Article

Enhancement of Detecting Permanent Water and Temporary Water in Flood Disasters by Fusing Sentinel-1 and Sentinel-2 Imagery Using Deep Learning Algorithms: Demonstration of Sen1Floods11 Benchmark Datasets

Yanbing Bai [1,†], Wenqi Wu [1,†], Zhengxin Yang [2], Jinze Yu [3], Bo Zhao [4], Xing Liu [5], Hanfang Yang [1,*], Erick Mas [6] and Shunichi Koshimura [6]

1   Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China; ybbai@ruc.edu.cn (Y.B.); 2015201627@ruc.edu.cn (W.W.)
2   China Huaneng Group Co., Ltd., Beijing 100031, China; zhx_yang@chng.com.cn
3   Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan; jzyu@ut-vision.org
4   School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, UK; bo.zhao@ed.ac.uk
5   Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan; ryu@vision.is.tohoku.ac.jp
6   International Research Institute of Disaster Science, Tohoku University, Sendai 980-8572, Japan; mas@irides.tohoku.ac.jp (E.M.); koshimura@irides.tohoku.ac.jp (S.K.)
*   Correspondence: hyang@ruc.edu.cn; Tel.: +86-10-6251-1318
†   These authors contributed equally to this work.

Abstract: Identifying permanent water and temporary water in flood disasters efficiently has mainly relied on change detection method from multi-temporal remote sensing imageries, but estimating the water type in flood disaster events from only post-flood remote sensing imageries still remains challenging. Research progress in recent years has demonstrated the excellent potential of multi-source data fusion and deep learning algorithms in improving flood detection, while this field has only been studied initially due to the lack of large-scale labelled remote sensing images of flood events. Here, we present new deep learning algorithms and a multi-source data fusion driven flood inundation mapping approach by leveraging a large-scale publicly available Sen1Flood11 dataset consisting of roughly 4831 labelled Sentinel-1 SAR and Sentinel-2 optical imagery gathered from flood events worldwide in recent years. Specifically, we proposed an automatic segmentation method for surface water, permanent water, and temporary water identification, and all tasks share the same convolutional neural network architecture. We utilize focal loss to deal with the class (water/non-water) imbalance problem. Thorough ablation experiments and analysis confirmed the effectiveness of various proposed designs. In comparison experiments, the method proposed in this paper is superior to other classical models. Our model achieves a mean Intersection over Union (mIoU) of 52.99%, Intersection over Union (IoU) of 52.30%, and Overall Accuracy (OA) of 92.81% on the Sen1Flood11 test set. On the Sen1Flood11 Bolivia test set, our model also achieves very high mIoU (47.88%), IoU (76.74%), and OA (95.59%) and shows good generalization ability.

Keywords: deep learning; Sen1Floods11 datasets; multi-source data fusion; Sentinel-1; Sentinel-2; permanent water; temporary water; flood

1. Introduction

Natural hazards such as floods, landslides, and typhoons pose severe threats to people's lives and property. Among them, floods are the most frequent, widespread, and deadly natural disasters. They affect more people globally each year than any other disaster [1–3]. Floods not only cause injuries, deaths, losses of livelihoods, infrastructure damage, and asset losses, but they can also have direct or indirect effects on health [4]. In

the past decade, there have been 2850 disasters that triggered natural hazards globally, of which 1298 were floods, accounting for approximately 46% [1]. In 2019, there were 127 floods worldwide, affecting 69 countries, causing 1586 deaths and more than 10 million displacements [2]. In the future, floods may become a frequent disaster that poses a huge threat to human society due to sea-level rise, climate change, and urbanization [5,6].

The effective way to reduce flood losses is to enhance our ability of flood risk mitigation and response. In recent years, timely and accurate flood detection products derived from satellite remote sensing imagery are becoming effective methods of responding flood disaster, and city and infrastructure planners, risk managers, disaster emergency response agency, and property insurance company are benefiting from it in worldwide [7–11], while identifying permanent water and temporary water in flood disasters efficiently is a remaining challenge.

In recent years, researchers have done a lot of work in flood detection based on satellite remote sensing images. The most commonly used method to distinguish between water and non-water areas is a threshold split-based method [12–15]. However, the optimal threshold is affected by the geographical area, time, and atmospheric conditions of image collection. Therefore, the generalization ability of the above methods is greatly limited. In recent years, the European Space Agency (ESA) developed a series of Sentinel missions to provide free available datasets including Synthetic Aperture Radar (SAR) data from Sentinel-1 sensor and optical data from Sentinel-2 sensor, taking into account the respective advantages of optical imagery and SAR imagery in flood information extraction [5,14,16–18], and the investigation of combining SAR images with optical images for more accurate flood mapping is of great interest for researchers [16,19–22]. Although the data fusion method improves the accuracy of flood extraction, it is still very challenging to distinguish permanent water from temporary water problems in flood disasters. The identification of temporary waters in flood disasters and permanent waters mainly rely on multi-temporal change detection methods [23–28]. This type of approach requires at least one pair of multi-temporal remote sensing scenes which were acquired before and after a flood event. Although the method based on multi-temporal change detection can better detect the temporary water in flood disaster events, the multi-temporal method was greatly limited due to the mandatory demand for satellite imagery before disasters.

Deep learning methods represented by convolutional neural networks have been proven to be effective in the field of flood damage assessment, and related research has grown rapidly since 2017 [29]. However, most algorithms are focusing on affected buildings in flood events [22,30], with very few examples of flood water detection. The latest research focuses on the application of deep learning algorithms for enhancing flood water detection [31,32]. Early research focused on the extraction of surface water [33–35]. Furkan et al. proposed a deep-learning-based approach for surface water mapping from Landsat imagery. The results demonstrated that the deep learning method outperform the traditional threshold and Multi-Layer Perceptron model. Considering the difference in the characteristics of surface water and floods in satellite imagery, which will increase the difficulty of flood extraction. Maryam et al. developed a semantic segmentation method for extracting the flood boundary from UAV imagery. The semantic segmentation-based flood extraction method was further applied to identify the flood inundation caused by mounting destruction [36]. Experimental results validate the efficiency and effectiveness of the proposed method. Muñoz et al. [37] combined the multispectral Landsat imagery and dual-polarized synthetic aperture radar imagery to evaluate the performance of integrating convolutional neural network and data fusion framework for generating compound flood mapping. The usefulness of this method was verified by comparing with other methods. These studies show that deep learning algorithms play an important role in enhancing flood classification. However, research in this field is still in its infancy, due to the lack of high-quality large-scale flood annotation satellite datasets.

Recent development in earth observation has contributed a series of open-sourced large scale disaster related satellite imagery datasets, which has greatly spurred the advance

of leveraging deep learning algorithm for disaster mapping from satellite imagery. For building damage classification, the xBD dataset has provided large scale satellite imagery data that collected from the multi-type disasters with four category damage level labels to worldwide researchers, and the research spawned by this public data has also verified the great potential of deep learning in building damage recognition [38,39]. For flooded building damage assessment in Hurricane disaster events, FloodNet provides a high-resolution UAV image dataset and has done the same task [40]. The recent release of the large-scale open-source Sen1Floods11 dataset [5] is boosting the research of utilizing deep learning algorithms for water type detection in flood disasters [41]. For water type detection in flood disaster events, Sen1Flood should take on a similar role. Unfortunately, so far, only one preliminary work has been conducted.

With the purpose of developing an efficient benchmark algorithm for distinguishing between permanent water and temporary water in flood disasters based on the Sen1Flood dataset to boost the research in this area, the contributions and originality of this research are as follows.

- Effectiveness: To the best of our knowledge, in terms of the sen1flood11 dataset, the accuracy of our proposed algorithm is the highest so far.
- Convenience: All of the sentinel-1 and sentinel-2 imagery utilized in the model come from post-flood imagery, and this greatly reduces the reliance on satellite imagery data before the flood.
- Refinement: We introduced salient object detection algorithm to modify the convolutional neural network classifier, in addition, the multi-scale loss algorithm and data augmentation were adopted to improve the accuracy of the model.
- Robustness: the robustness of our proposed algorithm was verified in a new Bolivia flood dataset.

2. Sen1Floods11 Dataset

We utilize the Flood Event Data in the Sen1Floods11 dataset [5] to train, validate, and test deep learning flood algorithms. This dataset provides raw Sentinel-1 SAR images (IW mode, GRD product), Sentinel-2 MSI Level-1C images, classified permanent water, and flood water. There are 4831 non-overlapping 512×512 tiles from 11 flood events. This dataset helps map flood at the global scale, covering 120,406 square kilometers, spanning 14 biomes, 357 ecological regions, and six continents. Locations for flood events are shown in Figure 1.

