AUTOMATIC FLOOD DETECTION FROM SENTINEL-1 DATA USING DEEP LEARNING ARCHITECTURES

B. Ghosh[1,4,]*, S. Garg[2], M. Motagh[1,3]

[1] GFZ German Research Centre for Geosciences, Department of Geodesy, Section of Remote Sensing, 14473 Potsdam, Germany
[2] Future Infrastructure and Built Environment, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom
[3] Institut of Photogrammetry and GeoInformation (IPI), Leibniz University Hannover, Germany
[4] Department of Computer Science, Technical University Berlin, Berlin, Germany

KEY WORDS: Flood detection, NASA, synthetic aperture radar, transfer learning, deep learning, UNet, Feature Pyramid Network (FPN)

ABSTRACT:

Floods are the most frequent, costliest natural disasters having devastating consequences on people, infrastructure, and the ecosystem. During flood events near real-time satellite imagery has proven to be an efficient management tool for disaster management authorities. However one of the challenges is accurate classification and segmentation of flooded water. The generalization ability of binary segmentation using threshold split-based method, is limited due to the effects of backscatter, geographical area, and time of image collection. Recent advancements in deep learning algorithms for image segmentation has demonstrated excellent potential for improving flood detection. However, there have been limited studies in this domain due to the lack of large scale labeled flood event dataset. In this paper, we present two deep learning approaches, first using a UNet and second, using a Feature Pyramid Network (FPN), both based on a backbone of EfficientNet-B7, by leveraging publicly available Sentinel-1 dataset provided jointly by NASA Interagency Implementation and Advanced Concepts Team, and IEEE GRSS Earth Science Informatics Technical Committee. The dataset covers flood events from Nebraska, North Alabama, Bangladesh, Red River North, and Florence. The performances of both networks were evaluated with multiple training, testing, and validation. During testing, the UNet model achieved the meanIOU score of 75.06% and the FPN model achieved the meanIOU score of 75.76%.

1. INTRODUCTION

Flooding is a widespread and dramatic natural disaster that affects lives, infrastructures, economics and local ecosystems all over the world. Floods often cause loss of life and substantial property damage. Moreover, the economic ramifications of flood damage disproportionately impact the most vulnerable members of society. Due to their imaging capabilities that allow data acquisition regardless of illumination and weather conditions, satellite Synthetic Aperture Radar (SAR) data have become the most widely used Earth Observation (EO) data for operational flood monitoring (Martinis et al., 2015). However, current operational services are mainly focused on inundated rivers and/or mapping of water extent in rural areas, where the specular reflection occurring on smooth water surfaces results in most cases in a dark tone in SAR data, making floodwater distinguishable from other land surfaces. In this context, the German Aerospace Center (DLR) developed a TerraSAR-X (Martinis et al., 2015) and Sentinel-1 Flood Service (Twele et al., 2016) for the automatic near-real time monitoring of rural flooded surfaces using hierarchical tile-based thresholding and fuzzy logic-based post-classification refinement.

Urban areas with low slopes and a high percentage of impervious surfaces are vulnerable to flooding. The increased risk of loss of human lives and damage to economic infrastructure makes urban flood mapping greatly valuable in terms of disaster risk reduction. However, flood detection in urban areas using SAR data is challenging due to the complex backscatter mechanisms associated with varying building types and heights,

vegetation areas, and different road topologies. Except for a few research studies, the full potential of SAR data for operational flood monitoring in urban and vegetated regions are not fully exploited yet. This is a very demanding task, considering the availability of a vast amount of Sentinel-1 data that have been globally available since October 2014. In practice, however, systems that routinely analyse the full potential of Sentinel-1 data for flood detection do not exist.

Flood mapping algorithms are usually based on thresholding algorithms, such as Otsu thresholding (Otsu, n.d.) and histogram leveling, for the initialization of the classification process in SAR amplitude data. These thresholding processes are followed by clustering techniques like K-means (Macqueen, n.d.) or ISODATA (Memarsadeghi et al., 2007) for improving the segmentation of water and non-water areas. These methods are capable of extracting the flood extent if there is a significant contrast between water and non-water areas in the SAR data. However, the result may lead to false positives (overestimation), if non-water areas are characterized by a similar low backscatter as open water surfaces. The main aim of our approach is to develop an automated system capable of extracting and detecting flooded areas in a very short time for near-real time for generation of flood maps for rapid response activities in case of flood emergencies.

However one of the challenges is accurate classification and segmentation of flooded water and permanent water. Binary segmentation using the threshold split-based method is commonly used in this regard, however, the generalization ability of this method is limited due to the effects of backscatter, geo-

---

* Corresponding author

graphical area, and time of image collection. Recent advances in computer vision and the rapid increase of commercially and publicly available medium and high resolution satellite imagery have given rise to a new era area of research at the interface between machine learning and remote sensing. For flood mapping applications, techniques like Bayesian network fusion, and deep convolutional networks have been applied for extraction of flooded areas from optical as well as SAR satellite images, although there have been limited studies in this domain due to the lack of large scale labeled flood event dataset (Li et al., 2019), (Wieland and Martinis, 2019).

