DeepFlood: A deep learning based flood detection framework using feature-level fusion of multi-sensor remote sensing images

A. Emily Jenifer
(School of Computing, SASTRA University, Thanjavur, Tamilnadu, India
https://orcid.org/0000-0002-5469-1412, emilyjenifer@sastra.ac.in)

Sudha Natarajan
(School of Computing, SASTRA University, Thanjavur, Tamilnadu, India
https://orcid.org/0000-0003-1729-6699, sudha@cse.sastra.edu)

Abstract: Flooding is the most common natural disaster in many countries. Remote sensing images are very much useful in disaster monitoring. The different image modalities from different satellites provide varied information about the earth. The synergistic use of optical and radar data helps in precise flood detection. The central focus of this paper is to identify the flooded regions using a dual patch-based Fully Convolutional Network (FCN) for performing deep learning-based feature fusion. The learned features of FCNs trained independently with Synthetic Aperture Radar (SAR) and Multispectral (MS) images are concatenated to represent the flooding better. A random forest classifier is employed to identify the flood from the fused features. The information retrieved is very much valuable in undertaking necessary rescue efforts in flood-affected areas. The proposed network shows superior performance in flood detection on the images from the SEN12-FLOOD dataset with an accuracy as high as 94.17%.

Keywords: flood detection, synthetic aperture radar, multispectral, convolutional neural network, deep learning.

1 Introduction

Flood is one of the natural disasters that affects the humanity to a greater extent. Both tropical and temperate regions are most commonly affected by floods. They damage crops and property, and also cause loss of human life [Brivio et al. 2002]. Therefore, it is necessary to obtain immediate information of the flood-affected area. Remote sensing has become a very suitable technique to detect flooding without having any direct contact with the land area. The open-access availability of multi-temporal and multi-sensor data makes remote sensing efficient in flood monitoring [Chawan et al. 2020]. The active sensors like (e.g. Sentinel-1) operating in the microwave region and the passive sensors (e.g. Sentinel-2) operating in the infrared and visible regions of the electromagnetic spectrum provide all important information of flood-affected areas [Sanyal and Lu 2004].
The development of satellites with different types of sensors makes it possible to collect multi-modal data in real-time. Thus, the shortfall of single sensor information could be supplemented with another sensor’s information [Seo et al. 2018]. Data fusion is an effective technique that combines data from multiple sources to generate high-quality information [Zhang 2010].

Initially, flood detection techniques were limited to aerial images. With the advent of satellite technology, SAR and optical images came into existence [Nazir et al. 2014]. Sentinel-1 (S1) satellite is equipped with SAR active sensor operating in C-band with 5-20m resolution [Geudtner et al. 2014]. It acquires the dual-polarized (Vertical-Vertical (VV) and Vertical-Horizontal (VH)) images of earth with the help of microwaves. SAR is a reliable data acquisition technique during monsoon and high cloud cover because it penetrates cloud and dust [Ban et al. 2010]. The signal depends upon the property of the incoming wave, roughness and dielectric property of the earth surface [Woodhouse 2017]. Smooth surface like water tend to exhibit specular reflection, thereby appears darker in the produced SAR image [Landuyt et al. 2020]. The Multispectral (MS) images consist of various bands with different wavelengths. The Sentinel-2 (S2) satellite is a medium to high resolution (10-60 m) satellite that acquires MS images with 11 spectral bands [Notti et al. 2018]. The natural colour band combination is red (band 4), green (band 3) and blue (band 2). This band combination gives the equivalent images of how our eyes see. Thus water appears in a shade of blue [Du et al. 2016]. The European Space Agency (ESA) has launched these S1 and S2 satellites, which has good global coverage for remotely sensed images [Bangira et al. 2019]. Both S1 and S2 sensor images are freely available from geo data portals like United States Geological Survey (USGS) Earth Explorer [USGS] and Alaska Satellite Facility Distributed Active Archive Centre (ASF DAAC) [ASF].

