FLOOD AREA DETECTION USING SAR IMAGES WITH DEEP NEURAL NETWORK DURING, 2020 KYUSHU FLOOD JAPAN

Vaibhav Katiyar (1), Nopphawan Tamkuan (1), Masahiko Nagai (1) (2)

[1] Graduate School of Sciences and Technology for Innovation, Yamaguchi University, 2-16-1, Tokiwadai, Ube, Yamaguchi, 755-8611, Japan
[2] Center for Research and Application of Satellite Remote Sensing, Yamaguchi University, 2-16-1, Tokiwadai, Ube, Yamaguchi, 755-8611, Japan
Email: vkatiyar@yamaguchi-u.ac.jp; ntamkuan@yamaguchi-u.ac.jp; nagaim@yamaguchi-u.ac.jp

KEY WORDS: Flood Mapping, Semantic Segmentation, Deep Learning, Synthetic Aperture Radar (SAR)

ABSTRACT: Disasters are on the rise and with global warming, the frequency of floods has seen an exponential increase. Flood area detection becomes necessary to prioritize the focus area for reducing the response time and saving more lives. As flood inundates larger area (several hundreds of square-km), remote sensing becomes important to monitor this affected area. However, flood and bad weather are in general occurs simultaneously, this restricts us to use Optical images and thereby Synthetic Aperture Radar (SAR) images become the obvious choice. In the recent Kyushu flood in Japan, which occurred due to concentrated rain in the first week of July 2020, led to the flooding of Kuma River and its tributaries. SAR images of previous flood events in Japan were utilized as training data. Due to the limited training data and binary class (flooded vs non-flooded area) segmentation problem, this study chooses U-Net, as it has shown better results in the above-mentioned situation. In this study, U-Net with alternate 5x5 convolution to 3x3 convolution in the encoder part has been used to grasp more contextual information. The result has been compared with the GSI's (Geospatial Information Authority of Japan) flood inundation map as a reference and thresholding-based approach as a comparative result. Our proposed method shows a significant improvement compared to the thresholding method. Also, the detected flooded regions are more connected and homogenous in comparison to the patchy result of the thresholding method.

1. INTRODUCTION

Flood is one of the most devastating types of a natural disaster. As per the report, *'Economic Losses, Poverty & Disasters 1998-2017'* from the United Nations Office of Disaster Risk reduction (UNDRR) (Wallemacq, 2018), out of total affected people by all major natural disasters, flood alone has affected more than 45% of the people worldwide between 1998 and 2017. In general, flood affects the larger area and Remote Sensing fulfil the requirement of monitoring the huge swath at the same time. Specifically, Synthetic Aperture Radar (SAR) images are best suited during the flood event due to coincidence of flood and cloudy conditions. As SAR is unaffected by cloud cover, it has been used to map the flood-affected area during the recent Kyushu flood in Japan which has occurred due to excessive rain in the first week of July 2020. As archive data is not always available during disaster time, this study is using only 'during-flood' images to reduce response time.

During the disaster, remote sensing experts use automatic thresholding technique such as Otsu method to classify the flooded areas (Ohki, 2019; Tiwari, 2020). However, thresholding has a limitation of the presence of bimodality of the data. If the flooded region occupies a smaller portion in the scene then the segmentation accuracy of thresholding methods such as Otsu thresholding declines rapidly (Lee, 1990).

Deep learning techniques had been applied successfully with SAR images for various purposes such as for land cover classification (Sirirattanapol, 2020), water bodies extraction (Katiyar, 2019) etc. Semantic segmentation has been one of the major techniques to exploit in the automated analysis of Remote Sensing Images for finding an exact outline of the specific object/area. With the popularity of Deep Convolutional Neural Networks (DCNNs), many new network architectures for segmentation have been proposed. Some of them are Multi-scale or pyramid network-based models such as PSPNet (Zhao, 2017), Dilation convolution-based model such as DeepLab (Chen, 2017), Attention-based models such as DANet (Fu, 2019), Encoder-decoder based models such as HRNet (Wang, 2020), UNet (Ronneberger, 2015) etc. This study selected U-Net based architecture to segment the flooded region as this architecture is fulfilling the requirement of 'low complexity and high accuracy' (Bahl, 2019) as well as their proven history of better segmentation in single-band medical images (Ronneberger, 2015) and object extraction in SAR images (Katiyar, 2019).