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 (Bai et al., 2018). Recently, the development of deep learning in the image processing field, especially deep convolutional neural networks (DCNNs), has enabled the development of new methods for automated extraction of flood extent from SAR images, as proposed in (Zhang et al., 2019), (Katiyar et al., 2021). The latest research focuses on the application of deep learning algorithms for enhancing flood water detection (Kang et al., 2018), (Li et al., 2019). Early research focused on the extraction of surface water (Chen et al., 2020), (Wangchuk and Bolch, 2020), (Zhang et al., 2020). (Isikdogan et al., 2017) 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. The semantic segmentation-based flood extraction method was further applied to identify the flood inundation caused by mounting destruction (Sunkara et al., 2020). Experimental results validate the efficiency and effectiveness of the proposed method. (Muñoz et al., 2021) 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, which brings us to the real problem of near-real-time flood mapping with deep learning techniques: the absence of a SAR-based global flood dataset that provides enough diversity to generalize the model.

Recent development in earth observation has contributed a series of open-sourced large scale disaster related satellite imagery datasets (Bonafilia et al., 2020), which has greatly spurred the advance of leveraging deep learning algorithm for disaster mapping from satellite imagery. For building damage classification, the xBD dataset (Gupta et al., 2021) has provided large scale satellite imagery data that collected from the multitype 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 (Bai et al., 2020), (Su et al., 2020). For flooded building damage assessment in Hurricane disaster events, FloodNet provides a high-resolution UAV image dataset and has done the same task (Rahnemoonfar et al., 2020). The recent release of the large-scale open-source Sen1Floods11 dataset (Bonafilia et al., 2020) is boosting the research of utilizing deep learning algorithms for water type detection in flood disasters (Konapala and Kumar, 2021).

In this work, our aim is to design models and train them on the labelled flood data from some specific geographical regions and then test the performances of the trained models on the data from the other geographical regions. This is to test if it is possible to detect flood in certain parts of the world, even though the model has been trained on flood data from a completely different geographical area.

In this paper, we utilize the publicly available Sentinel-1 dataset provided jointly by NASA Interagency Implementation and Advanced Concepts Team and IEEE GRSS Earth Science Informatics Technical Committee. The dataset is composed of 66,810 tiles of 256×256 pixels, and cover flood events from Nebraska, North Alabama, Bangladesh, and Florence. For our analysis, we compare two convolutional neural networks (CNN), one a Unet and the other an FPN architecture, both based on a backbone of EfficientNetb7. The performance of both networks were evaluated with multiple training, testing, and validation.

This paper is organized as follows. First, the details of the NASA dataset are discussed, along with the test site details and the data used for the test site. Then, the network architectures and training strategy are elaborated. After this, the testing steps, as well as validation data generation on the test site and the performance measures used in the study, are discussed. Finally, the results of the different models' performance are discussed using statistical metrics. The current study should help guide the remote sensing community in developing robust strategies for model development, and model validation.

2. MATERIALS AND METHODS

2.1 Dataset

To use artificial intelligence applications for earth observation datasets we require a huge amount of benchmark datasets to train and test their performance and effectiveness. However, currently, there is a scarcity of benchmark datasets in the remote sensing community (some reference). To address this issue, NASA Disaster team in collaboration with Alaska SAR Facility - Distributed Active Archive Centers (ASF -DAAC) who are specialists in synthetic aperture radar (SAR) data collection, processing, archiving, and distribution, organized a data science challenge for flood extent mapping. To label the SAR datasets provided by ASF DAAC, NASA IMPACT Machine learning team (NASA's IMPACT Collaborates on Global Flood Detection Challenge | Earthdata, n.d.) coordinated with students across the world and guided the students on a weekly basis to generate the flood extent datasets. These labeled datasets provide the necessary 'truth' for developing, validating, and comparing various algorithms and also maximize the potential use of earth-observation data for Artificial Intelligence applications (ETCI 2021 Competition on Flood Detection, n.d.).

The dataset is quite diverse and more representative of the different variations of geographical areas which were affected by flood, including agricultural land and urban settings. The dataset covered five flood events from Nebraska, North Alabama, Bangladesh, and Florence. A total of 54 Geotiff images, (total size 5.3 Gigabytes) were converted into tiles of 256×256 pixels. More comprehensive details about the flood evenets are detailed in Table 1.