As far as flood detection is concerned, methods such as visual interpretation [Chambenoit et al. 2003], segmentation using fuzzy logic [Giordano et al. 2005], thresholding [Moser and Serpico 2006], chromatic and textural analysis [Zhao et al. 2011] and Normalized Difference Water Index (NDWI) [Soltanian et al. 2019] were used. These traditional methods have limitations in performance for detecting the flood in complex environment [Hu et al. 2020]. Machine learning (ML) algorithms helped to overcome this weakness. Algorithms such as K-Nearest Neighbour (KNN) [Shahabi et al. 2020], artificial neural network (ANN) [Kia et al. 2021], logistic regression [Tien Bui et al. 2019, Pradhan 2010] and decision tree [Chen et al. 2020a] were applied. However, the ML methods are time-consuming and feature-dependent [Mosavi et al. 2020]. Therefore with the emergence of deep learning (DL) technology, the abundant remote sensing data can be processed to learn the task-relevant features efficiently [Zhang et al. 2016].

In this paper, we design a dual patch Fully Convolutional Network (FCN) architecture for processing the SAR and multispectral images separately. A feature-level fusion followed by a random forest classifier is applied to detect the flood more accurately. The entire methodology is called DeepFlood. The main contributions in this paper are

I. Fully convolutional neural networks are designed for extracting the features from the patches of SAR and multispectral images.

II. Feature-level fusion is applied and the flood is detected from the fused features using a random forest classifier.

III. The proposed DeepFlood architecture is evaluated using bi-temporal images from the SEN12-FLOOD dataset.
2 Related work in flood detection

As deep learning and its fusion frameworks have gained importance in recent years [Muñoz et al. 2021], it has been applied to remote sensing extensively. Various SAR and MS-based deep learning approaches for flood detection are discussed in this section.

2.1 SAR-based methods

Due to specular reflection of SAR signals, water surface appears dark in the radar data [Martinis and Ricke 2015]. This property of SAR facilitates in identifying the flooded region. The authors in [Bonafilia et al. 2020] presented a SAR-based flood detection dataset, namely SEN1Floods11, to train and test deep learning models. A fully convolutional neural network was employed to map the flood effectively. In another work [Katiyar et al. 2021], segmentation architectures such as SegNet and UNet were used to improve flood mapping. The authors applied transfer learning for precisely detecting the flooded regions in Kerala, India. Mapping the flood in open area is easier than urban flood mapping due to complex scattering behaviour of urban structures. The work in [Li et al. 2019] introduced an active self-learning convolution neural network framework to map flooding in Houston, USA after a hurricane. Flood mapping with auxiliary hydrological data such as digital elevation model and rainfall information from the meteorology department using VGG16 deep learning network was performed in [Kang et al. 2018].

Flood occurrence is associated with large cloud cover. SAR is capable of penetrating clouds and it is suitable for acquisition in any weather conditions [Dwivedi et al. 2000]. However, the multiplicative noise in SAR makes it difficult to interpret the data effectively [Yu et al. 2018]. This is overcome by fusing images from other sensors such as Sentinel-2.

2.2 Multispectral-based methods

Among all bands of multispectral data, high spatial resolution bands (such as bands 3 to 15) provide supplementary information in flood mapping. These MS band images along with SAR data were used for enhanced flood mapping [Quan et al. 2020]. In [Peng et al. 2019], a patch similarity convolutional neural network was developed for urban flood mapping with spectral reflectance as input to the network. The significance of this deep learning framework is that it does not require any handcrafted flood-related features for training. In order to map flooding in congested areas, a global spatial-spectral convolutional neural network (GSSC) was proposed [Chen et al. 2020b] to extract the water information from MS images effectively. Though the study paves the way to fuse bi-temporal remote sensing images, the single sensor MS images have an impact on the accuracy of the flood mapping due to the presence of clouds. In another work [Wieland et al. 2019a], rapid segmentation of flooded areas on MS images was done with convolutional neural network for situational awareness in emergency response. Although
this study focuses on urban flood detection, the polarimetry information from other sensors in addition can improve the results. Semantic water segmentation is effectively done with the CNN (U-Net) [Wieland et al. 2019b], and the results show a good accuracy of 92%. Most of the flood scenes have challenges in processing due to the presence of clouds and shadows. These challenges can be addressed by augmenting the MS data with other sensor information.