For each selected flood event, the time interval between the acquisition of Sentinel-1 imagery and Sentinel-2 imagery shall not exceed two days. The Sentinel-1 imagery contains two bands, VV and VH, which are backscatter values; the Sentinel-2 imagery includes 13 bands, and all bands are TOA reflectance values. The imagery is projected to the WGS-84 coordinate system.The ground resolution of the imagery is different on different bands. In order to fuse images, the ground resolution is sampled to 10 m on all bands. Each band is visualized in Figure 2.

Due to the high cost of hand labels, 4370 tiles are not hand-labeled and exported with annotations automatically generated by the Sentinel-1 and Sentinel-2 flood classification algorithms, which can serve as weakly supervised training data. The remaining 446 tiles are manually annotated by trained remote sensing analysts for high-quality model training, validation and testing. The weakly supervised data contain two types of surface water labels. One is produced by the histogram thresholding method based on the Sentinel-1 image; the other is generated by the Normalized Difference Vegetation Index (NDVI), MNDWI and thresholding method based on the Sentinel-2 image. All cloud and cloud shadow pixels were masked and excluded from training and accuracy assessments. Hand labels include all water labels and permanent water labels. For all water labels, analysts exploited Google Earth Engine to correct the automated labels using Sentinel-1 VH band, two false color composites from Sentinel-2 and the reference water classification from Sentinel-2 by removing uncertain areas and adding to the water classification. For the

permanent water label, with the help of the JRC (European Commission Joint Research Center) surface water data set, Bonafilia et al. [5] labeled the pixels that were detected as water at both the beginning (1984) and end (2018) of the dataset as permanent water pixels. The pixels never observed as water during this period are treated as non-water pixels. The remaining pixels are masked. Examples of water label are visualized in Figure 3.
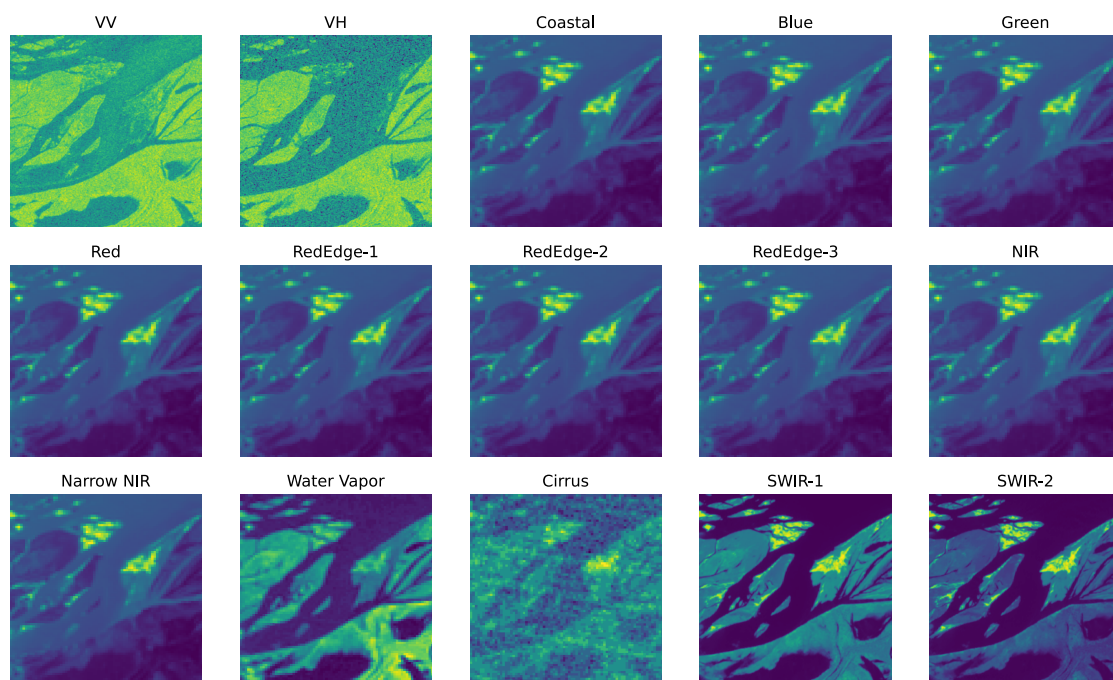

Figure 1. Sample locations of flood event data.
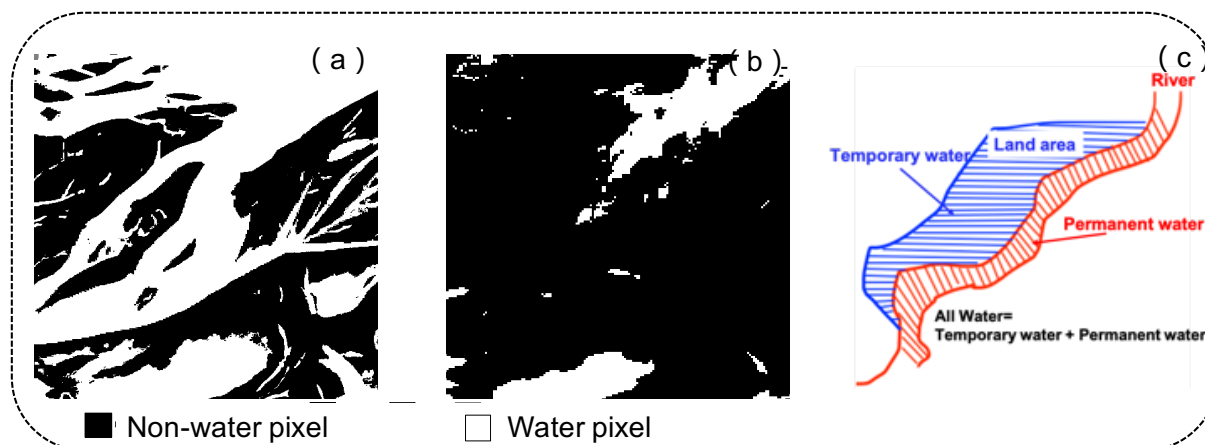

Figure 2. Example of all the bands.

Figure 3. Illustration of water label. (**a**) example of hand labeled data of all water; (**b**) example of hand labeled data of permanent water; (**c**) the illustration of temporary water, permanent water, and all water.

Like most existing studies [42], the Sen1Floods11 dataset shows the highly imbalanced distribution between flooded and unflooded area. As shown in Table 1, for all water, water pixels account for only 9.16%, and non-water pixels account for 77.22%, which is about eight times the number of surface water pixels. The percentages of water pixels and non-water pixels in permanent waters are 3.06% and 96.94%, respectively, and the number of non-water pixels is about 32 times that of non-water pixels.

Table 1. Water, Non-Water, and Ignored area proportion in pixel level

Label	Water	Non-Water	Ignored
All Water	9.16%	77.22%	13.63%
Permanent Water	3.06%	96.94%	0.00%

The dataset is split into three parts: training set, validation set, and test set. All 4370 images automatically labeled are used as the weakly supervised training set. The hand-labeled data are first randomly split into training, validation, and testing data in the proportion 6:2:2. In order to test the model's ability to predict unknown flood events, all hand-labeled data related to the Bolivia flood event is held out for a distinct test set. Rest hand-labeled data in training, validation, and testing sets are composed of final training, validation, and testing set, respectively. Correspondingly, all data from Bolivia in the weakly-supervised training set are also excluded and do not participate in model training. The overall composition of the dataset is shown in Table 2.