2. STUDY AREA AND DATA USED

2.1 Focus Area

This study has selected the Kuma river basin which is one of the major rapid of the japan, as a focused area. One of the big cities of the Kumamoto prefecture, Hitoyoshi city has been impacted quite drastically due to levee breach in the month of July 2020. Kumamoto University's Center for Water cycle Marine environment and Disaster management (CWMD) in their disaster investigation report has mentioned that the flood level in the Hitoyoshi City even breached the previous record of July 1965 flood (CWMD, 2020). This led to the evacuation of thousands of people and scores of them have been died too. This is why Hitoyoshi city (Figure 1) has been chosen as the focus area, the city has been surrounded by mountains from all sides and in past also flash-flood has occurred in this region.

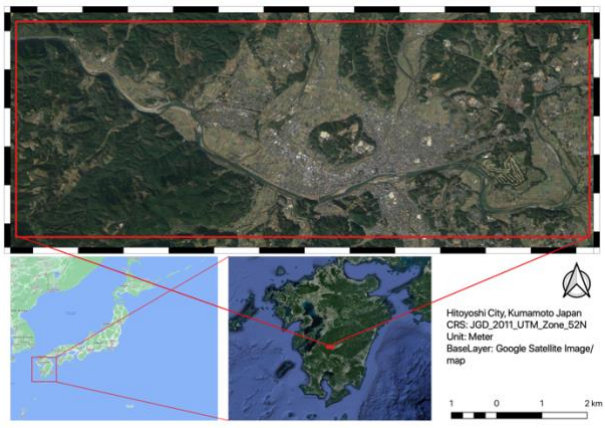


Figure 1: Study Area (Hitoyoshi city Kumamoto Japan)

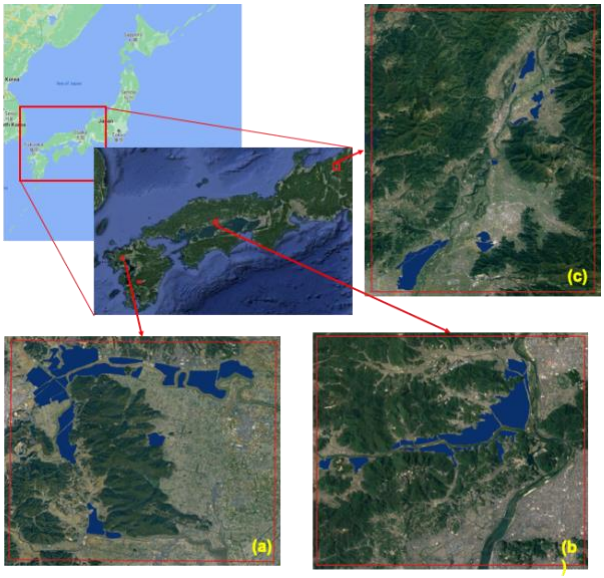


Figure 2: Areas used for training. (a) Saga city (b) Kurashiki city, Kumamoto (c) Nagano city. The blue colour in the area images is showing the flooded area

2.2 Data Used

This study has used ALOS-2 images during the flood. These images have been accessed through the Sentinel Asia portal, which is the international cooperation project with a focus on the disaster management in the Asia-pacific region and Yamaguchi University is part of it. All the scenes selected here are of HH polarization, while orbit looking is the mix of ascending and descending, solely based on the availability of image on the day of the flood. Other than the satellite images, the estimated flood inundation map generated by the Geospatial Information Authority of Japan (GSI) has been used to create the ground truth for flood-affected areas.