Each tile includes 3 RGB channels generated from Sentinel-1 C-band synthetic aperture radar (SAR) imagery data acquired

Table 1. Table showing the different flood events covered in the dataset

Area name	Total area covered	Flood start date	Flood end date
Bangladesh	7,150 sq. km.	14.03.2017	12.07.2017
Nebraska	1,741 sq. km.	08.01.2017	22.12.2017
North Alabama	13,789 sq. km.	02.03.2019	27.12.2019
Florence	7,197 sq. km.	01.10.2018	05.10.2018

in Interferometric Wide Swath mode in 5m * 20m resolution using Hybrid Pluggable Processing Pipeline "hyp3". The "hyp3" system takes the Sentinel archive and creates a set of processes to get to a consistent method of generating the VV-VH amplitude or power imagery. The imagery is then converted to a 0 - 255 grayscale image. The whole dataset consists of approximately 66000 tiled images from these various geographic locations. The dataset was split across 29 root folders named region_datetime, region being the region and datetime being the date and time of the flood event. Each root folder includes 4 sub-folders: VV, VH, flood_label and water_body_label with 2,068 files each. VV and VH correspond to the VV and VH bands of the satellite images and images in the flood_label and water_body_label folder provide reference ground truth (NASA's IMPACT Collaborates on Global Flood Detection Challenge | Earthdata, n.d.).

The Bangladesh geographic area which is predominant in the dataset, is primarily an agricultural hub and recently harvested fields can look similar to floods due to low backscatter in both VV and VH polarizations. Similarly the dataset from Florence has a primarily urban setting. Such varying backscatter is relevant for performance optimizations and generalizability to test imagery. A sample image from the dataset is depicted in Figure 1.
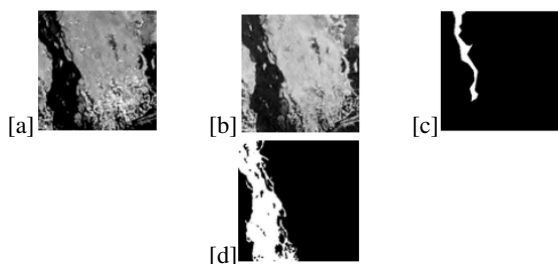

Figure 1. Visual example of SAR image and corresponding label mask in the dataset: (a) Sentinal-1 VV (b) Sentinal-1 VH (c) Water body data and (d) Flood label

The image tiles are generated by cropping the Geotiffs. Owing to the viewing geometry, there are some artifacts particularly at the edges- where the sentinel-1 Geotiffs do not exactly align with sliding cropping window resulting in noisy tiles. These data are then filtered from the whole dataset. After filtering, the remainder dataset consists of 33,405 image tiles covering flood events from all the geographical areas under consideration. The water body channel of each data tile was replaced by the channel containing the value of (vv+VH)/(VV-VH) for every pixel. This reason behind this was for better adaptation of the model so that only the VV and VH bands of Sentinel-1 SAR data can be used for flood detection, without the necessity of a separate water body layer, which sometimes maybe difficult to obtain.

In this work, our aim is to design models and train them on the labelled flood data from some of these geographical regions and then test the performances of the trained models on the data from the other remaining geographical regions. This is to test if it possible to detect flood in certain parts of the world, even though the model has been trained on flood data from a completely different geographical area.

For our evaluation, we keep the entire Florence dataset as our test set. 8382 images from the Florence dataset is separated from the rest of the data, and the remaining approximately 25000 image tiles are used for training. The remaining 25000 image tiles are shuffled randomly and split into two parts, taking 75% for training and 25% for validation.

2.2 Model Architecture

In this work, two network combinations of a densely supervised encoder and decoder are applied for semantic segmentation of flooded areas from the dataset. The encoder–decoder network can fuse abstract high-level information and detailed low-level information and is mainly responsible for water body segmentation (Bai et al., 2021). Unet (Ronneberger et al., 2015), (Bizopoulos et al., 2021) combines an encoder that scales down the features to a lower dimensional bottleneck and a decoder that scales them up to original dimensions. It also uses skip connections that were proven to improve image segmentation results (Drozdzal et al., 2016). Feature Pyramid Network (FPN) (Lin et al., 2017) is also similar to Unet with the difference of applying a 1 × 1 convolution layer and adding the features instead of copying and appending them as done in the Unet architecture. UNet is one of the most fundamental semantic segmentation networks. It was originally intended to be used on biomedical images, however it finds increasing relevance in nearly all areas of interest today including remote sensing (Hu et al., 2020). In case of UNet, the encoder is used for multi-level feature extraction and the decoder combines learnt features and resolution through a sophisticated stacking, taking both localization and feature representation into account (Ronneberger et al., 2015). On the other hand, FPN works by creating two pyramids, and combines them to generate feature-rich segmentation maps at each level (Figure 2).
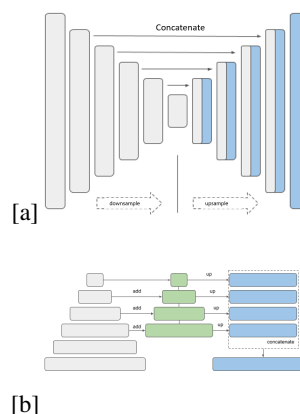

