

Urban Flood Mapping with Residual Patch Similarity Learning

Bo Peng

bo.peng@wisc.edu

Department of Geography
University of Wisconsin-Madison
Madison, Wisconsin

Zonglin Meng

zmeng32@wisc.edu

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

Xinyi Liu

xliu636@wisc.edu

Department of Geography
University of Wisconsin-Madison
Madison, Wisconsin

Qunying Huang*

qh Huang46@wisc.edu

Department of Geography
University of Wisconsin-Madison
Madison, Wisconsin

ABSTRACT

Urban flood mapping is essential for disaster rescue and relief missions, reconstruction efforts, and financial loss evaluation. Much progress has been made to map the extent of flooding with multi-source remote sensing imagery and pattern recognition algorithms. 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 often has heterogeneous background involving various ground objects (e.g., vehicles, buildings, roads, and trees), making traditional classification algorithms 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 residual patch similarity convolutional neural network (ResPSNet) to map urban flood hazard zones using bi-temporal high resolution (3m) pre- and post-flooding multispectral surface reflectance satellite imagery. Besides, remote sensing specific data augmentation was also developed to remove the impact of varying illuminations due to different data acquisition conditions, which in turn further improves the performance of the proposed model. Experiments using the high resolution imagery before and after the 2017 Hurricane Harvey flood in Houston, Texas, showed that the developed ResPSNet model, along with associated remote sensing specific data augmentation method, can robustly produce flood maps over urban areas with high precision (0.9002), recall (0.9302), F1 score (0.9128),

and overall accuracy (0.9497). The research sheds light on multi-temporal image fusion for high precision image change detection, which in turn can be used for monitoring natural hazards.

CCS CONCEPTS

• **Information systems** → *Geographic information systems*; • **Computing methodologies** → **Computer vision**; **Machine learning**; • **Applied computing** → *Earth and atmospheric sciences*.

KEYWORDS

Flood mapping, patch similarity, residual learning, deep learning, flood extent estimation

ACM Reference Format:

Bo Peng, Xinyi Liu, Zonglin Meng, and Qunying Huang. 2019. Urban Flood Mapping with Residual Patch Similarity Learning. In *3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI'19)*, November 5, 2019, Chicago, IL, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3356471.3365235>

1 INTRODUCTION

Natural disasters pose a great 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 [3, 14, 22]. The estimated global financial losses, with projected socio-economic change alone, will increase close to \$52 billion by 2050 [9]. 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 [24]. Therefore, an urgent call exists to map flood areas in near real-time for aiding in 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 inspection [3, 5, 7, 26]. Multispectral surface reflectance imagery often contains important spectral information for flood-water detection [1, 13, 16, 22, 25]. By virtue of the radar's active imaging property and its long-wavelength signal with penetration

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GeoAI'19, November 5, 2019, Chicago, IL, USA

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6957-2/19/11...\$15.00

<https://doi.org/10.1145/3356471.3365235>

power, the space-borne synthetic aperture radar (SAR) is able to collect data over the flooded regions day and night regardless of weather conditions [8, 14, 22, 23]. Based on different types of data, various frameworks have been developed for mapping the flood extent. Feng [5] mapped flood inundation areas based on VHR aerial optical images acquired by a mini unmanned aerial vehicle (UAV) using a Random Forest classifier performed in the spectral-texture feature space. Li [13] 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). Giustarini [8] introduced a statistical model to estimate the uncertainty in flood mapping delineation on satellite SAR images. Rudner [22] fused multisource satellite imagery, including VHR, multispectral, and radar data, in a convolutional neural network to detect flooded buildings.

However, both disaster relief management organizations (e.g., Federal Emergency Management Agency (FEMA)) and private sectors (e.g., insurance companies), have posed further demands on urban flood mapping with high resolution at the scale of street block level. In fact, this still remains a major challenge due to three constraints: (1) First, the VHR optical remote sensing imagery often has heterogeneous background involving various ground objects (e.g., vehicles, buildings, roads, and trees) [26]. Such heterogeneous background prohibitively prevents conventional pixel-based classifiers, such as random forest (RF), support vector machine (SVM), and maximum likelihood (ML), from capturing the underlying spatial correlation between neighboring pixels within the flood hazard area, and therefore from performing well. For example, a floodwater pixel can hardly be identified without considering its neighboring pixels; (2) Second, with the advancement of data acquisition technologies, current remotely sensed data are often in a huge volume, such as satellite images from DigitalGlobe [2] and Planet Labs [20], and aerial imagery collected by National Oceanic and Atmospheric Administration (NOAA [18]). With an increasing data volume, how to fully leverage the big data and to develop more robust and scalable algorithms for flood extent mapping are still not well addressed; and (3) Finally, due to inconsistent illumination conditions at the time of data acquisition, pixels of the same objects in VHR 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.

