Y. A Patch-Based Light Convolutional Neural Network for Land-Cover Mapping Using Landsat-8 Images. *Remote Sens.* **2019**, *11*, 114. [[CrossRef](#)]
30. Planet Team. *Planet Application Program Interface: In Space for Life on Earth*; Planet Team: San Francisco, CA, USA, 2018.
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
32. FEMA. *Federal Emergency Management Agency Flood Mapping Products*; FEMA: Washington, DC, USA, 2019.
33. Zbontar, J.; LeCun, Y. Stereo matching by training a convolutional neural network to compare image patches. *J. Mach. Learn. Res.* **2016**, *17*, 2.
34. Manning, C.D.; Manning, C.D.; Schütze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.
35. Raghavan, V.; Bollmann, P.; Jung, G.S. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Trans. Inf. Syst.* **1989**, *7*, 205–229. [[CrossRef](#)]
36. Manning, C.; Raghavan, P.; Schütze, H. Introduction to information retrieval. *Nat. Lang. Eng.* **2010**, *16*, 100–103.
37. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
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. In Proceedings of the Thirty-First Conference on Neural Information Processing Systems (NIPS-W), Vancouver, BC, Canada, 4–9 December 2017.
39. Benoudjit, A.; Guida, R. A Novel Fully Automated Mapping of the Flood Extent on SAR Images Using a Supervised Classifier. *Remote Sens.* **2019**, *11*, 779. [[CrossRef](#)]
40. Sublime, J.; Kalinicheva, E. Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami. *Remote Sens.* **2019**, *11*, 1123. [[CrossRef](#)]
41. Zagoruyko, S.; Komodakis, N. Deep compare: A study on using convolutional neural networks to compare image patches. *Comput. Vis. Image Underst.* **2017**, *164*, 38–55. [[CrossRef](#)]
42. Zagoruyko, S.; Komodakis, N. Learning to Compare Image Patches via Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4353–4361.
43. Han, X.; Leung, T.; Jia, Y.; Sukthankar, R.; Berg, A.C. MatchNet: Unifying Feature and Metric Learning for Patch-Based Matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3279–3286.
44. Chang, C.I. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*; Springer Science & Business Media: Berlin, Germany, 2003; Volume 1.