To inherit the advantages of different sensors in terms of spatial and spectral characteristics and to improve flood mapping for precisely assessing the damage, multi-sensor image fusion is considered in this paper. We use the feature-level fusion of multispectral and SAR images for identifying the flood. To our knowledge, there is no work exclusively on deep learning-based image fusion to identify the flood in the images of SEN12-FLOOD dataset.

3 Proposed DeepFlood architecture

The proposed DeepFlood architecture for flood detection consists of dual patch Fully Convolutional Network (FCN) for feature extraction, followed by a feature fusion and a random forest classifier, as shown in Figure 1. The two patch FCNs are separately trained with SAR and multispectral images. A feature-level fusion is performed to concatenate the learned features of SAR and MS. Then, a random forest classifier [Leo 2001] is used to classify the Flood and No Flood patches. In order to classify the entire image of a region, maximum vote of patch classes is considered. The different stages of DeepFlood architecture are explained in detail next.

![Figure 1: Proposed DeepFlood architecture](image)

3.1 Dual patch FCN for feature extraction

The proposed dual patch FCN (Figure 2) consists of two 10-layered identical fully convolutional networks. One FCN processes $75 \times 75$ SAR image patches, and the other processes $75 \times 75$ MS image patches. The inputs of sizes $75 \times 75 \times 2$ of SAR and $75 \times 75 \times 11$ of multispectral are given separately to two FCNs. A sequence of 5 pairs of
convolution and max-pooling layers processes these patches. Each convolution layer uses a filter kernel of size $3 \times 3$ for learning the features and each max-pooling layer considers blocks of size $2 \times 2$ for dimensionality reduction. The convolutions are performed with zero padding and stride 2. The number of filters is doubled at each layer to increase the number of feature maps. The first convolutional layer has 64 filters of size $3 \times 3$. The second convolutional layer has 128 filters of size $3 \times 3$. The remaining three layers have 128, 256 and 512 filters each of size $3 \times 3$. Further, regularization techniques like batch normalization and dropouts are added to improve the generalization of the network. The independently learned SAR and MS features are flattened and concatenated at last for feature fusion.

For learning the features, the networks are independently trained with their respective images. A fully connected layer is added at the end of each network to train the networks for Flood and No Flood classification. The fully connected layer has 4608 input units ($3 \times 3 \times 512$). Once the training is over, these fully connected layers are discarded and the convolutional networks are capable of extracting relevant features from the input images.

![Figure 2: Proposed dual patch FCN](image)

3.2 Feature-level fusion and classification

The feature fusion is applied by concatenating the flattened feature maps of SAR ($F_{sar}$) and MS ($F_{ms}$) images obtained from the trained dual patch FCN. The fused feature vector ($F_{fus}$) contains the needful information from both sensor images for classification.

$$F_{fusion} : \{F_{sar}, F_{ms}\} \rightarrow F_{fus} \quad (1)$$

Finally, the fused features are fed to a random forest classifier for Flood and No Flood classification. Random forest is an ensemble learning which combines many decision trees classifiers to provide the solution to a complex problem. For the classification problem, the trees in random forest cast their vote [Gislason et al. 2006]. The output class is based on the majority votes of the trees. In the DeepFlood architecture, a random forest classifier with 250 estimators (trees) is added after the FCN for the final classification of Flood and No Flood images.
4 Experiments and results

The DeepFlood architecture was implemented in Python with TensorFlow library and run on a workstation with 16 GB RAM, Intel Core i7-9700K CPU230 and NVIDIA Titan 3840 XP GPU.