Figure 2. Segmentation models investigated in the study: (a) UNet (b) FPN

2.3 Encoder Architecture

As mentioned in the paper (Tan and Le, 2020), Convolutional Neural Networks are commonly developed at a fixed resource

budget, and then scaled up for better accuracy if more resources are available. To maximize the model accuracy for any given resource constraints, an optimazation problem can be designed with parameters w, d, r, which are coefficients for scaling network width, depth, and resolution. The authors of the paper (Tan and Le, 2020) also use neural architecture search to design a new baseline network and scale it up to obtain a family of models, called EfficientNets, which achieve much better accuracy and efficiency than previous Convolutional Networks.

In the paper(Tan and Le, 2020), the authors introduce a new compound scaling method, which use a compound coefficient ϕ to uniformly scale network width (β^{ϕ}), depth (α^{ϕ}), and resolution (γ^{ϕ}), where α, β, γ are constants that can be determined by a small grid search. Intuitively, ϕ is a user-specified coefficient that controls how many more resources are available for model scaling, while α, β, γ specify how to assign these extra resources to network width, depth, and resolution respectively.
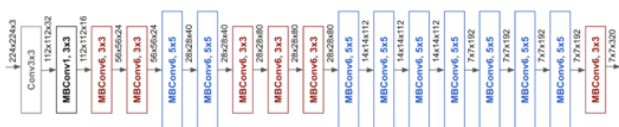

Figure 3. Network architecture of EfficientNet-B0 (Tan and Le, 2020)

Figure 3 shows the architecture of EfficientNet-B0. Its main building block is mobile inverted bottleneck MBConv (Sandler et al., 2019), to which squeeze-and-excitation optimization (Hu et al., 2019) is added. Starting from the baseline EfficientNet-B0, a compound scaling method is added to scale it up with two steps:

- $\phi = 1$ is first fixed, assuming twice more resources available, and a small grid search of α, β, γ is done.

- α, β, γ are fixed as constants and the baseline network is scaled up with different ϕ, to obtain EfficientNet-B1 to B7.

In particular, the EfficientNet-B7 has achieved state-of-the-art 84.3% top-1 accuracy on the ImageNet dataset (Deng et al., 2009), while being 8.4x smaller and 6.1x faster on inference than the best existing convolution network. The EfficientNet also transfers well and achieves state-of-the-art accuracy on CIFAR-100 (91.7%) (Krizhevsky, n.d.), and Flowers (98.8%) datasets (Nilsback and Zisserman, 2008).

2.4 Model training details

For our analysis, we use the EfficientNet-B7 as the encoder backbone for both the Unet as well as the FPN networks. The classifier is retrained on our filtered dataset and the weights of the pre-trained network are also fine-tuned by continuing the back-propogation. For the entire study, the mini-batch size was selected as 10 and iterated over the whole dataset for 14 epochs. The Adam optimizer (Kingma and Ba, 2017) was used for training optimization, with an initial learning rate of 0.01. The loss function used was dice loss for training both Unet as well as the FPN networks. The minimum value for learning has been fixed to 0.0001.

2.5 Model evaluation metrics

In a binary segmentation study such as flood inundation, two outcomes which correspond to water and non-water regions are possible. The output can be classified as (1) True Positive (TP): where water pixels are correctly classified as water; (2) True Negative (TN), where non-water pixels are correctly classified as non-water regions; (3) False Positive (FP): non-water pixels incorrectly classified as water (4) False Negative (FN): water pixels incorrectly classified as non-water (Konapala and Kumar, 2021).

Based on these outputs, pixel accuracy which determines the percentage of pixels correctly classified can be computed. However, as accuracy computes this percentage irrespective of classes, it can be misleading when the class of interest (i.e. water) has relatively low number of pixels. To avoid this, Precision, Recall, and F1 scores are commonly used. Precision and Recall are interdependent measures of over and under-segmentation Low values of Precision and Recall indicates over-segmentation and under-segmentation, respectively. F1 score is the harmonic mean of Precision and Recall scores capturing both the aspects as a single metric. Intersection over Union (IoU) is the ratio between the area of overlap and the area of union between the ground truth and the predicted areas. The mIoU is the average between the IoU of the segmented objects over all the images of the dataset.