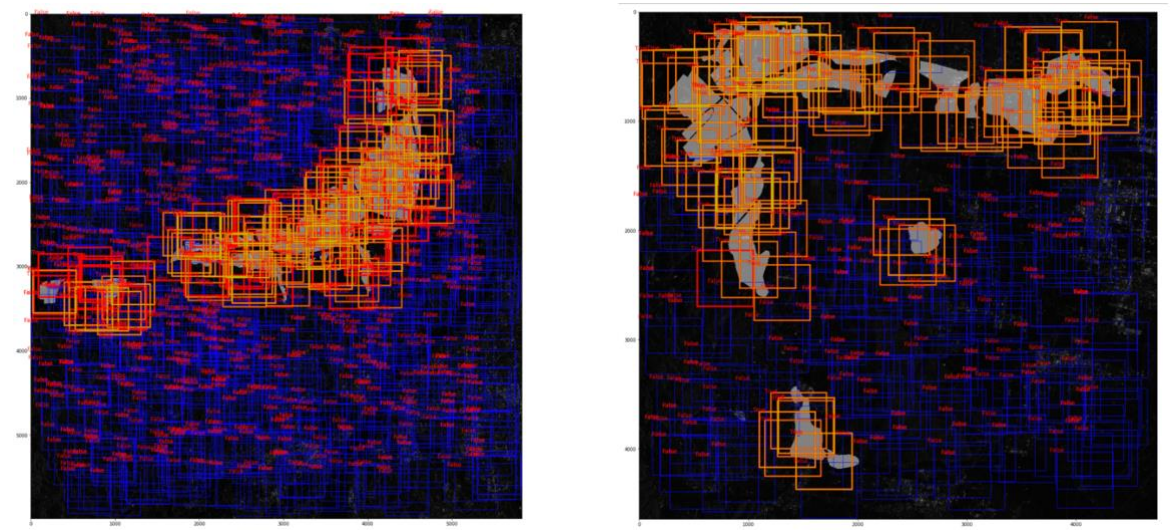
2.2.1 Training data: Availability of good training data is the basic need for the deep learning approaches to work better. This study has used scenes from three previous flood disasters: Kurashiki city area from Okayama prefecture during 2018 japan floods (7 July 2018), Saga city area from 2019 floods (28 Aug 2019) and Nagano city area during the Hagibis typhoon-fueled flood 2019 (13 Oct 2019) as shown in figure 2. With many different scenes which belongs to the different areas have an advantage of capturing more ground diversity in the training data rather than just using one scene or one area images. Ground truth has been created by using the estimated flood inundation map provided by GSI and by the help of SAR image expert. Still, there is the possibility of some areas being left out which was flooded and some of the non-flooded areas may be included as flooded. These problems during the creation of ground truth from SAR data happens due to the complex nature of SAR images. Some of the reasons for the problems are- 1) Geometry distortion due to the 'look angle' 2) Shadow areas in hilly regions 3) Paddy field which has been flooded for agriculture purposes and bare plain field that returns low to negligible backscatter and makes hard to distinguish between these areas with the flooded areas.

2.2.2 Testing data: For testing, Single polarized ALOS-2 image (HH) of 4th July 2020 has been used. The desired area for the testing site, Hitoyoshi city has been extracted in QGIS by the respective ROI.