Motivated by these key issues, this paper proposed a residual patch similarity convolutional neural network (ResPSNet) 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. Besides, remote sensing specific data augmentation was also developed to remove the impact of varying illuminations due to different data acquisition conditions, which in turn further improves the performance of the proposed ResPSNet model. Experiments comparing with the uni-temporal (i.e., with only post-flooding data) convolutional neural network (CNN) and support vector machine



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

(SVM), using the flooding event caused by the 2017 Hurricane Harvey in Houston, Texas, as a case study, showed that the developed ResPSNet model with associated data augmentation method in this study can robustly produce flood maps over urban areas with high precision, recall, F1 score, and overall accuracy.

In summary, major contributions of this study include:

- The ResPSNet was developed that 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-temporal CNN and SVM, the proposed model achieved superior performance in all evaluation metrics.
- 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 and the appended water related spectral index NDWI reduced the impact of irrelevant changes caused by data acquisition conditions (e.g., illumination).
- Remote sensing specific data augmentation was developed, which further improves the performance of ResPSNet by enhancing the sensitivity of the network to floodwaters in the post-flooding imagery.
- 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 METHODOLOGY

2.1 Preliminaries

Flood mapping is a task to extract the flooding zones (i.e., land areas submerged in floodwaters), which may have different definitions in the literature [5, 7, 14, 26]. For example, FEMA's National Flood Mapping Products [4], define flood hazard zones as land areas

that are either covered or surrounded by floodwaters. 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 [26]. Since



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

we cannot label and validate these invisible flooding areas, this paper uses FEMA’s definition of floodwaters as previous work [14].

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 as input of a mapping model M , the goal of the work is to develop a classification model that can output a binary flood hazard map O , $O = M(I_1, I_2)$. Each pixel in the output O was assigned with either 1 (flood, FL) or 0 (non-flood, NF).

To develop the classification model M for incorporating bi-temporal data for flood mapping, there are three major issues: (1) First, although pre- and post-flooding imagery are co-registered, corresponding pixels, $I_1(i, j)$ and $I_2(i, j)$, at the same geographical location, may not exactly point to the same ground object. This is inevitable especially for time-series imagery collected at different seasons when trees grow differently, and over urban areas where moving objects are common; (2) Second, satellite imagery may not be perfectly orthorectified due to various terrains and inaccurate geometric corrections; and (3) Third, pixel-wise dense classification, also known as semantic segmentation, requires high-resolution

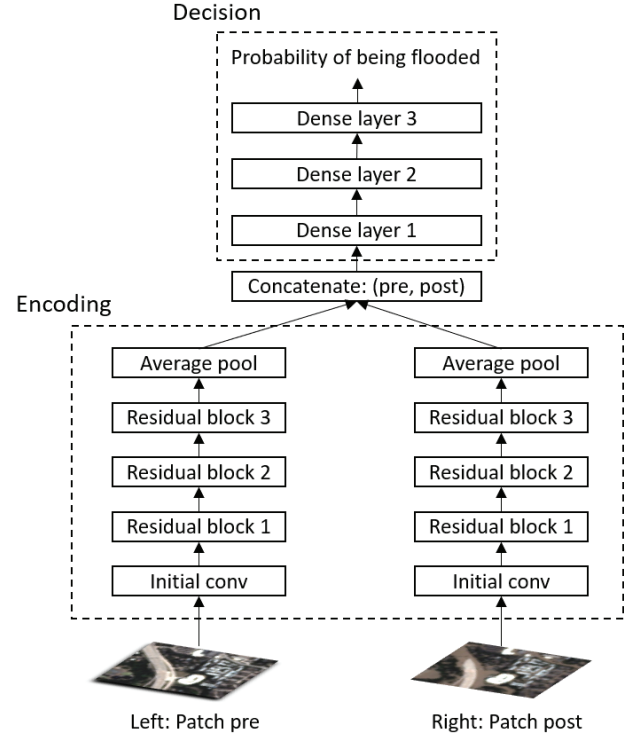


Figure 3: The Siamese residual patch similarity convolutional neural network (ResPSNet).

flooding masks corresponding to the study area for model training, resulting in expensive human annotation. As such, we did not apply time-series pixel-wise dense classification or semantic segmentation to urban flood mapping.