Precision illustrates how many of the predicted water pixels matched the water pixels in the annotated labels. It can be calculated as

$$Precision = TP/(TP + FP) \qquad (1)$$

Whereas Recall denotes how many have been predicted as water pixels by our deep learning model. It can be defined as:

$$Recall = TP/(TP + FN) \qquad (2)$$

For an image to be classified accurately, both Precision and Recall should be high. For this purpose, F1 score and mIoU, is often used as a tradeoff metric to quantify both over- and under-segmentation into one measure. While training, a sum of F1 score and mIoU, which is the model evaluation score is used a metric for evaluating the model while training.

$$F1score = 2 * (Precision * Recall)/(Precision + Recall) \qquad (3)$$

$$F1score = TP/(TP + (FP + FN)/2) \qquad (4)$$

$$IoU = TP/(FN + FP + TP) \qquad (5)$$

$$mIoU = IoU/N \qquad (6)$$

where N is the number of images in the dataset.

A modified K-fold cross validation approach was used to evaluate the performance of both models. For each one of two models, the Unet and the FPN, the filtered dataset was randomly divided into 10 equal subsets. For each round, the model was trained on 9 of subsets randomly and validated on the remaining subset. The process was repeated for k=10 by randomly selecting 9 susbsets for training and the remaining one for validation. The model with the highest model evaluation score was selected and its performance was evaluated on the test dataset. As mentioned, this process was repeated for both models separately and the best models from each of the two architectures, the Unet and the FPN were then selected and their performances were assessed on the test dataset. The training was performed on a server with three nVidia GP104GL [Quadro P4000] GPUs, with driver NVIDIA UNIX x86.64 Kernel Module 460.56. The whole model development and training were performed using the Tensorflow platform (Abadi et al., 2015) along with the Keras library in Python.

3. RESULTS

3.1 Training results

As mentioned in section 2.2, a K-fold corss-validation was performed using both models, the Unet and the FPN separately using the filtered training dataset. The two best models, of both architectures were selected based on the highest model evaluation score. The progression of training and validation loss and training and validation IoU, for the best Unet model with Efficient B7 as encoder, over 15 epochs, after K-fold cross-validation, is depicted in Figure 4. Similar curves for the best FPN model, after K-fold cross-validation, is shown in Figure5.
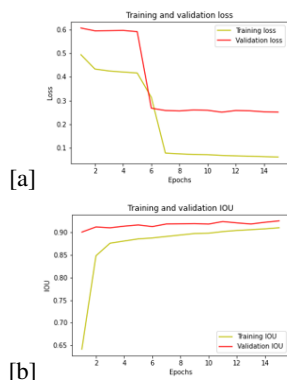

Figure 4. (a) shows the progression of training and validation loss and (b) shows the progression of training and validation IoU, for the Unet model with Efficient B7 as encoder, over 15 epochs.

The results of the performances of the best models of both the Unet and the FPN, after K-fold cross-validation, on a few of the training images are shown in Figure 6.

3.2 Results on test data

As mentioned in section 2.1, 8382 tiles of the Florence dataset was used to test and evaluate the performances of both models. During testing, data augmentation was done by rotating each tile by 90, 180 and 270 degrees and predicting the flood mask by applying the trained model on each of the augmented tile. The final prediction was calculated by averaging the mean of
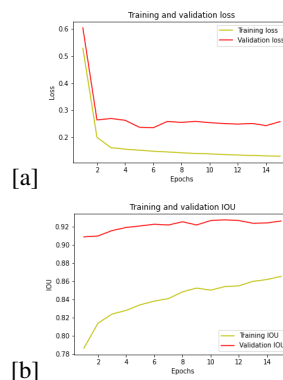

Figure 5. (a) shows the progression of training and validation loss and (b) shows the progression of training and validation IoU, for the FPN model with Efficient B7 as encoder, over 15 epochs.

Performance metric	UNet model	FPN model
Precision	97.2%	97.2%
Recall	95%	97.5%
F1 score	96.1%	97.3%
Mean IOU	75.06%	75.76%

Table 2. Comparison of performance metrics of the best models of the UNet and the FPN on the test data from Florence

all the separate predictions. This whole process is shown in Figure 7.

The results of the performances of the best models of both the UNet and the FPN, after K-fold cross-validation, on a few of the test images from Florence are shown in Figure 8.

Based on the labels from the test data, metrics like precision, recall, F1 score and mean IOU were calculated for both models and the results are depicted in Table 2.

4. DISCUSSION AND CONCLUSION

In this work, the main objective was to leverage the huge amount of publicly available Sentinel-1 dataset to delineate open water bodies which can be further used in flood extent mapping. In this work, two deep learning segmentation models, the UNet and the FPN, both with the same EfficientNet-B7 encoder architecture were designed and trained against a set of labeled SAR datasets from certain geographical areas and then tested for detection of flooded pixels for a different test case in another geographical area. Based on the labels from the test data, metrics like precision, recall, F1 score and mean IOU were calculated for both models and the results are depicted in Table 2.