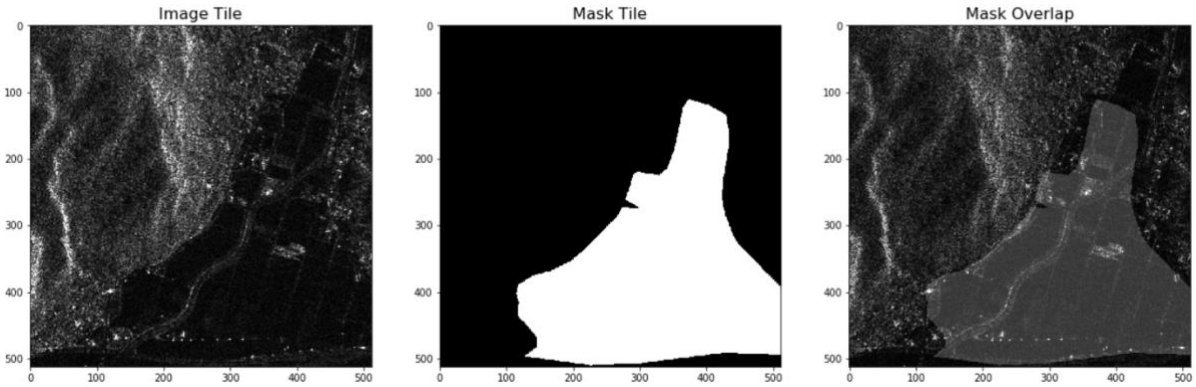
3. METHODOLOGY

3.1 Training Data Preparation

The ground truth of flood that was created as a polygon on QGIS software, has been saved as a mask raster file (GeoTIFF image) with the binary value of 1 (flooded area) and 0 (non-flooded area). Iterative Random sampling on this mask image has been carried out to create 512x512 pixel tiles and only those tiles which fulfils the condition that at least 10% of its pixels belong to the flooded area has been classified as valid tiles. Out of the total valid tiles, 200 of them has been selected randomly from each training area this makes a total of 600 tiles, together with the mask the corresponding tiles from images have also been saved.



(a) (b)



(c)

Figure 3: (a) and (b) are showing the training data creation using flood mask. Blue rectangles show not valid tiles(<10% flood area pixels), Red rectangle is valid and Yellow rectangles are the one chosen out of red ones. (c) is showing the image and mask tiles, created through the above-shown procedure

3.2 Network and Architecture

Due to the limited number of training tiles and binary classes (flood and non-flood class), we have selected U-Net with the modification of using 5x5 kernel size with 3x3 kernel size in the alternate setting during the encoder phase. This helps to increase the receptive field (RF) of the network. As flood is an event which affects the larger area with diverse landscape so larger RF will help in capturing local context details, that in turn will help in increasing the accuracy of the flooded-area segmentation.

For this study, we have used five encoder and five decoder blocks along with one bottleneck block. Loss has been calculated by binary cross-entropy, which is optimized by Adam optimizer with the initial learning rate 0.0001. The adaptive learning rate modified by 0.001 on each round after the training hit plateau for more than five epochs. Starting feature size is chosen as 16 which getting double at each encoder block till 512 features in the bottleneck block after which it starts getting half at each up-sampling. All the layers are using ReLU as an activation function except the last one which is using sigmoid as an activation function.