2.2 Model development

Instead of using pixel-wise classification, this study splits the raw bi-temporal imagery I_1 and I_2 , into non-overlapping patches of the same size [14]. It is worth noting that we changed pixel classification into patch classification, resulting in a lower resolution flooding map. Nevertheless, this is a more efficient approach to creating flood maps over heterogeneous urban area as discussed in [14], which are also in consistent with flood maps from FEMA. Leveraging the spatial context within the patches, we proposed the Siamese ResPSNet to learn the nonlinear mapping from the pre- and post-flooding patches to one of the two predefined classes (i.e., either FL or NF). The corresponding pair of patches (P_{ij}^1, P_{ij}^2) from the pre- and post-flooding imagery (I_1, I_2) were the inputs of the Siamese network.

The architecture of the ResPSNet is shown in Figure 3. It consists of two major modules. First, **Encoding** on the left and right paths extracts features from pre- and post-flooding image patches using a residual CNN. The residual block used in this study is shown in Figure 4b. Note that the rectified linear unit (ReLU) activation is not used after adding the identity mapping information [15]. Inspired

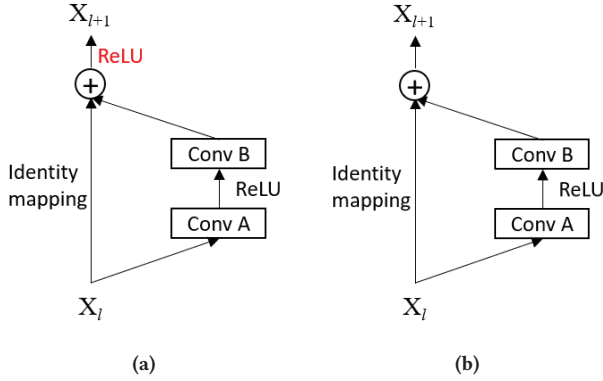


Figure 4: Different residual convolution blocks (a) with ReLU activation [11], and (b) without ReLU activation [15] (used in this study), after addition on top.

by the same hidden nodes in recurrent neural networks (RNN) for time series data processing, the left and right sub-networks shared the same set of weights, and joined at the top via feature vector concatenation as designed in [27].

Second, **Decision** consists of three dense layers taking the concatenated feature vectors as the input. The probability of being flooded was then computed via fully connection operations. Since the pre- and post-flooding images were collected recently before and shortly after the flooding event, respectively, the major changes between the pair of patches were resulted from flooding. The flooding probability can be interpreted as the inverse of the similarity score of the two input patches. This means that the higher probability of being flooded, the less similar of the two patches.

It is worth noting that the similarity of the two patches was mainly determined by the change of spectral reflectance instead of spatial structures of the ground objects. To keep the spectral contrasts between floodwaters in the post-flooding patch and the ground objects (e.g., roads and grass) in the pre-flooding patch, we did not use batch normalization throughout the entire network.

All hyperparameters of the ResPSNet after tuning are listed in Table 1.

In addition to the ResPSNet, we also developed a non-residual version of Siamese patch similarity convolutional neural network (PSNet) for urban flood mapping. The left and right sub-networks contain a set of convolution operations without residual connection. The network architecture was designed similar to ResPSNet as shown in Figure 5, where the Conv block is designed as Figure 6. Hyperparameters were set based on the CNN architecture used in [14], as summarized in Table 2.

2.3 Remote sensing specific data augmentation

For remote sensing imagery with surface spectral reflectance information, as demonstrated in [17], normalized class-specific radiometric indices obtained from spectral reflectance are invariant with respect to the illumination at different data acquisition conditions. Such invariance enables the neural network to generalize well on testing data which are not seen during the training process.

