DEEP LEARNING BASED URBAN FLOOD MAPPING FROM HIGH RESOLUTION CAPELLA SPACE SAR IMAGERY

Philip Popien, Olivier D'Hondt, Veda Sunkara and Subit Chakrabarti

Floodbase, Brooklyn, USA

ABSTRACT

In this paper, we address the problem of flood detection in urban environments from single Capella Space X-Band High-Resolution Synthetic Aperture Radar (HR SAR) images. Our approach uses a Convolutional Neural Network (CNN) architecture and proposes two innovations to deal with a limited amount of training data and the burden of manual labeling. We introduce a two-stage labeling process utilizing threshold-based predictions to reduce the amount of manual interaction in the annotation process. Moreover, we use several auxiliary inputs to give more context to the CNN. We train our model on 28 flooded and non-flooded scenes and evaluate its performance on 4 out-of-sample test images. We show that our approach outperforms threshold-based segmentation in all test scenes.

Index Terms— Flood Detection, SAR, Deep Learning, Semantic segmentation

1. INTRODUCTION

Flooding is a natural hazard that has a tremendous deleterious impact on the population and causes more damage than any other type of disaster [1]. Due to global climate change, floods are becoming more frequent and of increasing intensity. For instance, in August 2022, flooding in Pakistan resulted in more than 1700 deaths and nearly 40 billion dollars in damage [2]. Readily available remote sensing satellite data allows for scalable, low cost, near-real time disaster monitoring and mapping in order to support response and relief. In this paper we introduce a CNN [3] based method to map urban floods inundation using HR SAR imagery.

SAR is a type of active Earth observation sensor that operates in the microwave domain and produces imagery independent of weather and illumination conditions. For this reason, SAR is a powerful tool to map flood inundation when no optical imagery is available due to persistent cloud cover. Moreover, the resolution of the Capella Space high-resolution imagery modes (between 0.5m and 1.2m) potentially enables the detection of water at a street level which makes these sensors appealing for urban flood mapping and property-level insights. It is well known that water is characterized in SAR images by a low backscatter. Using this principle, backscatter thresholding-based flood-mapping methods have been developed and successfully used in the past [4]. The main limitation of such methods is that they produce false positives in other areas that also appear dark such as asphalt, concrete, and building and terrain shadows. Conversely, when a road is flooded, water cannot be easily distinguished from asphalt, leading to false negatives. Although some of these issues could be mitigated by using change detection methods [5], acquiring pre-event images on a global scale and updating these images regularly to avoid the influence non-flood related changes would be prohibitive in terms of cost and storage.

In this paper, we propose a deep learning based solution that operates on single images to overcome these limitations. CNNs have been very successful at tackling various complex computer vision problems [6]. They have been used for semantic segmentation of remote sensing and, in particular, SAR imagery [7]. CNNs extract high level hierarchical spatial features that lead to better segmentation results than traditional Machine Learning algorithms. In this study, we use the UNet++ [8] architecture that has reported excellent performance on semantic segmentation problems. One shortcoming of CNN-based techniques is that performance critically depends on the number of training samples. In our case, because we are using commercial SAR imagery which is expensive to acquire, the number of images we can use for training is limited. Moreover, creating manual semantic segmentation labels on large amounts of examples is extremely time-consuming.

Our contribution is to address these limitations by introducing two innovations: 1) to decrease the amount of manual annotation, we propose an efficient two-stage labeling process that combines manually drawn labels with thresholding predictions and 2) to cope with the small amount of training data and the backscattering ambiguities inherent to SAR, we introduce several auxiliary input layers to inform the CNN in difficult cases.

2. METHODS

In the present work, we use the Capella Space constellation that acquires high-resolution X-band imagery in HH

The authors would like to thank Capella Space for supporting this research through their Analytics Partner program.

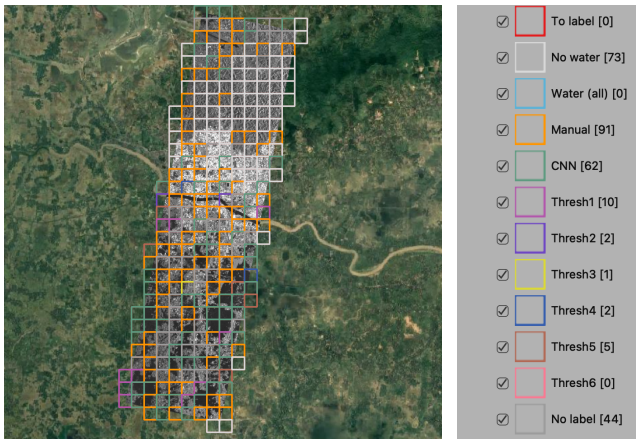

Fig. 1. Pre-labeled image. Only orange chips have to be further manually labeled. Others are labeled using different thresholds or early CNN predictions.