4.1 Dataset

SEN12-FLOOD dataset [Rambour et al. 2020b], available at IEEEdataport, was used to evaluate the efficacy of the proposed DeepFlood framework. It is a recently created fairly large dataset for flood detection studies [Rambour et al. 2020a] using deep learning. The dataset consists of pre-flood and post-flood images of 336 flood events in West and South-East Africa, Middle-East countries, and Australia. The images are acquired by the Sentinel 1 and 2 sensors. The Sentinel 1 SAR (with Vertical Vertical - VV and Vertical Horizontal - VH polarization) data were acquired in interferometric wide swath mode with a resolution of 10 × 10 m and pre-processed for radiometric calibration and range-Doppler terrain correction. The Sentinel 2 multispectral (12 bands) data were pre-processed for Level 2A atmospheric correction. Both the SAR (VV and VH) and multispectral (all 12 bands) images were resized to 512 × 512. The mean image of all the bands in the case of multispectral and the mean of VV and VH images in the case of SAR were computed. They were fed as inputs to FCN. The sample post-flood SAR and multispectral images from the SEN12-Flood dataset and their mean images are shown in (Figures 3 and 4)

![Sample post-flood SAR images](image)

4.2 Dual patch FCN training

The inputs to the dual patch FCN are SAR and MS images patches of size 75 × 75. One of the patch FCNs was trained with SAR image patches while the other with multispectral patches. From the SEN12-FLOOD dataset, 480 cloud-free MS images (240 Flood and 240 No Flood) and their corresponding 480 SAR images were taken for training. The network was trained with K-Fold (K=8) cross-validation. 120 SAR and 120 MS (60 Flood and 60 No Flood) images were considered for testing. The FCNs were trained with a batch size of 32 and a learning rate of 0.01. RMSprop optimizer was chosen for optimization.
4.3 Performance evaluation

The commonly used metrics such as precision, recall, F1-score and classification accuracy derived from the confusion matrix are used to evaluate the classification performance of the proposed DeepFlood architecture. True Positives (TP), False Positives (FP) and False Negatives (FN) are computed for Flood and No Flood classes. In the case of Flood class, TP is the number of Flood images that are classified correctly. FN is the number of Flood images that are classified as No Flood. FP is the number of No Flood images that are classified as a Flood. Precision (Producer Accuracy-PA) measures the fraction of the identified positives that are true positives and hence in the case of Flood class, PA is the fraction of patches classified as Flood by the network actually belongs to Flood. Recall (User Accuracy-UA) gives the fraction of correctly identified positives. It is the fraction of Flood patches that are classified as a Flood by the network. F1-score considers both PA and UA. The metrics are computed using TP, FN, FP and TN values.

\[
\text{Recall(UA)} = \frac{TP}{TP + FN} \tag{2}
\]

\[
\text{Precision( PA)} = \frac{TP}{TP + FP} \tag{3}
\]

\[
F1 \text{ - score} = \frac{2}{\frac{1}{UA} + \frac{1}{PA}} \tag{4}
\]

\[
\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{5}
\]
<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flood</td>
<td>0.83</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td>NoFlood</td>
<td>0.86</td>
<td>0.82</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 1: Classification performance metrics of SAR patch FCN

For the independently trained SAR and MS networks, the confusion matrices are shown in Figures 5 and 6 respectively. The confusion matrix (fused) giving the classification results of DeepFlood architecture is shown in Figure 7. The metrics computed are listed in Tables 1, 2 and 3. The feature maps generated from the first convolutional layer of SAR and MS-based FCNs are shown in Figure 8. From these results, it can be inferred that the fused network performs better than the independent networks. Further, SAR-based FCN performs better than MS-based FCN because SAR images are more responsive to water pixels than MS images due to their sensitivity to water and moisture. However, the additional spectral features of the MS images contribute to the improved results in DeepFlood architecture. The flood detection results of a region before and after a flood event are shown in Figure 9 on the underlying mean SAR images. The results are given for 75 × 75 patches of the entire 512 × 512 image of the region. The patches identified as Flood by the DeepFlood architecture are denoted by ‘F’. Most of the flooded areas are identified after flooding while in the image of the region before the flood event, no patch is identified as Flood.