Table 2. Sen1Floods11 dataset composition

Dataset		Sample Size
Training Set	Hand-Labeled	252
	Weakly-Labeled	4160
Validation Set		89
Testing Set		90
Bolivia Testing Set		15

3. Method

Figure 4 depicts a flowchart of our work. In this work, the benchmark Sen1Floods11 dataset that contains 4831 samples with 512 × 512-pixel size from both Sentinel-1 and Sentinel-2 imagery were utilized to develop the algorithm. The spatial resolution resampling and pixel value normalization were adopted to fuse sentinel-1 and sentinel-2 imagery. The model input is a stack of fused image bands with permanent water and temporary water annotations. The network used is BASNet proposed by Qin et al. [43] is used. BASNet architecture is shown in Figure 5. The network combines a densely supervised encoder–decoder network similar to U-Net and a new residual refinement module. The encoder–decoder produces a coarse probability prediction map from the image input, and the Residual Refinement Module is responsible for learning the residuals between the coarse probability prediction map and the ground truth. We apply the network to remote sensing data sets, adapt, train, and optimize it to better predict flood areas.
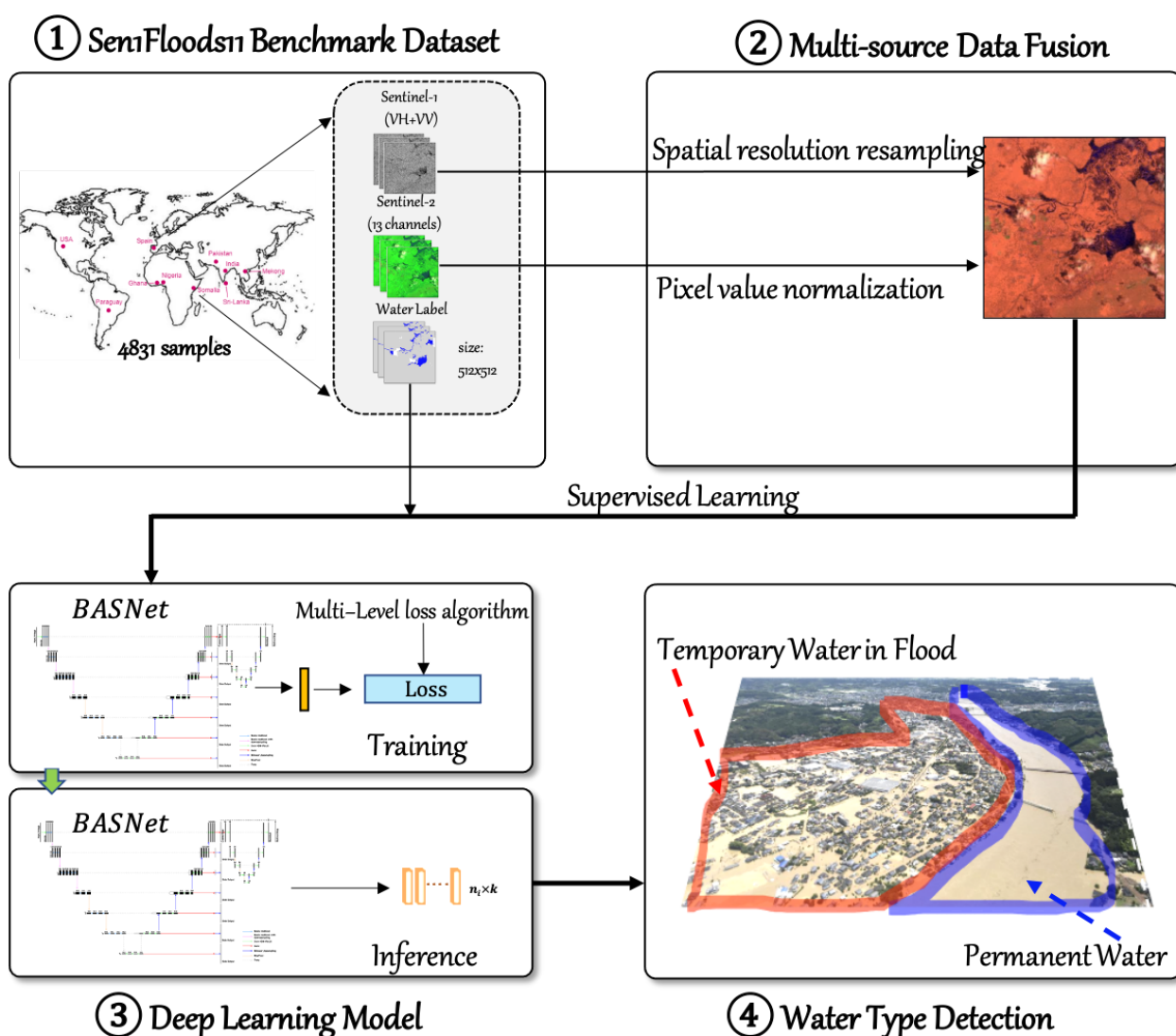

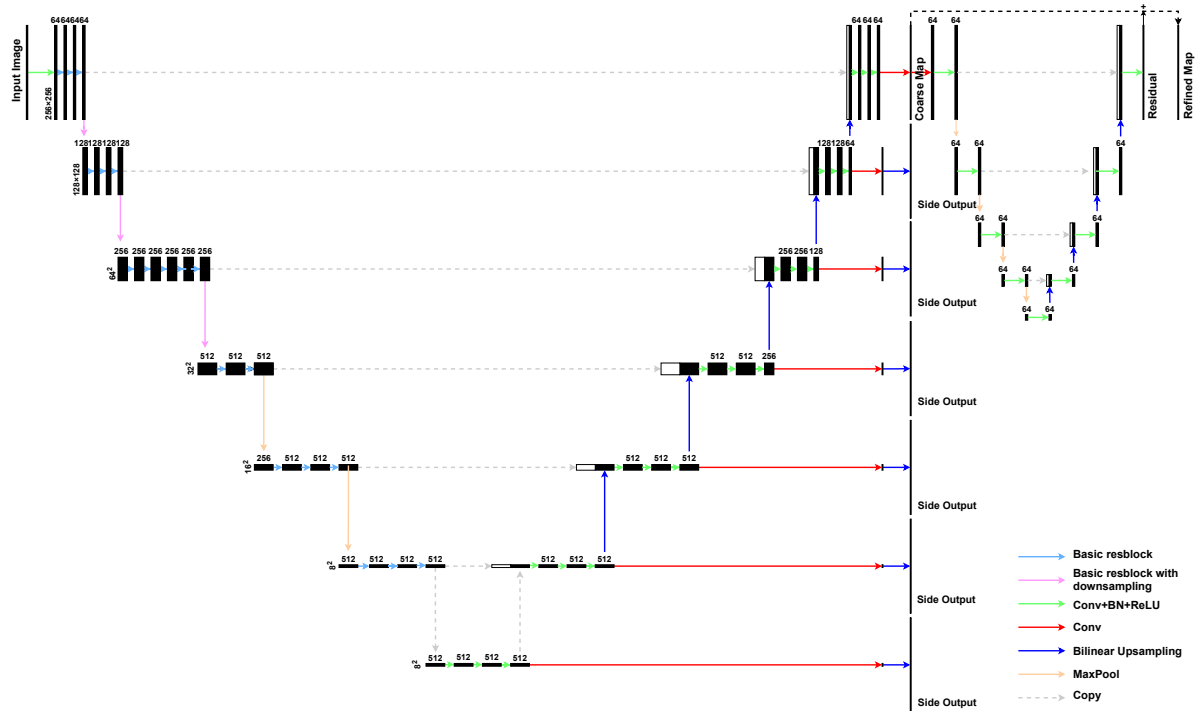**Figure 4.** Flowchart of water type detection proposed in this work.

Figure 5. Architecture of the BASNet used in this study.

3.1. Encoder–Decoder Network

The encoder–decoder network can fuse abstract high-level information and detailed low-level information and is mainly responsible for water body segmentation. The encoder part contains an input convolution block and six convolution stages consisting of basic res-blocks. The input convolution block comprises a convolution layer with batch normalization [44] and Rectified Linear Unit (ReLU) activation function [45]. The size of the convolution kernel is 3×3, and the stride is 1. This convolution block can convert an input image of any number of channels to a feature map of 64 channels. The first four convolution stages directly use the four stages of ResNet34 [46]. Except for the first residual block, each residual block will double the feature map's channels. The last two convolution stages have the same structure, and both consist of three basic res-blocks with 512 filters and a 2×2 max pooling operation with stride 2 for downsampling.

Compared with traditional convolution, atrous convolution can obtain a larger receptive field without increasing parameter amount and capture long-range information. In addition, atrous convolution can avoid the reduction of feature map resolution caused by repeated downsampling and allow a deeper model [47,48]. To capture global information, Qin et al. [43] designed a bridge stage to connect the encoder and the decoder. This stage is comprised of three atrous convolution blocks. Each atrous convolution block consists of a convolution layer with 512 atrous 3×3 filters with dilation 2, a batch normalization, and a ReLU activation function.