Layers	Parameters
Input patch	$C@14 \times 14$, $C: n_channels$
Initial conv	64, 3×3 , stride=1, pad=1, ReLU
Residual block 1	Conv A - 64, 3×3 , stride=1, pad=1, ReLU Conv B - 64, 3×3 , stride=1, pad=1
Residual block 2	Conv A - 128, 3×3 , stride=2, pad=1, ReLU Conv B - 128, 3×3 , stride=1, pad=1
Residual block 3	Conv A - 256, 3×3 , stride=2, pad=1, ReLU Conv B - 256, 3×3 , stride=1, pad=1
Pooling	Adaptive average pooling, (1×1)
Concatenation	Feature vector concatenation, pre + post
Dense layer 1	Fully connection, $512 \rightarrow 512$, ReLU
Dense layer 2	Fully connection, $512 \rightarrow 256$, ReLU
Dense layer 3	Fully connection, $256 \rightarrow 1$, Sigmoid

Table 1: Hyperparameters of the ResPSNet

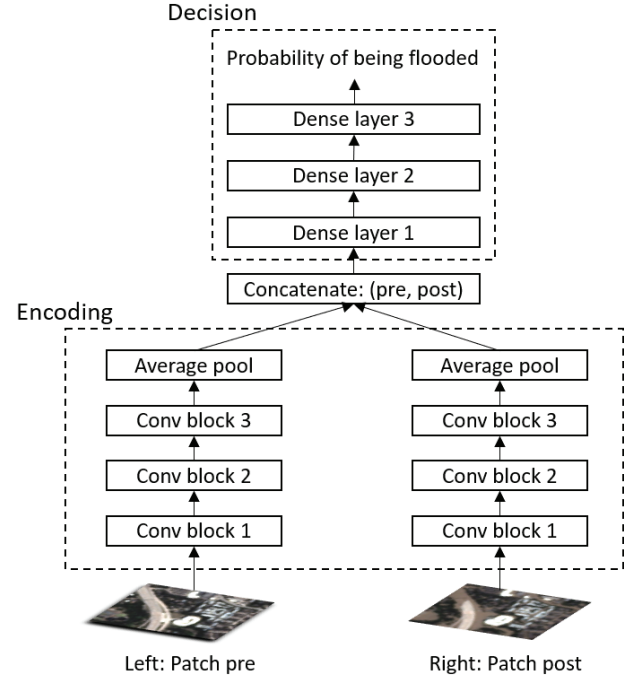


Figure 5: The non-residual version of Siamese patch similarity convolutional neural network (PSNet).

Therefore, to train a model with the capability of being sensitive to floodwaters, we augment the input patches by concatenating another feature map, the normalized difference water index (NDWI) [6], defined as

$$NDWI = \frac{G - NIR}{G + NIR} \quad (1)$$

where G and NIR represent the spectral reflectance bands of green and near infrared in the input patches respectively. The resulting NDWI value is within the range of $(-1, 1)$. For consistent inputs of

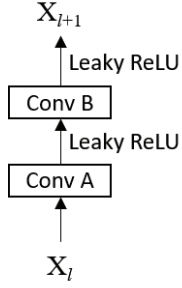


Figure 6: The convolution block used in PSNet

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

the network, we rescaled the NDWI value into (0, 1), the same as the range of spectral reflectance (0, 1).

2.4 Model training

During the training process, we randomly selected corresponding pre- and post-flooding patches (P_{ij}^1, P_{ij}^2) along with the label (FL or NF) as the input of the network. Adam optimizer [12] with batch size of 16 was used for training. The loss function used in this study is the weighted binary cross entropy loss, which is defined in Equation 2.

$$L(x, y) = \frac{1}{N} \sum_{i=1}^N l_i \quad (2)$$

$$l_i = -w_i [y_i \log x_i + (1 - y_i) \log (1 - x_i)]$$

where N is the total number of samples in a batch, w_i denotes the loss weight associated with the i^{th} sample, and $x = (x_1, \dots, x_N)$, $y = (y_1, \dots, y_N)$ represent the output of the network and the corresponding true labels (i.e., 1 for FL and 0 for NF), respectively. Considering the high class imbalance of our training data, the loss weight w_i for the two classes were set as the complementary of the occurrence frequency r_c of class c , where $c = \text{FL or NF}$. More specifically, as r_{FL} and r_{NF} are equal to 0.2978 and 0.7022 in our training set, class FL and NF were assigned with weights $1 - r_{\text{FL}} = 0.7022$

and $1 - r_{\text{NF}} = 0.2978$ respectively. We used a weight decay of $1e-5$ and momentum parameters $(\beta_1, \beta_2) = (0.9, 0.999)$. The training process started with the learning rate of $1e-4$ and then divided by 10 after the validation loss plateaued, and the model was trained for up to 300 epochs.