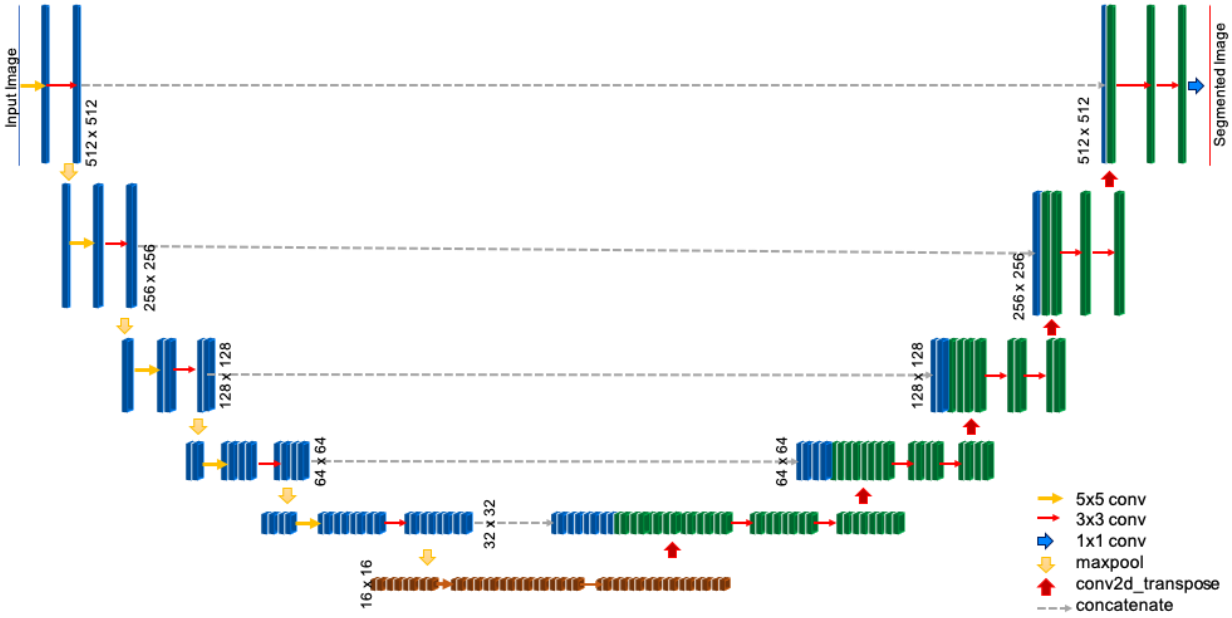


Figure 4: U-Net architecture used in the study

3.3 Training

As the total number of training tiles are less for training a deep neural network, this study has used image augmentation technique to increase the samples, mainly horizontal and vertical flip has been used which keep maintaining the data integrity. For validation data, the available data has been divided at runtime which has been set up as 20% of total tiles available which means practically 480 tiles has been used for training and 120 for the validation. However, these 480 tiles do not remain fixed but keep changing on each epoch due to validation split is happening at runtime. The network has been trained for 100 epochs with the batch size of 16 on two Nvidia Titan-V GPU. U-Net does not have many complex calculations and binary cross-entropy is much simpler loss function to optimize, this is why total training has taken less than one hour.

3.4 Testing:

The size of the test image is 5102x2319 (width x height). The ‘same’ padding i.e. zero value pixel has been added at the border of the tile during training to maintain the size of the output after convolution, but the output value at the border will surely be biased due to padding (zero or other value) (Huang B., 2018). This does create a big problem in the testing of remote sensing images as the training data and testing data always have a huge difference in their spatial extent. Due to this reason few techniques that have been proposed by Huang et al. (2018) have been used to minimize the effect of tiling on the result- 1) The output sized has been maximized as per the GPU memory, in our case (2048 x 2048) tile can be processed at the same time. 2) The translational variance has been handled by using overlapped tiles for evaluation and during stitching time just neglected the affected boundary pixels, that has been calculated by the number of times pooling has been done. Also, for other overlapped pixels we have taken weighted sum with the limit of max value as one because the output of the network is the probability of the pixels to belongs to a particular class and that has to be within the range of zero to one.

3.5 Evaluation metrics:

Following metrics has been used to measure the accuracy of the result as well as to compare with the other methods.