In the decoder part, each decoder stage corresponds to an encoder stage. As shown in the Formulas (1) and (2), g_i is the merge base, f_i is the feature map to be fused, h_i is the merged feature map, and $conv_{3 \times 3}$ is a convolution layer followed by batch normalization and ReLU activation, RRM represents the Residual Refinement Module and the operator $[\cdot; \cdot]$ represents concatenation along the channel axis. There are three convolution layers with batch normalization and a ReLU activation function in each decoder stage. The feature map from the last stage is first fed to an up-sampling layer to get g_{i+1}, and then concatenated with the current feature map f_i. In order to alleviate overfitting, the last layer of the bridge stage and each stage of the encoder is fed to a 3×3 convolution layer followed by a bilinear up-sampling layer and sigmoid activation function to generate a prediction map and then supervised by the ground truth:

$$g_i = \begin{cases} RRM(h_i) & \text{if } i = 1 \\ upsample(h_i) & \text{if } i \geq 2 \\ h_i & \text{if } i = 7 \end{cases} \tag{1}$$

$$h_i = \begin{cases} conv_{3\times3}(conv_{3\times3}(conv_{3\times3}([g_{i+1}; f_i]))) & \text{if } i \neq 7 \\ f_i & \text{if } i = 7 \end{cases} \tag{2}$$

3.2. Residual Refinement Module

The residual refinement module learns the residuals between the coarse maps and the ground truth and then adds them to the coarse maps to produce the final results. By fine-tuning the prediction results, the fuzzy and noisy boundaries can be made sharper. The probability gap between water and non-water pixels can be increased. Compared to the encoder–decoder network, the residual refinement module has a simpler architecture containing an input layer, a four-stage encoder–decoder with a bridge, and an output layer. Each stage has only one convolution layer followed by a batch normalization and a ReLU activation function. The convolution layer has 64 3×3 filters with stride 1. In addition, down-sampling and up-sampling are performed through non-overlapping 2×2 max pooling and bilinear interpolation in the encoder–decoder network.

3.3. Hybrid Loss

Training loss is defined as the summation of all outputs' losses:

$$L = \sum_{k=1}^{K} l^{(k)} \tag{3}$$

where l is the loss of the k-th output. Here, $K = 8$, including seven side outputs from the bridge stage and decoder and one final output from the refinement module. Each loss is comprised of three parts: focal loss [49], Structural SIMilarity (SSIM) loss [50], and Intersection over Union (IoU) loss [51]:

$$l = l_{focal} + l_{ssim} + l_{iou} \tag{4}$$

We replace Binary Cross Entropy (BCE) loss in Qin et al. (2019) [43] with focal loss. It is defined as:

$$l_{focal} = \begin{cases} -\alpha(1-p)^\gamma log(p) & \text{if } y = 1 \\ -(1-\alpha)p^\gamma log(1-p) & \text{if } y = 0 \end{cases} \tag{5}$$

where y specifies the ground-truth class and p is the model's estimated probability for the water class. The focal loss is designed to address the extreme imbalance between water and non-water classes during training. On the one hand, using weighting factor $\alpha \in [0, 1]$ for water and $1 - \alpha$ for non-water to balance the importance of water/non-water pixels, focal loss can avoid non-water pixels dominating the gradient. Larger α puts more weight on water pixels. On the other hand, with the modulating factor γ, focal loss can reduce the loss contribution from easy examples and thus focus training on hard non-waters (e.g., boundary pixels). Focal loss is a pixel-level measure and can be utilized to maintain a smooth gradient for all pixels.

Taking each pixel's local neighborhood into account, SSIM loss is a patch-level measure and is developed to capture the structural information in an image. It gives a higher loss around the boundary when the predicted probabilities on the boundary and the inner pixels are the same. Thus, it can drive the model to focus training on the boundary pixels, which are usually harder to classify. It is defined as:

$$l_{ssim} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

where μ_x, μ_y, and σ_x, σ_y are the mean and standard deviations of patch x and y, respectively, σ_{xy} is their covariance. C_1 and C_2 are constants which are used to avoid zero denominators.

IoU loss is a map-level measure. It gives more focus on the water body, whose error rate is usually higher than non-water area. For water pixels, a larger $p_{r,c}$ stands for higher confidence of the network prediction. The higher the model prediction's confidence of the water is, the lower the loss is. It is defined as:

$$l_{iou} = 1 - \frac{\sum_{r=1}^{H} \sum_{c=1}^{W} p_{r,c} y_{r,c}}{\sum_{r=1}^{H} \sum_{c=1}^{W} [p_{r,c} + y_{r,c} - p_{r,c} y_{r,c}]} \tag{7}$$

where H, W are the height and width of the image, respectively.

4. Experimental Analysis

As introduced in the data section, we first use the 4160 automatically annotated images to pre-train our network and then use 252 manually annotated images to fine-tune it. We monitor convergence and overfitting during training on the validation set while evaluating the model performance on the test set and Bolivia test set.

4.1. Implementation Detail and Experimental Setup

The backbone used in BASNet is ResNet-34 [46] for all the experiments, which was pre-trained on ImageNet [52]. Other convolution layers are initialized by Xavier [53]. For hyperparameters introduced by focal loss, γ is set to 2.5, and α is set to 0.25. We utilize the Adam optimizer [54] to train our network and all hyperparameters are set by default, namely lr $= 0.001$, betas $= (0.9, 0.999)$, eps $= 1 \times 10^{-8}$, weight dacay $= 0$. A "poly" learning rate policy is used to adjust the learning rate; that is, the learning rate is multiplied by $(1 - \frac{step}{maxstep})^{power}$ with power $= 0.9$. The image batch size during training is 8.

We apply some data augmentation methods to enhance the generalization ability of our model. Specifically, random horizontal and vertical flip and random rotation of $45 \times k, (k = 1, 2, 3, 4)$ degrees with a probability of 0.5 are performed on each image. For extra data preprocessing, we randomly crop the image into a fixed size of 256×256, and the pixel value is normalized to a value in the range of $[0, 1]$.

We implement our network based on the Pytorch deep learning framework. Both training and testing are conducted on Public Computing Cloud, Renmin University of China using a single NVIDIA Titan RTX GPU and an Inter Xeon Silver 4114 @2.20GHz CPU.

4.2. Evaluation Metrics

We use five measures to evaluate our method: Intersection over Union (IoU), mean Intersection over Union (mIoU, equal weighting of all tiles), Overall Accuracy (OA), omission error rates (OMISSION), and commission error rates (COMMISSION) [55]. IoU, mIoU, and OA are standard evaluation metrics for image segmentation tasks. Omission and commission error rates are reported for comparison to remote sensing literature [5,56]. Omission rates are false negative water detection rates, and commission rates are false positive water detection rates. When the number is larger, the model's performance is worse. Following the common practice, we used the mIoU as the primary metric to evaluate methods. We calculate all the above metrics for water body segmentation, permanent water body segmentation, and temporary water body segmentation. The definitions are as follows:

$$IoU_i = \frac{TP_i}{FN_i + FP_i + TP_i} \tag{8}$$

$$mIoU = \frac{1}{N} \sum_{i=1}^{N} IoU_i \tag{9}$$

$$IoU = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(FN_i + FP_i + TP_i)} \quad (10)$$

$$Accuracy_i = \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i} \quad (11)$$

$$OA = \frac{1}{N}\sum_{i=1}^{N} Accuracy_i \quad (12)$$

$$OMISSION = \frac{\sum_{i=1}^{N} FN_i}{\sum_{i=1}^{N}(FN_i + TP_i)} \quad (13)$$

$$COMMISSION = \frac{\sum_{i=1}^{N} FP_i}{\sum_{i=1}^{N}(FP_i + TN_i)} \quad (14)$$

where i is the index of image. For the i-th sample, TP_i, TN_i, FP_i, FN_i is the number of accurately categorised water pixels, accurately categorized non-water pixels, misclassified non-water pixels, and misclassified water pixels, respectively.

4.3. Ablation Study

In this section, we conduct experiments to verify each improvement measures' effectiveness in our model. The ablation study contains three parts: data augmentation ablation, loss ablation, and image fusion ablation. The baseline is the original BASNet network with only the Sentinel-1 SAR image as input. Tables 3 and 4 show summaries of the test set and Bolivia test set results, respectively.

Table 3. Evaluation on the Sen1Floods11 test set for ablation studies. Results showed on all water (AW), permanent water (PW), and temporary water (TW).

Task	Method	Augments	Focal Loss	Image Fuse	mIoU (%)	IoU (%)	OA (%)
All Water	Baseline				29.39	35.08	79.49
	+Augments	✓			30.36	42.27	84.04
	+Focal Loss		✓		42.77	54.10	90.58
	+Image Fuse			✓	47.60	63.13	92.85
	+Augments+Focal Loss	✓	✓		42.95	53.86	90.80
	+Augments+Focal Loss+Image Fuse	✓	✓	✓	58.73	64.52	93.38
Permanent Water	Baseline				40.09	45.81	91.50
	+Augments	✓			34.10	38.14	88.18
	+Focal Loss		✓		50.06	35.48	91.47
	+Image Fuse			✓	65.04	50.81	93.49
	+Augments+Focal Loss	✓	✓		59.74	50.88	93.50
	+Augments+Focal Loss+Image Fuse	✓	✓	✓	68.79	52.03	93.84
Temporary Water	Baseline				25.69	31.68	83.52
	+Augments	✓			27.81	33.44	85.60
	+Focal Loss		✓		34.64	39.19	89.13
	+Image Fuse			✓	50.53	51.19	92.97
	+Augments+Focal Loss	✓	✓		34.75	38.99	88.19
	+Augments+Focal Loss+Image Fuse	✓	✓	✓	52.99	52.30	92.81

Table 4. Evaluation on Sen1Floods11 Bolivia test set for ablation studies. Results shown on all water (AW), permanent water (PW), and temporary water (TW).