4.4 Ablation studies

Ablation studies have been conducted to analyse the performance of random forest classifier. We analysed the classification performance for increasing number of estimators to decide the optimal number. It is observed from Figure 10 that the classification accuracy
increases up to 250 estimators and then begin to decrease. The number of estimators is therefore fixed to 250.

5 Comparision studies

The recent works that apply the SAR and multispectral images for flood detection were considered for comparison. The results of the proposed DeepFlood architecture and exist-

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flood</td>
<td>0.79</td>
<td>0.73</td>
<td>0.76</td>
</tr>
<tr>
<td>NoFlood</td>
<td>0.75</td>
<td>0.80</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 2: Classification performance metrics of MS patch FCN
ing flood detection approaches are provided in Table 4. One approach [Jacinth Jennifer et al. 2020] derives Normalized Difference Water Index (NDWI) for detecting flood. It is a statistical method but not very effective in identifying floods in urban areas. In another work, Relevance Vector Machine (RVM) [Sharifi 2020] has been applied on SAR to map flooded areas. Although the flood detection accuracy is reasonably high for SAR alone, taking advantage of another sensor information might help to increase the accuracy. A few deep learning based solutions are available in recent works. A Convolutional Neural Network (CNN) [Bhadra et al. 2020] architecture has been designed for identifying flood from multi-sensor data. It however achieves a lower accuracy due to smaller training
and testing datasets (with only 100 and 10 images respectively). On the SEN12-Flood dataset, a RESNET [Rambour et al. 2020a] has been applied independently for SAR and MS. On the same dataset, our proposed DeepFlood architecture gives better flood detection accuracies for both SAR and MS. By taking advantage of the feature fusion, the accuracy is further improved. The proposed fusion architecture gives the best accuracy when compared to other methods.

6 Conclusion

In this paper, a deep learning-based methodology, namely DeepFlood, was developed to perform feature-level fusion for Flood and No Flood classification of regions from their multisensor images. The multisensory (SAR and multispectral) image patches were initially fed to a dual patch FCN. The dual patch FCN is used for feature extraction and fusion of multi-sensor images. The SAR and multispectral perform differently in identifying the flood features. A suitably designed dual patch-based FCN along with a random forest classifier performs better than existing standard deep networks in flood detection. The proposed DeepFlood framework achieves the best accuracy of 94.17% in classifying the images in SEN12-FLOOD dataset.
<table>
<thead>
<tr>
<th>Approach</th>
<th>Study area/Dataset</th>
<th>Data</th>
<th>Flood detection accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDWI [Jacinth Jen-</td>
<td>Alappuzha region, Kerala</td>
<td>SAR and MS</td>
<td>83% (SAR+MS)</td>
</tr>
<tr>
<td>nifer et al. 2020]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RVM [Sharifi 2020]</td>
<td>Aqqala, Iran</td>
<td>SAR</td>
<td>89% (SAR)</td>
</tr>
<tr>
<td>CNN [Bhadra et al.</td>
<td>Barpeta and Kamrup of Assam, India</td>
<td>SAR and MS</td>
<td>80% (SAR+MS)</td>
</tr>
<tr>
<td>2020]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResNet-50 [Rambour et al. 2020a]</td>
<td>SEN12-FLOOD Dataset</td>
<td>SAR and MS</td>
<td>79% (MS), 75% (SAR)</td>
</tr>
<tr>
<td>DeepFlood</td>
<td>SEN12-FLOOD Dataset</td>
<td>SAR and MS</td>
<td>94.17% (SAR+MS), 84.17% (SAR), 76.67% (MS)</td>
</tr>
</tbody>
</table>


[Zhao et al. 2011] Zhao, M., Shang, H., Huang, W., Zou, L., Zhang, Y.: “Flood area extraction from rgb aerophotograph based on chromatic and textural analysis”; In International Conference on Advanced Geographic Information Systems, Applications and Services GeoProcessing (pp. 46-52), (2011).