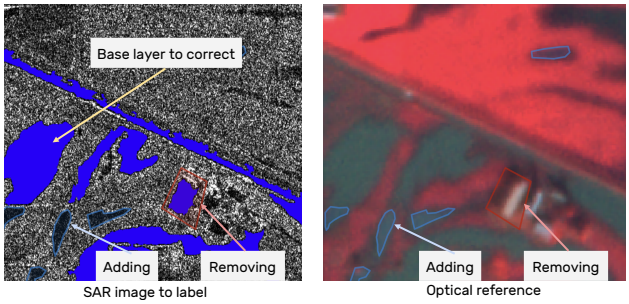

Fig. 2. Manual labeling is done by adding and removing regions to a base prediction. The optical reference is from ©PlanetScope (color coding R = NIR, G = Red, B = Green). Water appears in green.

polarization. To get the largest possible footprints, we use geocoded products in Stripmap mode that covers 5×20km terrain patches and are at approximately 1.2 m resolution. This allows us to resolve water pixels even in dense urban areas. Getting access to a constellation of 7 sensors allows for a theoretical sub-daily revisit time, increasing the chances of observing short duration floods.

2.1. Collection of training data

In order to get the largest possible amount of training data we had two possible sources at hand: 1) the "archive" images that are already acquired and available from Capella's database and 2) the "tasked" images which we have collected by tasking the satellite through Capella's console. For the latter, we were continuously monitoring major flood events based on news and warnings issued by the National Weather Service.

2.2. Data annotation

From a total of 67 acquired Stripmap swaths we have selected 32 images to be annotated based on their flood/permanent water, level of urbanness, and the availability of optical reference data. Manual annotation is a time-consuming task: depending on the content, labeling a single swath can take up to a week.

To speed up the process we have introduced two innovations in the labeling framework. During a preliminary assessment of global thresholding, we have noticed that for many areas, these methods produced satisfactory labels. Therefore, we have introduced a pre-labeling stage prior to the manual labeling task to reduce the amount of data to be annotated by hand. This pre-task consists of the following steps: first, several pseudo-labeled images are computed from global thresholding methods with different parameter settings as well as an early CNN prediction obtained from a few fully manually labeled scenes. The image is divided into 768×768 pixels chips. For each chip, the annotator decides if one of the pseudo-labeled maps can be used as a training label or if the chip needs to be manually labeled. Fig. 1 illustrates this process.

The second stage consists of labeling the chips that have been marked for manual annotation with a GUI based tool developed internally at Floodbase that allows the addition and removal of polygons on a base prediction chosen to be the "overall best" during the pre-labeling task, as illustrated in Fig. 2. This way, manual annotation is only a label correction step which saves a substantial amount of time.

2.3. Deep Learning model and training

The CNN model architecture is based on a U-Net++ architecture [8] with an EfficientNet-B1 [9] backbone pre-trained on ImageNet and an Adam optimizer with an initial learning rate of 10^{-3} and no weight decay. The average of dice and cross-entropy was used as the loss function. Capella provides tasked imagery at a high resolution at the cost of smaller footprints. For this reason, the number of images used for labeling was relatively small, and it is well known that CNNs perform better with a large training dataset [6].

To cope with the relatively small number of training samples we have added additional inputs to the model, to provide the CNN with more context about the scene. The following input channels were used: in addition to the HH Capella dB intensity, we have added a cloud-free and flood-free 10 meter ESA Sentinel-2 optical image prior to the SAR acquisition to introduce a general scene understanding. We used the blue, green, red, NIR, SWIR16, SWIR22 channels which are relevant to water detection. The rich multi-spectral content of such data allows the model to disambiguate low backscatter areas that look like water in SAR data such as roads and some large roofs. The occurrence layer of the Global Surface Water (GSW) model [10] produced by the Joint Research Commission (JRC) indicates how frequently water was observed