From the results in Table 2, it can be seen that both models performed really well in identifying the flood pixels. However, the metrics show that the FPN model outperformed the UNet model to some extent. This may be due to the fact that FPN model architecture was able to capture the minute details from the three bands much efficiently compared to the UNet. Also, the results indicate that the EfficientNet-B7 encoder performed really well as is evident from the fact that both models achieved a meanIOU score of more than 75% on the test dataset. It is also shown that our method can enable scalable training with data distribution drifts, as is evident from the fact that both
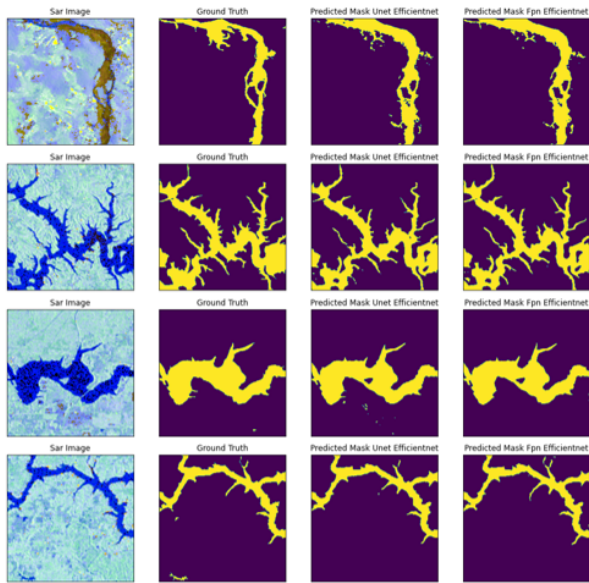
Figure 6. Figure showing the SAR composite image, the corresponding ground truth and the prediction results for both the UNet and the FPN models for the training and validation dataset from Bangladesh, Nebraska and North Alabama flood events

models were trained on data from three geographical areas - Bangladesh, North Alabama and Nebraska and tested on a different dataset from Florence, and in-spite this the performance of both the models were quite commendable.

Further improvements to the models can be made with access to better datasets in the future, such as more specific classes for floods (open floods, flooded vegetation, and urban floods) rather than only one general class. Also, it will be very interesting to evaluate and analyze the results of classification and segmentation of the derived flooded areas, using a visual analytics approach to explain the causality behind the classification of the flooded areas by the deep learning models.

5. ACKNOWLEDGMENTS

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P.,

Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Bai, Y., Gao, C., Singh, S., Koch, M., Adriano, B., Mas, E., Koshimura, S., 2018. A Framework of Rapid Regional Tsunami Damage Recognition From Post-event TerraSAR-X Imagery Using Deep Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 15(1), 43–47. Conference Name: IEEE Geoscience and Remote Sensing Letters.

Bai, Y., Hu, J., Su, J., Liu, X., Liu, H., He, X., Meng, S., Mas, E., Koshimura, S., 2020. Pyramid Pooling Module-Based Semi-Siamese Network: A Benchmark Model for Assessing Building Damage from xBD Satellite Imagery Datasets. *Remote Sensing*, 12(24), 4055. https://www.mdpi.com/2072-4292/12/24/4055. Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

Bai, Y., Wu, W., Yang, Z., Yu, J., Zhao, B., Liu, X., Yang, H., Mas, E., Koshimura, S., 2021. 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. *Remote Sensing*, 13(11), 2220. https://www.mdpi.com/2072-4292/13/11/2220. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Bizopoulos, P., Vretos, N., Daras, P., 2021. Comprehensive Comparison of Deep Learning Models for Lung and COVID-19 Lesion Segmentation in CT scans. *arXiv:2009.06412 [cs, eess]*. http://arxiv.org/abs/2009.06412. arXiv: 2009.06412.

Bonafilia, D., Tellman, B., Anderson, T., Issenberg, E., 2020. Sen1Floods11: a georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 835–845. ISSN: 2160-7516.

Chen, L., Zhang, P., Xing, J., Li, Z., Xing, X., Yuan, Z., 2020. A Multi-Scale Deep Neural Network for Water Detection from SAR Images in the Mountainous Areas. *Remote Sensing*, 12(19), 3205. https://www.mdpi.com/2072-4292/12/19/3205. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. ISSN: 1063-6919.

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C., 2016. The Importance of Skip Connections in Biomedical Image Segmentation. *arXiv:1608.04117 [cs]*. http://arxiv.org/abs/1608.04117. arXiv: 1608.04117.

ETCI 2021 Competition on Flood Detection, n.d.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2021. xbd: A dataset for assessing building damage from satellite imagery.