Task	Method	Augments	Focal Loss	Image Fuse	mIoU (%)	IoU (%)	OA (%)
All Water	Baseline				45.31	64.54	91.40
	+Augments	✓			45.51	64.95	91.53
	+Focal Loss		✓		47.12	70.63	94.07
	+Image Fuse			✓	46.39	67.37	92.22
	+Augments+Focal Loss	✓	✓		47.42	71.23	94.20
	+Augments+Focal Loss+Image Fuse	✓	✓	✓	54.07	78.90	95.79
Permanent Water	Baseline				38.78	36.10	95.31
	+Augments	✓			43.37	40.43	94.32
	+Focal Loss		✓		67.29	71.79	99.06
	+Image Fuse			✓	72.30	75.42	99.23
	+Augments+Focal Loss	✓	✓		70.23	76.45	99.22
	+Augments+Focal Loss+Image Fuse	✓	✓	✓	75.27	78.80	99.39
Temporary Water	Baseline				40.69	64.32	91.76
	+Augments	✓			40.70	64.66	92.08
	+Focal Loss		✓		41.12	66.47	92.79
	+Image Fuse			✓	43.53	70.60	94.23
	+Augments+Focal Loss	✓	✓		40.65	64.13	92.20
	+Augments+Focal Loss+Image Fuse	✓	✓	✓	47.88	76.74	95.59

4.3.1. Data Augmentation

As can be seen from Table 3, after applying data augmentation, on the test set, the mIoU, IoU, and OA of the water body segmentation task are increased by 0.97%, 7.19%, and 4.55%, respectively. Those of the temporary water body segmentation task increase by 2.12%, 1.76%, and 2.08%, respectively. There is a decrease on the permanent water body detection task. Data augmentation improves mIoU, IoU, and OA on all three tasks on the Bolivia test set (Table 4).

4.3.2. Loss Function

As shown in Table 5, the number of water pixels is far more than that of non-water pixels on all tasks. Specifically, the number of non-water pixels is 8.43, 17.49, and 11.81 times that of surface, permanent and temporary water pixels, respectively. We can see from Tables 3 and 4, on the test set, that focal loss outperforms cross-entropy loss with respect to mIoU, IoU, and OA on all tasks, especially the permanent water mapping task. For permanent water body extraction, focal loss brings improvements of mIoU by 25.64%, IoU by 12.74%, and OA by 5.32% (Table 3). On the Bolivia test set, focal loss significantly improved the permanent water detection, showing an improvement for mIoU of 26.87%, IoU of 36.02%, and OA of 4.90% (Table 4). Meanwhile, it achieves the best performance on the test set's surface and temporary water detection tasks and the Bolivia test set's surface water detection task. In addition, it achieves competitive performance in temporary water detection on the Bolivia test set.

Table 5. Water, Non-Water, and Ignored area proportion at the pixel level of all water (AW), permanent water (PW), and temporary water (TW).

Label	Water	Non-Water	Ignored
All Water	9.16%	77.22%	13.63%
Permanent Water	4.26%	74.49%	21.25%
Temporary Water	6.54%	77.22%	16.25%

Comparison with other loss functions: Besides focal loss, distributional ranking (DR) loss [57] and normalized focal loss [58] are also proposed to address the class imbalance problem.

DR loss [57] treats the classification problem as a ranking problem and improves object detection by distributional ranking supplementary. The distributional ranking model ranks the distributions between positive and negative examples in the worse-case scenario. As a result, this loss can handle a class imbalance problem and the problem of imbalanced hardness of negative examples as well as maintaining efficiency. In addition, it can separate the foreground (water) and background (non-water) with a large margin. We have the DR loss as:

$$\min_{\theta} \mathscr{L}_{DR}(\theta) = \sum_{i}^{N} l_{logistic}(\hat{P}_{i,j_-} - \hat{P}_{i,j_+} + \gamma) \tag{15}$$

$$l_{logistic}(z) = \frac{1}{L}\log(1 + \exp(Lz)) \tag{16}$$

$$\hat{P}_{i,j_-} = \sum_{j_-}^{n_-} \frac{1}{Z_-} \exp(\frac{p_{i,j_-}}{\lambda_-}) p_{i,j_-} = \sum_{j_-}^{n_-} q_{i,j_-} p_{i,j_-} \tag{17}$$

where j_+ and j_- denote the water and non-water pixels, respectively. $q_+ \in \Delta$ and $q_- \in \Delta$ denote the distributions over water and non-water pixels, respectively. P_+ and P_- represent the expected scores under the corresponding distribution. $\Delta = \{q : \sum_j q_j = 1, \forall j, q_j \geq 0\}$.

Zheng et al. [58] modified focal loss for balanced optimization. They adjust the loss distribution without a change of sum to avoid gradient vanishing. The paper introduced a normalization constant Z that guarantees

$$\sum l(p_j, y_j) = \frac{1}{Z} \sum (1 - p_j)^{\gamma} l(p_j, y_j) \tag{18}$$

where $l(p_j, y_j)$ denotes the j-th pixel's cross entropy loss. p_j represents the j-th pixel's predicted probability. In addition, y_j is its ground truth. Hence, for the loss of each pixel, they produce a new weight $\frac{1}{Z}(1 - p_j)^{\gamma}$.

We carry out experiments to compare these loss functions with focal loss. We use DR loss with $\lambda_+ = 1$, $\lambda_- = \frac{1}{log(3.5)}$, $L = 6$, $\tau = 4$ and set $\gamma = 2.5$, $\alpha = 0.25$ in focal loss and normalized focal loss. Table 5 shows that the permanent water detection task suffers the most from the class imbalance problem. Tables 6 and 7 compare the mIoU of the three loss functions on the three tasks. Focal loss obtains slightly inferior results in surface and temporary water detection on the test set. Still, both on the test and Bolivia test set, focal loss produces the highest mIoU in permanent water detection. With respect to normalized focal loss, focal loss (on the test set and the Bolivia test set) brings 2.74% and 3.16% performance gains in mIoU, respectively. For DR loss, focal loss gains 1.69% and 10.83% in mIoU, respectively. In addition, focal loss outperforms normalized focal loss and DR loss on the Bolivia test set across all other tasks.

Table 6. Comparison on different losses in terms of mIoU(%) on the Sen1Floods11 test set. Results shown on all water (AW), permanent water (PW), and temporary water (TW).

Loss	AW	PW	TW
Normalized Focal Loss	42.26	57.00	35.23
DR Loss	43.32	58.05	36.87
Focal Loss	42.95	59.74	34.75

Table 7. Comparison on different losses in terms of mIoU(%) on the Sen1Floods11 Bolivia test set. Results shown on all water (AW), permanent water (PW), and temporary water (TW).

Loss	AW	PW	TW
Normalized Focal Loss	47.36	67.08	40.33
DR Loss	43.99	59.41	36.08
Focal Loss	47.42	70.24	40.65

4.3.3. Image Fusion

After fusing Sentinel-2 optical imagery and Sentinel-1 SAR imagery, results improve significantly on all tasks, which demonstrates that optical imagery can provide useful supplementary information on water segmentation.

4.4. Comparison with General Methods

4.4.1. Evaluation on the Sen1Floods11 Test Set

To evaluate our method, we conduct comprehensive experiments on the Sen1Floods11 dataset. On the one hand, since Sen1Floods11 is a newly released data set, there are few existing methods to conduct experiments on this data set; on the other hand, existing methods use different evaluation metrics, and the experimental results of different methods cannot be directly compared. Therefore, we reproduce the classical methods in remote sensing and some CNN (convolutional neural network)-based methods from classical to state-of-the-art to compare with our model under uniform experimental conditions and with the same evaluation codes. These methods include the Otsu thresholding method based on the VH band [5], FCN-ResNet50 [5], Deeplab v3+ [59], and U^2-Net [60].