Further data augmentation was used to reduce the model generalization error on validation and testing data. Each pair of input patches were augmented with probability of 0.5, including horizontal and vertical flipping and rotation of $[0, 90, 180, 270]$ degrees. Then, before fed into the network, the reflectance of each pixel in the input pre- and post-flooding patches was normalized to $[0, 1]$.

Neural networks with multiple paths and layers usually require appropriate initialization of the model weights before training, since it helps avoid partial node activation [21]. For our network, we sampled the initial model weights from Gaussian distribution with zero mean and standard deviation of $\sqrt{V/2}$, where V is the number of parameters in each operation [10]. For example, in a $k \times k$ convolution with C incoming channels, $V = k^2 C$.

3 EXPERIMENTS AND RESULTS

3.1 Datasets

We tested our model on datasets collected over Houston, Texas, where dense residential, commercial, and industrial areas were severely flooded in August 2017 due to the Hurricane Harvey. The datasets contain two orthorectified multispectral (i.e., blue (B), green (G), red (R), and near infrared (NIR)) surface reflectance satellite imagery provided by Planet Lab [20]: 1) pre-flooding image on July 31, 2017, and 2) post-flooding image on August 31, 2017. Data specifications are summarized in Table 3.

Scene	Bands	Size (h, w)	Pixel size	Product
Pre	B, G, R, NIR	(1848, 3066)	3m	Reflectance
Post	B, G, R, NIR	(1848, 3066)	3m	Reflectance

Table 3: Planet Lab data characteristics

Both pre- and post-flooding images were co-registered and split into non-overlapping patches of the same 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 [14], 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 with the ones in [14]. 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. Three research assistants participated in the annotation of the dataset individually. The class of the post-flooding multispectral patch (14×14) was determined by visual inspection of the corresponding VHR patch (140×140). With regard to patches

annotated differently, the corresponding labels were finally assigned with agreement among all annotators.

As discussed in [14], patches with invisible floodwaters under trees were not considered as being flooded. In summary, we have a total of 28,908 labeled patches with 8,517 in class FL and 20,391 in class NF. As a result, the numbers of patches in different classes are highly imbalanced. The pre- and post-flooding images with the labeled ground truth over the whole study area are shown in Figure 7.

For model training and evaluation, we randomly sampled training, validation, and testing subsets of patch pairs without replacement with ratios of 40%, 10%, and 50%. Therefore, there are 11,563 patch pairs for training, 2,891 for validation, and 14,454 for testing.

3.2 Results and discussion

During the training, we evaluated the overall accuracy (OA), precision, recall, and F1 score as in [14, 26], and selected trained models with highest validation F1 scores for testing. For comparative analysis, we also trained a uni-temporal CNN model and SVM for post-flooding patch classification. To be more specific, the uni-temporal CNN used the same encoding architecture as one branch of the PSNet with hyperparameters in Table 2. Regarding the implementation of SVM, we used default settings in the scikit-learn package [19], in which the radial basis kernel was used and the class weight was set to be balanced considering the class imbalance. It should be noted that the baseline models (i.e., CNN and SVM) used only post-flooding data for training and testing. Therefore, we could evaluate whether bi-temporal data based models (i.e., ResPSNet and PSNet) would outperform uni-temporal ones (i.e., CNN and SVM). Additionally, since the CNN has the same architecture and hyperparameters as one branch of the PSNet, we could compare the classification performance between traditional machine learning algorithm (i.e., SVM) and the popular deep learning model (i.e., CNN).

Table 4 lists the binary classification results of the testing data. It shows that the ResPSNet and PSNet consistently outperformed the uni-temporal CNN and SVM across all evaluation metrics. Due to limited size and high class imbalance of training data, it remains a challenge for the uni-temporal model (i.e., CNN and SVM) to learn the flood relevant feature representations of the input patch. However, leveraging the bi-temporal information from pre- and post-flooding data, patch similarity is an important *a priori* for binary classification. In this study, we do not need to learn very deep architectures, which usually require a large number of training data. As a result, ResPSNet and PSNet can generalize well on testing data.