Scene	CNN (Average)	Threshold (Average)	CNN (NF+FB)	Threshold (NF+FB)	CNN (PW)	Threshold (PW)
Freeport, USA	87.0	66.7	75.3	49.2	98.7	84.1
Miami Beach, USA	87.7	60.9	75.7	27.9	99.7	94.0
London, UK	53.6	43.9	7.5	1.7	99.6	86.1
Brisbane, Australia	77.1	62.4	55.5	44.5	98.7	80.3

Table 1. CNN and thresholding IoU scores in percentage (see text for details) for our test scenes on Never Flooded and Flooded Before (NF+FB) and Permanent Water (PW) areas. "Average" denotes the mean of NF+FB and PW IoUs.

in each 30 meter pixel historically. The slope of the Copernicus Global Digital Elevation Model (DEM) [11] at 30 meter resolution helps avoid false positives in mountain shadows since water is rarely observed in areas of high slope. Finally, the VH and VV channels of the Sentinel-1 Global Backscatter Model [12] act like a SAR prior for each 10 meter pixel. For example, pixels with low backscatter in Capella and high backscatter in the prior have a higher probability of being flood water.

Our initial 32 image data set was split into: a training-validation dataset with 7968 chips from 28 scenes, and a testing dataset with 1341 chips from 4 withheld scenes. The size of the chips is 768×768 pixels. Then, the training-validation dataset was split into three folds for cross-validation. Chips with a lot of flood water were shown to the model more frequently than chips without flood water during training.

3. RESULTS

For quantitative evaluation, we have computed the Intersection over Union (IoU) score which is the ratio between the area of overlap and the area of union of the reference and predicted labels. This metric has been averaged across our 4 test scenes but computed separately for permanent water (PW) areas (where GSW $> 30\%$) and the combination of "never-flooded" and "flooded before" (NF+FB) areas (where $0 \leq$ GSW $\leq 30\%$). We have compared the output of our model to flood maps from a simple threshold based segmentation. In PW areas, our model's average IoU was 99.2% against 86.1% for thresholding. In NF+FB areas, the model's IoU was 53.6% against 30.8% for thresholding. We can observe that the CNN based method largely outperformed thresholding. However, the performance of both methods significantly drops on NF+FB compared to PW.

Table 1 displays the per scene IoUs on NF+FB and PW areas. These results show that the performance of both methods is relatively stable across scenes for PW areas whereas there is a significant variation in NF+FB area. In the last case we can also notice more disparity between the two methods where the performance gain in using the CNN is higher. The London image shows the worst performance for both methods, although the CNN still outperforms thresholding.

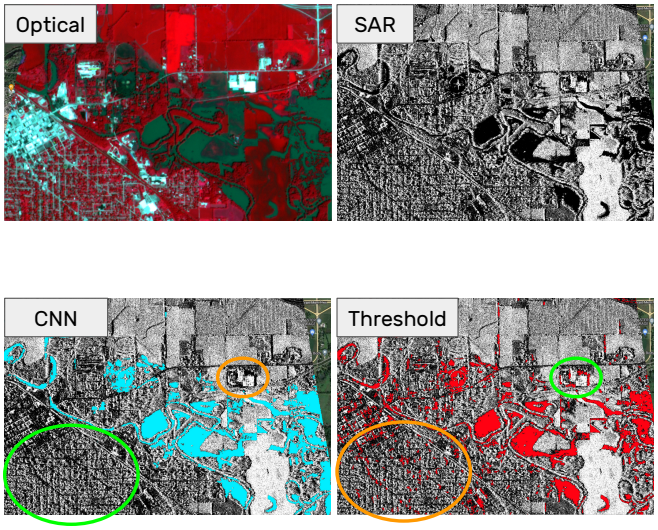

Fig. 3. CNN (cyan) and thresholding (red) predictions on Freeport, USA (crop). The optical reference is from ©PlanetScope (color coding R = NIR, G = Red, B = Green). Water appears in green. Green and orange circles highlight resp. correct and incorrect segmentation.