The Otsu thresholding method [5,15] is a widely used method in water body extraction. It can degenerate a grayscale image into a binary image with the best threshold to distinguish the two different types of pixels. The between-class variance is calculated according to the specific algorithm corresponding to Otsu [15]. Then, find the threshold corresponding to the largest between-class variance as the best threshold. This method is unsupervised, simple, and fast. ResNet [46] is utilized as a standard backbone in most networks. Bonafilia et al. [5] use the fully convolutional neural networks (FCNN) model with a ResNet50 backbone to map floods. Here, we compare our method with it. Our tasks can also be regarded as segmentation tasks. Chen et al. [48] proposed Deeplab for semantic segmentation. Here, we try to use it to map floods and compare it with our model. We use the latest version of Deeplab(Deeplabv3+ [59]) as our comparison. We replace its Aligned Xception with Resnet-50 [46] to decrease the parameters amount and computational complexity. Moreover, to account for the relatively small batch size, we convert all of the batch normalization layers to group normalization layers. Considering water bodies as the salient object, we can solve our problems by SOD models. U^2-Net [60] is a SOD network and has mainly two advantages compared with previous architectures. First, it allows training from scratch rather than from existing pre-trained backbones, which avoids the problem of distributional differences between RGB images and satellite imagery; second, it can achieve deeper architecture while maintaining high-resolution feature maps at a low memory and computation cost.

Quantitative comparison: We train and test all the models on the same data set. In addition, we use the same evaluation code to evaluate all the predicted maps for a fair comparison. Table 8 summarizes the mIoU, IoU, OMISSION, COMMISSION, and OA of all the methods on the Sen1Floods11 test set. We underline the best results under each metric. As can be seen, on the test set, the method proposed in this paper outperforms other methods by a large margin (over 18%) in surface water, permanent water as well as temporary water segmentation in mIoU. In terms of COMMISSION and OA, our model achieves the best result on all tasks. For IoU, the proposed method largely improves most tasks (over 8%) except that Otsu and FCN-ResNet50 are superior in permanent water segmentation on the test set. There is still room for improvements on the test set.

Table 8. Performance comparison with other methods on Sen1Floods11 test set. Results shown on all water (AW), permanent water (PW), and temporary water (TW). xx indicates the best performance.

Task	Method	mIoU(%)	IoU(%)	OMISSION(%)	COMMISSION(%)	OA(%)
All Water	Otsu	35.73	54.58	26.68	4.91	90.01
	FCN-ResNet50	30.9	49.32	28.49	8.26	88.56
	Deeplab v3	32.08	47.67	34.89	6.72	87.96
	U^2Net	38.39	52.03	32.95	5.30	89.96
	BASNet	58.73	64.52	31.19	1.16	93.38
Permanent Water	Otsu	47.78	62.81	1.22	4.90	92.83
	FCN-ResNet50	35.19	55.11	18.04	7.55	90.70
	Deeplab v3	37.54	44.60	43.55	4.12	90.18
	U^2Net	38.36	27.80	64.03	4.56	88.34
	BASNet	68.79	52.03	47.43	0.16	93.84
Temporary Water	Otsu	27.95	38.98	40.42	4.91	89.13
	FCN-ResNet50	22.55	30.30	24.49	19.93	79.50
	Deeplab v3	34.49	41.58	46.09	3.96	90.27
	U^2Net	31.05	44.08	36.07	6.01	89.77
	BASNet	52.99	52.30	43.15	1.16	92.81

Qualitative comparison: Figure 6 depicts the flooding maps of each model (binary maps for the Otsu method, probability maps for FCN-ResNet50, Deeplab, U^2Net, and our model) on some samples from the Sen1Floods11 test set. For the Otsu method, lots of details such as small river tributaries and fragmented land in the middle of rivers are missed. For FCN-ResNet50, Deeplab, and U^2Net, besides missing details, we can observe the large gray area in predicted probability maps, which demonstrates that these CNN-based models can only produce low confidence predictions and blurred boundary. However, our method produces both clear boundaries and sharp-contrast maps. Even in urban area, our method produce an accurate map. Compared with other models, the proposed method produces clearer and more accurate prediction maps.
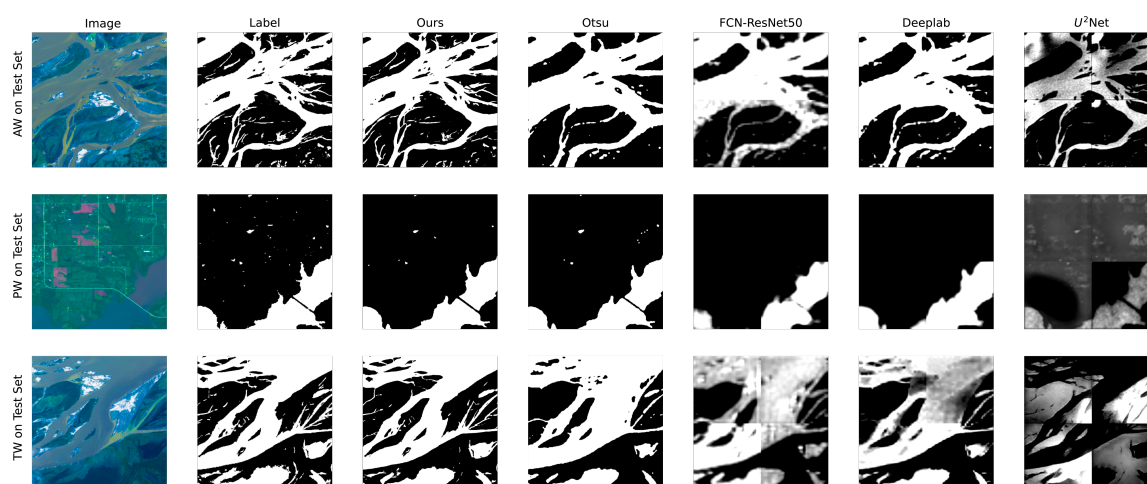

Figure 6. Qualitative comparison of proposed method with other methods. Each row represents one image and corresponding flooding maps (binary maps for the Otsu method, probability maps for FCN-ResNet50, Deeplab, U^2Net, and our model). Each column represents one method. Results shown on all water (AW), permanent water (PW), and temporary water (TW).

4.4.2. Evaluation of the New Scenario: Bolivia Flood Datasets

Quantitative comparison: Table 9 summarizes the mIoU, IoU, OMISSION, COMMISSION, and OA of all the methods. We underline the best results under each metric. As

can be seen, on the Bolivia test set, the method proposed in this paper increases mIoU by over 5% in surface water, permanent water, as well as temporary water segmentation. Our model achieves the best result on all tasks in COMMISSION and OA. For IoU, the proposed method improves over 7% on all tasks. Our method performs well on the Bolivia test set in terms of OMISSION.

Table 9. Performance comparison with other methods on Bolivia datasets. Results shown on all water (AW), permanent water (PW), and temporary water (TW). xx indicates the best performance.

Task	Method	mIoU(%)	IoU(%)	OMISSION(%)	COMMISSION(%)	OA(%)
All Water	Otsu	48.22	70.58	12.85	4.58	93.64
	FCN-ResNet50	43.52	57.32	10.19	10.68	89.53
	Deeplab v3	44.46	68.11	10.28	5.98	92.89
	U^2Net	40.11	56.60	9.10	11.43	88.53
	BASNet	54.07	78.90	7.66	3.21	95.79
Permanent Water	Otsu	36.26	35.20	55.36	5.25	93.47
	FCN-ResNet50	32.98	22.64	16.60	8.37	91.43
	Deeplab v3	36.09	34.50	11.83	4.86	93.08
	U^2Net	40.13	34.06	20.12	4.20	95.55
	BASNet	75.27	78.80	18.60	0.10	99.39
Temporary Water	Otsu	42.23	69.28	13.35	4.58	93.64
	FCN-ResNet50	36.21	38.72	10.48	23.10	78.74
	Deeplab v3	31.40	54.74	29.77	4.98	90.80
	U^2Net	35.29	53.06	7.47	13.10	87.60
	BASNet	47.88	76.74	9.85	3.08	95.59