Moreover, ResPSNet performed slightly better than PSNet as demonstrated in Table 4 with best results in bold. With respect to ResPSNet, as we augmented the input by appending another feature map NDWI to the original input feature maps, the precision and overall accuracy were further improved. Similar improvements were also observed for PSNet with augmented NDWI, where the precision, F1 score, and overall accuracy were all improved slightly as a result of additional NDWI, which is sensitive to floodwaters. However, as improvements associated with NDWI were not very significant, it remains to be addressed if NDWI really helps boost

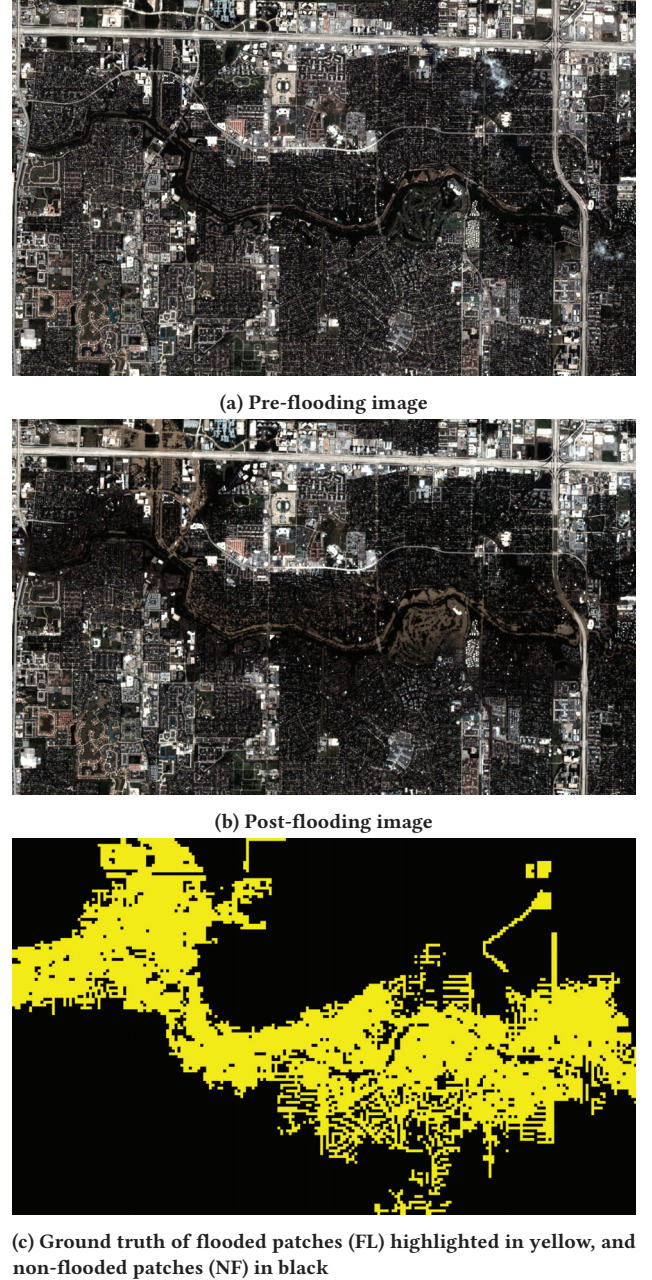


Figure 7: Optical view of pre- and post-flooding multispectral images with ground truth of flooded patches.

the performance of the proposed model by reducing the impact of varying illumination.

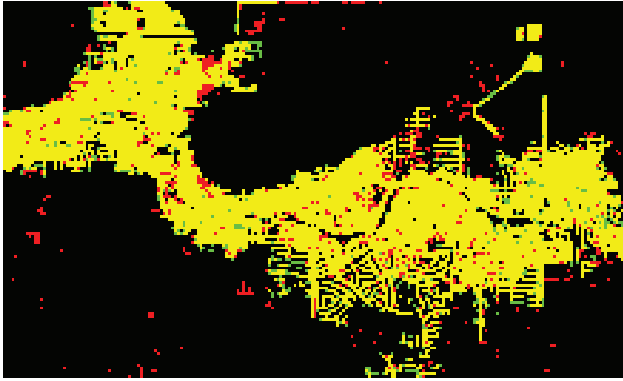
For visual inspection, the classification maps of the entire image scene over the study area are shown in Figure 8 for ResPSNet, and Figure 9 for PSNet. Patches in yellow represents the true positives of FL, indicating the correct predictions of flooded patches. Patches in red represent the false alarms of FL, which means non-flooded

Models	NDWI	Precision	Recall	F1	OA
ResPSNet	Yes	0.9002	0.9302	0.9128	0.9497
	No	0.8902	0.9413	0.9131	0.9489
PSNet	Yes	0.8984	0.9222	0.9079	0.9468
	No	0.8753	0.9445	0.9064	0.9443
CNN	Yes	0.8802	0.8789	0.8766	0.9296
	No	0.9044	0.8646	0.8812	0.9338
SVM	Yes	0.7573	0.8609	0.8058	0.8786
	No	0.7898	0.8606	0.8237	0.8923