Hu, J., Li, L., Lin, Y., Wu, F., Zhao, J., 2020. A Comparison and Strategy of Semantic Segmentation on Remote Sensing Images. *arXiv:1905.10231 [cs]*, 1074, 21–29. http://arxiv.org/abs/1905.10231. arXiv: 1905.10231.
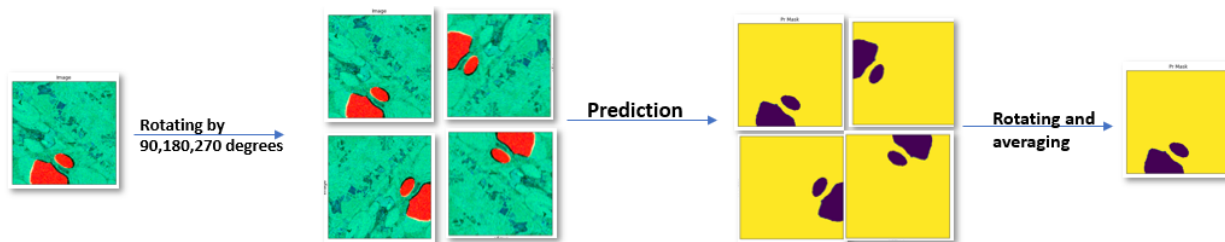
Figure 7. Augmentation of test data by rotating each tile by 90, 180 and 270 degrees and predicting the flood mask by applying the trained model on each of the augmented tile.
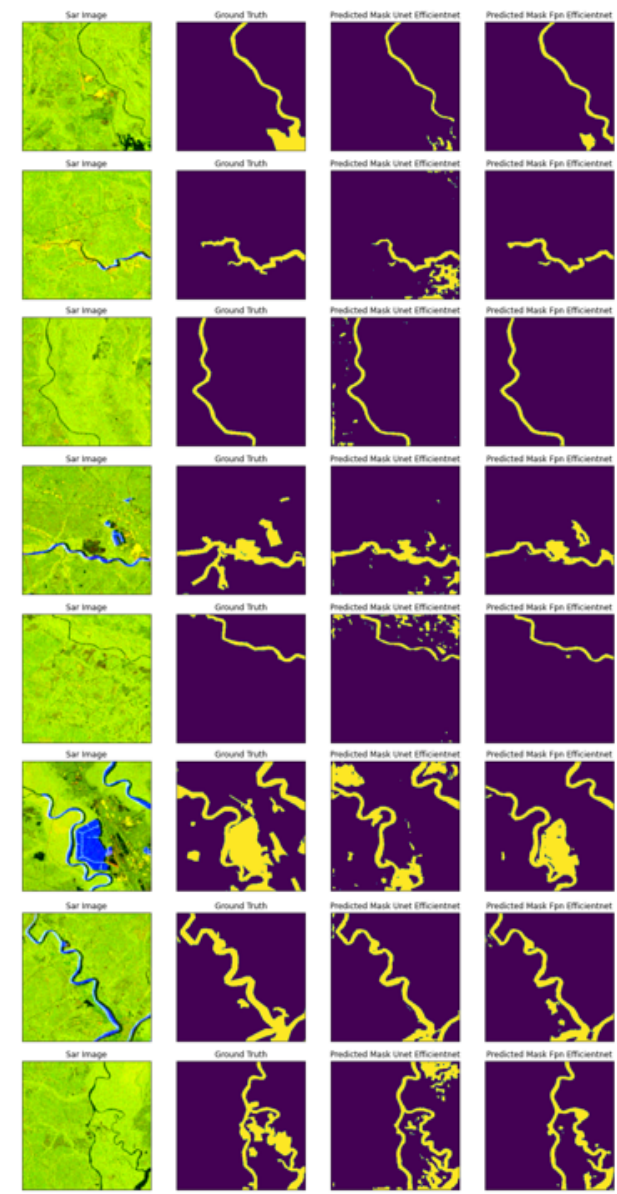

Figure 8. Figure showing the SAR composite image, the corresponding ground truth and the prediction results for both the UNet and the FPN models for the test dataset from Florence flood event

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2019. Squeeze-and-Excitation Networks. *arXiv:1709.01507 [cs]*. http://arxiv.org/abs/1709.01507. arXiv: 1709.01507.

Isikdogan, F., Bovik, A. C., Passalacqua, P., 2017. Surface Water Mapping by Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 4909-4918.

Kang, W., Xiang, Y., Wang, F., Wan, L., You, H., 2018. Flood Detection in Gaofen-3 SAR Images via Fully Convolutional Networks. *Sensors*, 18(9), 2915. https://www.mdpi.com/1424-8220/18/9/2915. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Katiyar, V., Tamkuan, N., Nagai, M., 2021. Near-Real-Time Flood Mapping Using Off-the-Shelf Models with SAR Imagery and Deep Learning. *Remote Sensing*, 13(12), 2334.