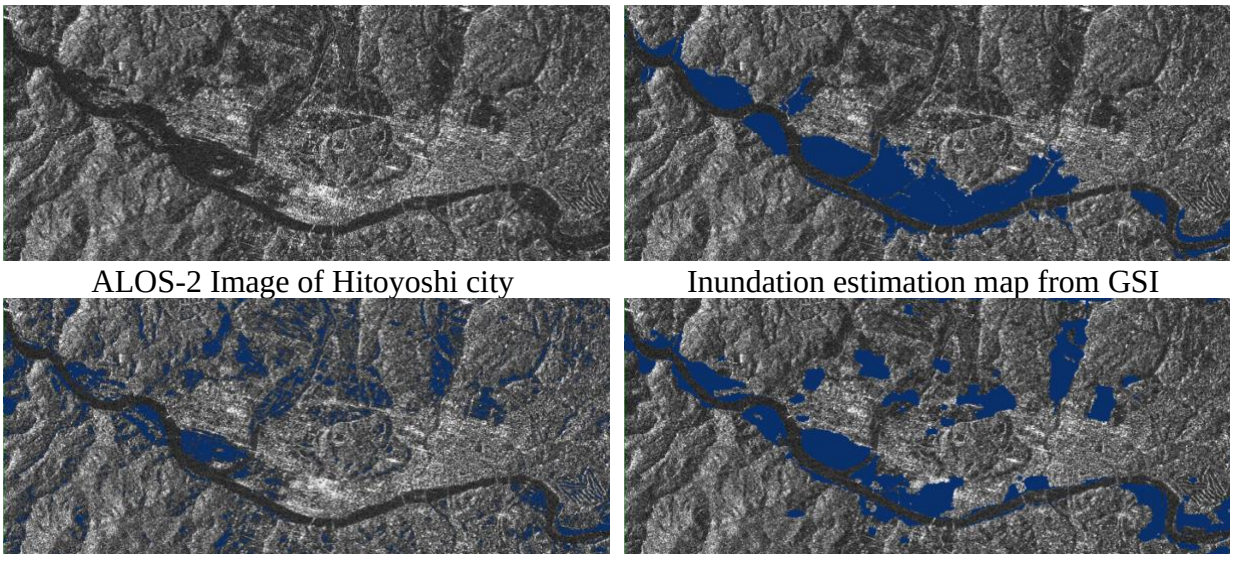
$$Accuracy = \frac{(True\ Positive + True\ Negative)}{Total\ Number\ of\ pixels}$$

$$\boldsymbol{F1\ Score = \frac{2 * Precision * Recall}{(Precision + Recall)}}$$

Where *precision* is the score for the proportion of positive identifications and *recall* is the proportion of correctly classified pixels in total positive pixels classified by the system.

4. RESULTS AND DISCUSSION

Our results show that U-Net is able to extract the flooded area in a more homogenous way in comparison to the thresholding method as can be seen in figure 5. The accuracy in comparison to inundation map provided by GSI Japan comes to 89.57%, improved from 84% by thresholding method, but, the major improvement can be seen in the F1 Score which has increases from 0.21 (thresholding method) to 0.43 (our method). Even though the F1 score improved by more than 100% from the thresholding method but still it is less, however, we need to recognize the lack of training data availability and also the inadequacy of the ground-truth credibility, as there are few areas which have been flooded but have not been included in the flooded-area due to limitation of ground data etc. and vice-versa as mentioned by GSI.