Qualitative comparison: Figure 7 depicts the flooding maps of each model (binary maps for the Otsu method, probability maps for FCN-ResNet50, Deeplab, U^2Net, and our model) on some samples from Sen1Floods11 Bolivia test set. In some challenging scenes, such as low-contrast foreground and cloud occlusion areas, our method still obtains robust results.
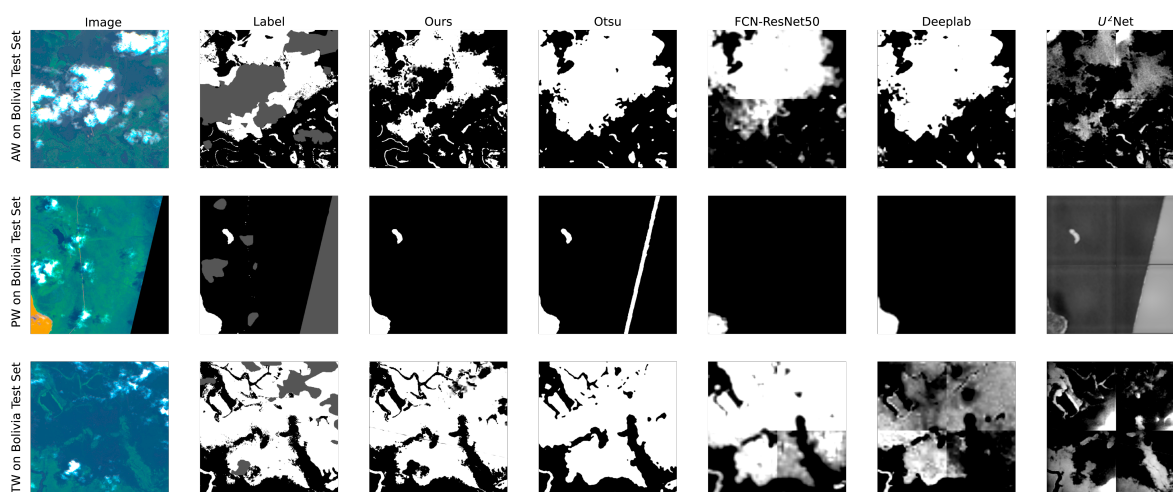

Figure 7. Qualitative comparison of proposed method with other methods. Each row represents one image and corresponding flooding maps (binary maps for the Otsu method, probability maps for FCN-ResNet50, Deeplab, U^2Net, and our model). Each column represents one method. Results shown on all water (AW), permanent water (PW), and temporary water (TW).

5. Discussion

Data augmentation can improve the generalization ability of the model, especially when there is only a little training data. Since the Bolivia test set consists of completely unknown flood event images, the gains indicate that the model's generalization ability has been improved. Focal loss [49] is designed to deal with the extreme imbalance between water/non-water, difficult/easy pixels during training. The ablation experimental results show the effectiveness of focal loss in dealing with a sample imbalance problem. Optical imagery contains information on the ground surface's multispectral reflectivity, which is widely used in water indices and thresholding methods. Image fusion aims to use optical image data to assist SAR image prediction. Our experimental results demonstrate that optical imagery can provide useful supplementary information on water segmentation.

However, all water and temporary water have poor mIoU scores. There are two reasons to explain this phenomenon. One reason is the training process, as the image of all water and temporary water data have more water pixels and fewer non-water pixels. The difference in the sample size leads to differences in the learning effect. From OMISSION and COMMISSION, we can see that the all water and temporary water tasks perform better than permanent water on water pixels, and perform worse on non-water pixels. On the whole, non-water pixels dominate our data. The poorer prediction of non-water pixels leads to worse overall results. The other reason is the difference in image characteristics, and the images in all water and temporary water data contain more small tributaries and scattered areas from newly flooded areas. These areas are usually more challenging to identify.

With the help of hybrid loss, our model pays more attention to boundary pixels and increasing the confidence of the prediction. As a result, our method can not only produce richer details and sharper boundaries but also distinguish water and non-water pixels with a larger probability gap. The excellent feature extraction ability of deep learning model enables our model to deal with some challenging scenes.

6. Conclusions

In this paper, we developed an efficient model for detecting permanent water and temporary water in flood disasters by fusing Sentinel-1 and Sentinel-2 Imagery using a deep learning algorithm with the help of benchmark Sen1Floods11 datasets. The BASNet network adopted in this can capture both large-scale and detailed structural features. By combining with focal loss, our model achieved state-of-the-art accuracy for hard boundary pixels' identification. The model's performance was further improved by fusing the multi-source information, and the ablation study verified the effectiveness of each improvement measures. The comparison experiment results demonstrated that the implemented method could detect permanent water and temporary water flood more accurately than other methods. The proposed model performed well on the unknown Bolivia test set, which verifies its robustness. Due to the network architecture's modularity, it can be easily adapted to data from other sensors. Finally, the method does not require prior knowledge, additional data pre-processing, and multi-temporal data, which significantly reduces the method's complexity and increases the degree of automation.

Ongoing and future works focus on training water segmentation models on high spatial resolution remote sensing imagery. High spatial resolution remote sensing imagery has more complex background information, objects with larger-scale variation, and more unbalanced pixel classes [58]. More sophisticated modules are required to extract and fuse richer image information. In addition, the existing pre-trained neural networks are all based on RGB images, and, directly applied to remote sensing images, may reduce the efficiency of transfer learning due to differences in the data distribution. McKay et al. [61] dealt with this problem by discarding deep feature layers. Qin et al. [60] designed a network that allows for training from scratch, but this lighter network may degrade the performance. Although we dramatically improved the results of flood mapping, there is still much work to do.

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

Abbreviations

The following abbreviations are used in this manuscript:

AW	All Water
BCE	Binary Cross Entropy
CNN	Convolutional Neural Network
COMMISSION	Commission Error Rates
DR	Distributional Ranking
ESA	European Space Agency
FCNN	Fully Convolutional Neural Network
GEE	Google Earth Engine
HAND	Height above Nearest Drainage
IoU	Intersection over Union
JRC	European Commission Joint Research Centre
mIoU	mean Intersection over Union
MNDWI	Modified Normalized Difference Water Index
NDFI	Normalized Difference Flood Index
NDFVI	Normalized Difference Flood in Vegetated Areas Index
NDVI	Normalized Difference Vegetation Index
OA	Overall Accuracy
OMISSION	Omission Error Rates
PW	Permanent Water
RAPID	RAdar-Produced Inundation Diary
ReLU	Rectified Linear Unit
RRM	Residual Refinement Module
SAR	Synthetic Aperture Radar
SOD	Salient Object Detection
SSIM	Structural SIMilarity
TW	Temporary Water
TOA	Top of Atmosphere