Kingma, D. P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. http://arxiv.org/abs/1412.6980. arXiv: 1412.6980.

Konapala, G., Kumar, S., 2021. Exploring Sentinel-1 and Sentinel-2 diversity for Flood inundation mapping using deep learning. Technical Report EGU21-10445, Copernicus Meetings. Conference Name: EGU21.

Krizhevsky, A., n.d. Learning Multiple Layers of Features from Tiny Images. 60.

Li, Y., Martinis, S., Wieland, M., Schlaffer, S., Natsuaki, R., 2019. Urban Flood Mapping Using SAR Intensity and Interferometric Coherence via Bayesian Network Fusion. *Remote Sensing*, 11(19), 2231. https://www.mdpi.com/2072-4292/11/19/2231. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection. *arXiv:1612.03144 [cs]*. http://arxiv.org/abs/1612.03144. arXiv: 1612.03144.

Macqueen, J., n.d. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. *MULTIVARIATE OBSERVATIONS*, 17.

Martinis, S., Kersten, J., Twele, A., 2015. A fully automated TerraSAR-X based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 203–212. https://linkinghub.elsevier.com/retrieve/pii/S0924271614001981.

Memarsadeghi, N., Mount, D. M., Netanyahu, N. S., Le Moigne, J., 2007. A fast implementation of the isodata clustering algorithm. *International Journal of Computational Geometry & Applications*, 17(01), 71–103. https://www.worldscientific.com/doi/abs/10.1142/S0218195907002252. Publisher: World Scientific Publishing Co.

Muñoz, D. F., Muñoz, P., Moftakhari, H., Moradkhani, H., 2021. From local to regional compound flood mapping with deep learning and data fusion techniques. *Science of The Total Environment*, 782, 146927. https://www.sciencedirect.com/science/article/pii/S0048969721019975.

NASA's IMPACT Collaborates on Global Flood Detection Challenge | Earthdata, n.d.

Nilsback, M.-E., Zisserman, A., 2008. Automated flower classification over a large number of classes. *Indian Conference on Computer Vision, Graphics and Image Processing*.

Otsu, N., n.d. A Tlreshold Selection Method from Gray-Level Histograms. 5.

Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., Murphy, R., 2020. FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. *arXiv:2012.02951 [cs]*. http://arxiv.org/abs/2012.02951. arXiv: 2012.02951.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*. http://arxiv.org/abs/1505.04597. arXiv: 1505.04597.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2019. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv:1801.04381 [cs]*. http://arxiv.org/abs/1801.04381. arXiv: 1801.04381.

Su, J., Bai, Y., Wang, X., Lu, D., Zhao, B., Yang, H., Mas, E., Koshimura, S., 2020. Technical Solution Discussion for Key Challenges of Operational Convolutional Neural Network-Based Building-Damage Assessment from Satellite Imagery: Perspective from Benchmark xBD Dataset. *Remote Sensing*, 12(22), 3808. https://www.mdpi.com/2072-4292/12/22/3808. Number: 22 Publisher: Multidisciplinary Digital Publishing Institute.

Sunkara, V., Purri, M., Saux, B. L., Adams, J., 2020. Street to Cloud: Improving Flood Maps With Crowdsourcing and Semantic Segmentation. *arXiv:2011.08010 [cs]*. http://arxiv.org/abs/2011.08010. arXiv: 2011.08010.

Tan, M., Le, Q. V., 2020. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:1905.11946 [cs, stat]*. http://arxiv.org/abs/1905.11946. arXiv: 1905.11946.

Twele, A., Cao, W., Plank, S., Martinis, S., 2016. Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing*, 37(13), 2990–3004. https://www.tandfonline.com/doi/abs/10.1080/01431161.2016.1192304. Publisher: Taylor & Francis.

Wangchuk, S., Bolch, T., 2020. Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier: Strengths and challenges. *Science of Remote Sensing*, 2, 100008. https://www.sciencedirect.com/science/article/pii/S2666017220300079.

Wieland, M., Martinis, S., 2019. A Modular Processing Chain for Automated Flood Monitoring from Multi-Spectral Satellite Data. *Remote Sensing*, 11(19), 2330. https://www.mdpi.com/2072-4292/11/19/2330. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Zhang, P., Chen, L., Li, Z., Xing, J., Xing, X., Yuan, Z., 2019. Automatic Extraction of Water and Shadow from SAR Images Based on a Multi-Resolution Dense Encoder and Decoder Network. *Sensors (Basel, Switzerland)*, 19(16), E3576.

Zhang, Y., Zhang, G., Zhu, T., 2020. Seasonal cycles of lakes on the Tibetan Plateau detected by Sentinel-1 SAR data. *The Science of the Total Environment*, 703, 135563.