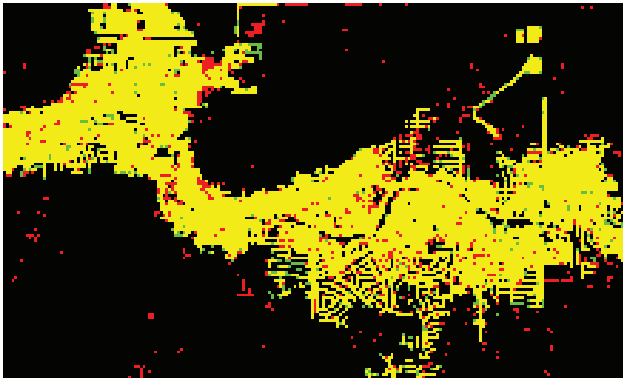
Table 4: Classification results of testing data.

patches were detected as being flooded. Patches in green show false negatives of FL, i.e., flooded patches were classified as being non-flooded.

Compared with ground truth shown in Figure 7c qualitatively, the proposed ResPSNet and PSNet could produce reasonable classification maps. Only a very small number of false alarms and false negatives were reported by ResPSNet and PSNet as demonstrated in Figure 8 and Figure 9 with very few patches in red and green.

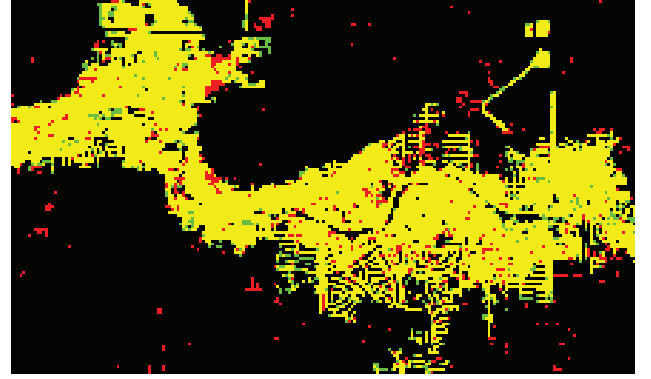


(a) ResPSNet with NDWI

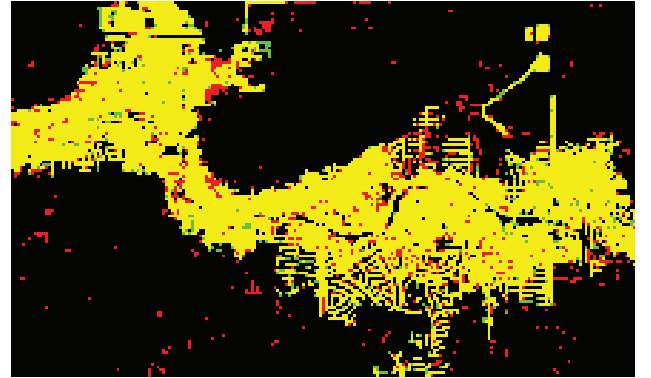


(b) ResPSNet without NDWI

Figure 8: Classification results by ResPSNet over the study area, with patches in yellow for true positives of FL, red for false alarms of FL, and green for false negatives of FL.



(a) PSNet with NDWI



(b) PSNet without NDWI

Figure 9: Classification results by PSNet over the study area, with patches in yellow for true positives of FL, red for false alarms of FL, and green for false negatives of FL.

4 CONCLUSION AND FUTURE WORKS

This paper addressed the challenge of urban flood mapping via patch similarity learning. We proposed the Siamese patch similarity convolutional neural network (ResPSNet) and its non-residual variant (PSNet) 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 or not. To further enhance the network's response to floodwaters, we applied remote sensing specific data augmentation by appending another feature map (i.e., NDWI layer) to the original input patches. Results showed that both ResPSNet and PSNet models developed in this study achieved high performance with approximately 90% accuracy on testing data in all evaluation metrics including overall accuracy, precision, recall, and F1 score.

In the future, we would experiment with data for other flooding events (e.g., the 2018 Hurricane Florence flood over the city of Lumberton, North Carolina) to further test the model generalization ability.

5 ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the National Science Foundation with project ID: 1940091.