References

1. IFRC. World Disaster Report 2020. Available online: https://media.ifrc.org/ifrc/world-disaster-report-2020/ (accessed on 18 January 2021).
2. IDMC. Global Report on Internal Displacement. Available online: https://www.internal-displacement.org/sites/default/files/publications/documents/2019-IDMC-GRID.pdf (accessed on 17 January 2021).
3. Aon. Weather, Climate & Catastrophe Insight 2019 Annual Report. Available online: http://thoughtleadership.aon.com/Documents/20200122-if-natcat2020.pdf?utm_source=ceros$&$utm_medium=storypage$&$utm_campaign=natcat20 (accessed on 18 January 2021).
4. FAO. The State of Food Security and Nutrition in the World. Available online: http://www.fao.org/3/I9553EN/i9553en.pdf (accessed on 18 January 2021).
5. Bonafilia, D.; Tellman, B.; Anderson, T.; Issenberg, E. Sen1Floods11: A Georeferenced Dataset to Train and Test Deep Learning Flood Algorithms for Sentinel-1. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 210–211.
6. Mason, D.C.; Speck, R.; Devereux, B.; Schumann, G.J.P.; Neal, J.C.; Bates, P.D. Flood Detection in Urban Areas Using TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 882–894. [CrossRef]
7. Alfieri, L.; Cohen, S.; Galantowicz, J.; Schumann, G.J.; Trigg, M.A.; Zsoter, E.; Prudhomme, C.; Kruczkiewicz, A.; de Perez, E.C.; Flamig, Z.; et al. A global network for operational flood risk reduction. *Environ. Sci. Policy* **2018**, *84*, 149–158. [CrossRef]
8. Zajic, B. How flood mapping from space protects the vulnerable and can save lives. *Planet Labs* **2019**, *17*.
9. Oddo, P.C.; Bolten, J.D. The value of near real-time earth observations for improved flood disaster response. *Front. Environ. Sci.* **2019**, *7*, 127. [CrossRef]
10. Enenkel, M.; Osgood, D.; Anderson, M.; Powell, B.; McCarty, J.; Neigh, C.; Carroll, M.; Wooten, M.; Husak, G.; Hain, C.; et al. Exploiting the convergence of evidence in satellite data for advanced weather index insurance design. *Weather. Clim. Soc.* **2019**, *11*, 65–93. [CrossRef]
11. Okada, G.; Moya, L.; Mas, E.; Koshimura, S. The Potential Role of News Media to Construct a Machine Learning Based Damage Mapping Framework. *Remote Sens.* **2021**, *13*, 1401. [CrossRef]
12. Martinis, S.; Twele, A.; Voigt, S. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 303–314. [CrossRef]
13. Mahoney, C.; Merchant, M.; Boychuk, L.; Hopkinson, C.; Brisco, B. Automated SAR Image Thresholds for Water Mask Production in Alberta's Boreal Region. *Remote Sens.* **2020**, *12*, 2223. [CrossRef]
14. Tiwari, V.; Kumar, V.; Matin, M.A.; Thapa, A.; Ellenburg, W.L.; Gupta, N.; Thapa, S. Flood inundation mapping- Kerala 2018; Harnessing the power of SAR, automatic threshold detection method and Google Earth Engine. *PLoS ONE* **2020**, *15*. [CrossRef]
15. Otsu, N. Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
16. Bioresita, F.; Puissant, A.; Stumpf, A.; Malet, J.P. Fusion of Sentinel-1 and Sentinel-2 image time series for permanent and temporary surface water mapping. *Int. J. Remote Sens.* **2019**, *40*, 9026–9049. [CrossRef]
17. Conde, F.C.; Munoz, M.D. Flood Monitoring Based on the Study of Sentinel-1 SAR Images: The Ebro River Case Study. *Water* **2019**, *11*, 2454. [CrossRef]
18. Huang, M.M.; Jin, S.G. Rapid Flood Mapping and Evaluation with a Supervised Classifier and Change Detection in Shouguang Using Sentinel-1 SAR and Sentinel-2 Optical Data. *Remote Sens.* **2020**, *12*, 2073. [CrossRef]
19. Markert, K.N.; Chishtie, F.; Anderson, E.R.; Saah, D.; Griffin, R.E. On the merging of optical and SAR satellite imagery for surface water mapping applications. *Results Phys.* **2018**, *9*, 275–277. [CrossRef]
20. 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]
21. DeVries, B.; Huang, C.Q.; Armston, J.; Huang, W.L.; Jones, J.W.; Lang, M.W. Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. *Remote Sens. Environ.* **2020**, *240*. [CrossRef]
22. Rudner, T.G.; Rußwurm, M.; Fil, J.; Pelich, R.; Bischke, B.; Kopačková, V.; Biliński, 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; Volume 33, pp. 702–709.
23. Schlaffer, S.; Matgen, P.; Hollaus, M.; Wagner, W. Flood detection from multi-temporal SAR data using harmonic analysis and change detection. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 15–24. [CrossRef]
24. Twele, A.; Cao, W.X.; Plank, S.; Martinis, S. Sentinel-1-based flood mapping: a fully automated processing chain. *Int. J. Remote Sens.* **2016**, *37*, 2990–3004. [CrossRef]
25. Schlaffer, S.; Chini, M.; Giustarini, L.; Matgen, P. Probabilistic mapping of flood-induced backscatter changes in SAR time series. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *56*, 77–87. [CrossRef]
26. Amitrano, D.; Di Martino, G.; Iodice, A.; Riccio, D.; Ruello, G. Unsupervised Rapid Flood Mapping Using Sentinel-1 GRD SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3290–3299. [CrossRef]
27. Moya, L.; Endo, Y.; Okada, G.; Koshimura, S.; Mas, E. Drawback in the change detection approach: false detection during the 2018 western Japan floods. *Remote Sens.* **2019**, *11*, 2320. [CrossRef]
28. Moya, L.; Mas, E.; Koshimura, S. Learning from the 2018 Western Japan Heavy Rains to Detect Floods during the 2019 Hagibis Typhoon. *Remote Sens.* **2020**, *12*, 2244. [CrossRef]

29. Bai, Y.; Gao, C.; Singh, S.; Koch, M.; Adriano, B.; Mas, E.; Koshimura, S. A framework of rapid regional tsunami damage recognition from post-event TerraSAR-X imagery using deep neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 43–47. [CrossRef]

30. Bai, Y.; Mas, E.; Koshimura, S. Towards operational satellite-based damage-mapping using u-net convolutional network: A case study of 2011 tohoku earthquake-tsunami. *Remote Sens.* **2018**, *10*, 1626. [CrossRef]

31. Kang, W.; Xiang, Y.; Wang, F.; Wan, L.; You, H. Flood detection in gaofen-3 SAR images via fully convolutional networks. *Sensors* **2018**, *18*, 2915. [CrossRef]

32. 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]

33. Chen, L.; Zhang, P.; Xing, J.; Li, Z.; Xing, X.; Yuan, Z. A Multi-Scale Deep Neural Network for Water Detection from SAR Images in the Mountainous Areas. *Remote Sens.* **2020**, *12*, 3205. [CrossRef]

34. Wangchuk, S.; Bolch, T. Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier: Strengths and challenges. *Sci. Remote Sens.* **2020**, *2*, 100008. [CrossRef]

35. Zhang, Y.; Zhang, G.; Zhu, T. Seasonal cycles of lakes on the Tibetan Plateau detected by Sentinel-1 SAR data. *Sci. Total Environ.* **2020**, *703*, 135563. [CrossRef]

36. Sunkara, V.; Purri, M.; Saux, B.L.; Adams, J. Street to Cloud: Improving Flood Maps With Crowdsourcing and Semantic Segmentation. *arXiv* **2020**, arXiv:2011.08010.

37. Muñoz, D.F.; Muñoz, P.; Moftakhari, H.; Moradkhani, H. From Local to Regional Compound Flood Mapping with Deep Learning and Data Fusion Techniques. *Sci. Total Environ.* **2021**, 146927. [CrossRef]

38. Bai, Y.; Hu, J.; Su, J.; Liu, X.; Liu, H.; He, X.; Meng, S.; Mas, E.; Koshimura, S. Pyramid Pooling Module-Based Semi-Siamese Network: A Benchmark Model for Assessing Building Damage from xBD Satellite Imagery Datasets. *Remote Sens.* **2020**, *12*, 4055. [CrossRef]

39. Su, J.; Bai, Y.; Wang, X.; Lu, D.; Zhao, B.; Yang, H.; Mas, E.; Koshimura, S. Technical Solution Discussion for Key Challenges of Operational Convolutional Neural Network-Based Building-Damage Assessment from Satellite Imagery: Perspective from Benchmark xBD Dataset. *Remote Sens.* **2020**, *12*, 3808. [CrossRef]

40. FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. *arXiv* **2020**, arXiv:2012.02951.

41. Konapala, G.; Kumar, S. Exploring Sentinel-1 and Sentinel-2 Diversity for Flood Inundation Mapping Using Deep Learning. Technical Report. Copernicus Meetings. 2021. https://doi.org/10.5194/egusphere-egu21-10445 (accessed on 4 March 2021).

42. Li, Y.; Martinis, S.; Plank, S.; Ludwig, R. An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 123–135. [CrossRef]

43. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. Basnet: Boundary-aware salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7479–7489.

44. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.

45. Hahnloser, R.H.; Seung, H.S.; Slotine, J.J. Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.* **2003**, *15*, 621–638. [CrossRef] [PubMed]

46. 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.

47. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.

48. 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. Intell. Mach.* **2017**, *40*, 834–848. [CrossRef]

49. Goyal, P.; Kaiming, H. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.

50. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2; pp. 1398–1402.

51. Máttyus, G.; Luo, W.; Urtasun, R. Deeproadmapper: Extracting road topology from aerial images. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3438–3446.

52. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

53. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.

54. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

55. Banko, G. *A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory*; International Institute for Applied Systems Analysis: Laxenburg, Austria, 1998.

56. Chang, C.H.; Lee, H.; Kim, D.; Hwang, E.; Hossain, F.; Chishtie, F.; Jayasinghe, S.; Basnayake, S. Hindcast and forecast of daily inundation extents using satellite SAR and altimetry data with rotated empirical orthogonal function analysis: Case study in Tonle Sap Lake Floodplain. *Remote Sens. Environ.* **2020**, *241*. [CrossRef]
57. Qian, Q.; Chen, L.; Li, H.; Jin, R. DR loss: Improving object detection by distributional ranking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 12164–12172.
58. Zheng, Z.; Zhong, Y.; Wang, J.; Ma, A. Foreground-Aware Relation Network for Geospatial Object Segmentation in High Spatial Resolution Remote Sensing Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 4096–4105.
59. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
60. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]
61. McKay, J.; Gag, I.; Monga, V.; Raj, R.G.; Ieee. What's Mine is Yours: Pretrained CNNs for Limited Training Sonar ATR. *arXiv* **2017**, arXiv:1706.09858.