Fig. 3 and 4 show visual examples of how the CNN performs in different situations. These two examples provide a good summary of the strengths and weaknesses of single-polarization, single-image HR SAR for urban flood mapping: water is accurately detected on open natural areas which are often dark when water is present, but models fall short in detecting water on smooth surfaces like streets and parking lots. Moreover, extreme cases of natural areas with low backscattering result in False Positives. In the case of the London scene, the CNN correctly predicts no water inside the city except for the Thames river and some lakes. Outside the city it avoids most False Positives in the fields. Thresholding overpredicts inside the city due to lots of low-backscatter noise pixels, but also does detect the Thames river inside the city. Outside the city it overpredicts in harvested fields, tree shadow lines, streets, building shadows, etc. For Freeport, the CNN predicts the flood water well except for some asphalted areas and some flooded streets between dense settle-

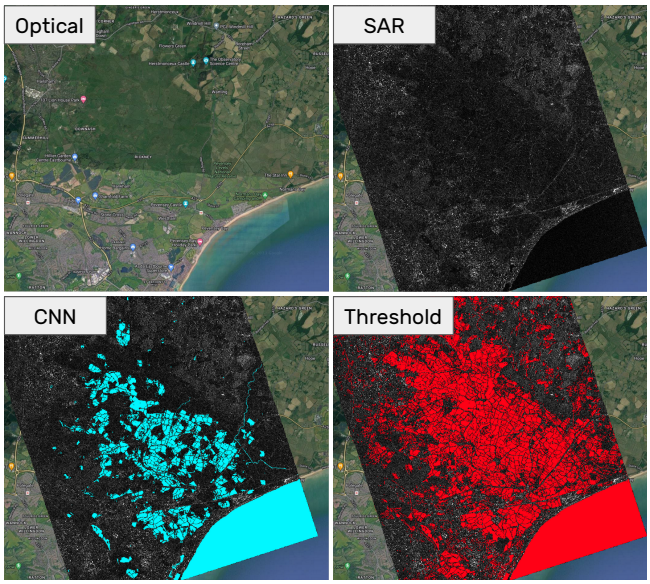

Fig. 4. Example of false positives due to low backscatter vegetation for CNN (cyan) and thresholding (red) on London, UK (crop). The optical reference is from ©Google Maps imagery.

ments. Thresholding detects the flooding, but overpredicts as it assigns all low backscatter regions to water which are typically streets and building shadows.

4. CONCLUSION

In this paper we have presented a deep learning approach to map urban floods using Capella Space HR SAR imagery. Our model can easily be used in a near-real time emergency activation scenario to provide high-resolution flood maps when optical observations are not available due to cloud cover. Several innovations were introduced to speed up the annotation process and to cope with the limited amount of training data. Our study shows that the CNN approach outperforms thresholding in all tested areas. Some limitations like false positives due to low-backscatter vegetation could be solved in the future by introducing similar scenes in the training data. False negatives in asphalted areas are due to the nature of SAR images where water cannot be distinguished from other smooth surfaces.

5. REFERENCES

[1] C. for Research on the Epidemiology of Disasters, "The human cost of weather related disasters 1995-2015," The United Nations Office for Disaster Risk Reduction, 2015.

[2] M. Ahmed, "Pakistan: World bank estimates floods caused $40b in damages." online article, Oct. 2022. [Online]. Available: https://apnews.com/article/floods-pakistan-south-asia-islamabad-25ee9dc0ec7aee6f4f2ef7b557216ee7

[3] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *ISCAS 2010*, May 2010, pp. 253–256.

[4] L. Landuyt, A. V. Wesemael, G. J. Schumann, R. Hostache, N. E. C. Verhoest, and F. M. B. V. Coillie, "Flood mapping based on synthetic aperture radar: An assessment of established approaches," *IEEE Trans. Geosci. Remote. Sens.*, vol. 57, no. 2, pp. 722–739, 2019.

[5] M. E. Tupas, F. Roth, B. Bauer-Marschallinger, and W. Wagner, "An intercomparison of sentinel-1 based change detection algorithms for flood mapping," *Remote. Sens.*, vol. 15, no. 5, p. 1200, 2023.

[6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, 2022.

[7] X. X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, pp. 143–172, 2021.

[8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[9] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *ICML 2019*, vol. 97, Jun. 2019, pp. 6105–6114.

[10] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, pp. 418–422, 2016.

[11] "Copernicus Digital Elevation Model (DEM), European Space Agency (ESA), accessed in 2021."

[12] B. Bauer-Marschallinger, S. Cao, C. Navacchi, V. Freeman, F. Reuß, D. Geudtner, B. Rommen, F. C. Vega, P. Snoeij, E. Attema, C. Reimer, and W. Wagner, "The normalised Sentinel-1 Global Backscatter Model, mapping Earth's land surface with C-band microwaves," *Scientific Data*, vol. 8, no. 1, Oct. 2021.