REFERENCES

- [1] Fabio Cian, Mattia Marconcini, and Pietro Ceccato. 2018. Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data. *Remote Sensing of Environment* 209 (may 2018), 712–730. <https://doi.org/10.1016/j.rse.2018.03.006>
- [2] DigitalGlobe. [n.d.]. Open Data Program. <https://www.digitalglobe.com/ecosystem/open-data>
- [3] Jigar Doshi, Saikat Basu, and Guan Pang. 2018. From Satellite Imagery to Disaster Insights. In *32nd Conference on Neural Information Processing Systems Workshop*.
- [4] FEMA. [n.d.]. Federal Emergency Management Agency Flood Mapping Products. <https://www.fema.gov/flood-mapping-products>
- [5] Quanlong Feng, Jiantao Liu, and Jianhua Gong. 2015. Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier A Case of Yuyao, China. *Water* 7, 12 (mar 2015), 1437–1455. <https://doi.org/10.3390/w7041437>
- [6] Bo-Cai Gao. 1996. NDWI normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment* 58, 3 (1996), 257–266.
- [7] Asmamaw Gebrehiwot, Leila Hashemi-Beni, Gary Thompson, Parisa Kordjamshidi, and Thomas Langan. 2019. Deep Convolutional Neural Network for Flood Extent Mapping Using Unmanned Aerial Vehicles Data. *Sensors* 19, 7 (mar 2019), 1486. <https://doi.org/10.3390/s19071486>
- [8] Laura Giustarini, Renaud Hostache, Dmitri Kavetski, Marco Chini, Giovanni Corato, Stefan Schlaffer, and Patrick Matgen. 2016. Probabilistic flood mapping using synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing* 54, 12 (2016), 6958–6969.
- [9] Stephane Hallegatte, Colin Green, Robert J Nicholls, and Jan Corfee-Morlot. 2013. Future flood losses in major coastal cities. *Nature climate change* 3, 9 (2013), 802.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [12] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. <http://arxiv.org/abs/1412.6980> cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
- [13] Linyi Li, Yun Chen, Xin Yu, Rui Liu, and Chang Huang. 2015. Sub-pixel flood inundation mapping from multispectral remotely sensed images based on discrete particle swarm optimization. *ISPRS Journal of Photogrammetry and Remote Sensing* 101 (mar 2015), 10–21. <https://doi.org/10.1016/j.isprsjprs.2014.11.006>
- [14] Yu Li, Sandro Martinis, and Marc Wieland. 2019. 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* 152 (jun 2019), 178–191. <https://doi.org/10.1016/j.isprsjprs.2019.04.014>
- [15] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 136–144.
- [16] Radosław Malinowski, Geoff Groom, Wolfgang Schwanghart, and Goswin Heckrath. 2015. Detection and delineation of localized flooding from WorldView-2 multispectral data. *Remote sensing* 7, 11 (2015), 14853–14875.
- [17] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. 2016. Pansharpening by convolutional neural networks. 8, 7 (2016), 594.
- [18] National Oceanic and Atmospheric Administration. [n.d.]. Data and Imagery from NOAA National Geodetic Survey. <https://www.ngs.noaa.gov/>
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [20] Planet Team. 2018. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. <https://api.planet.com>
- [21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Nassir Navab, Joachim Hornegger, William M Wells, and Alejandro F Frangi (Eds.). Springer International Publishing, Cham, 234–241.
- [22] Tim G. J. Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, Veronika Kopackova, and Piotr Bilinski. 2019. Multi³Net: Segmenting Flooded Buildings via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery. <https://arxiv.org/abs/1812.01756>
- [23] Xinyi Shen, Emmanouil N Anagnostou, George H Allen, G Robert Brakenridge, and Albert J Kettner. 2019. Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar. *Remote Sensing of Environment* 221 (feb 2019), 302–315. <https://doi.org/10.1016/j.rse.2018.11.008>
- [24] United Nations. 2018. The Sustainable Development Goals Report. <https://www.un.org/sustainabledevelopment/cities>
- [25] Peng Wang, Gong Zhang, and Henry Leung. 2018. Improving super-resolution flood inundation mapping for multispectral remote sensing image by supplying more spectral information. *IEEE Geoscience and Remote Sensing Letters* (2018).
- [26] Miao Xie, Zhe Jiang, and Arpan Man Sainju. 2018. Geographical hidden markov tree for flood extent mapping. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2545–2554.
- [27] Jure Zbontar and Yann LeCun. 2016. Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research* 17, 1-32 (2016), 2.