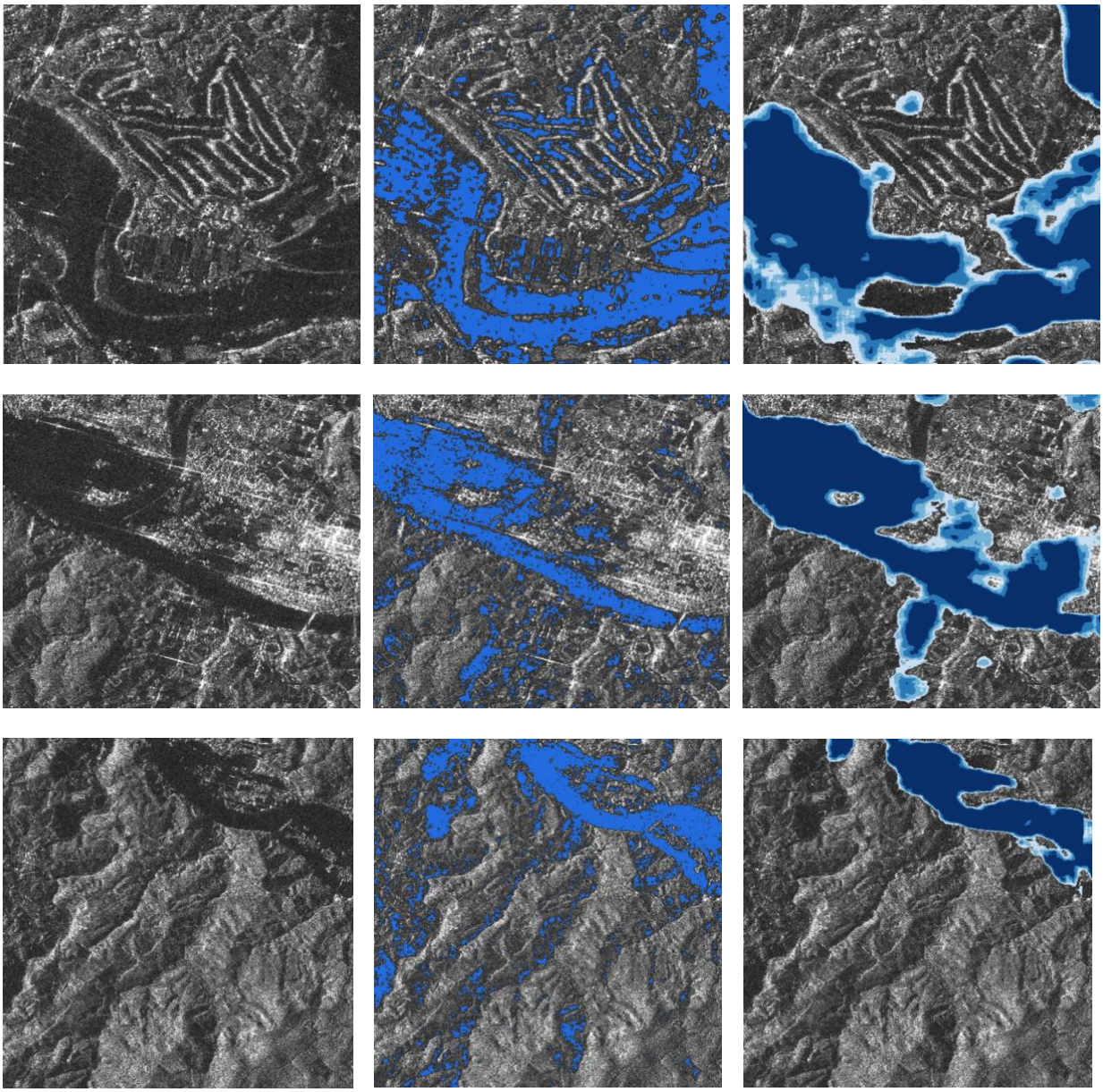
ALOS-2 Image of Hitoyoshi city | Inundation estimation map from GSI

Result by the thresholding method | Result by our method

Figure 5: Results, flooded-area in Blue colour

Figure-6 compares the result of our method with the thresholding method. The first row of figure-6 is showing the golf course which has been detected as water areas in thresholding methods due to the specular reflectance from the flat golf fields, however, our methods have been able to detect the difference. In the case of the middle row, thresholding was not able to detect urban areas which have been flooded but due to double-bounce, those areas were showing greater backscatter, here our model is able to detect some area as flooded area but still lot of flooded urban area has been left out. Another advantage of the method has also been seen in case of shadows in hilly regions (last row of figure 6) where our method precisely left out the smaller shadow regions. As Japan

has a lot of hilly regions that is why in general, pre-processing is required using Digital elevation model (DEM) to remove those shadow regions based on the slope of the area, nonetheless here we have not used any pre-processing over images. Although some very big shadow region may have been classified as a flooded region by our method too. Here we also want to emphasize that the data that has been used for training belongs to complete different areas with no particular geometry as we have used ascending and descending images without any priority and when applied on this test regions it has given a very good result, this shows that network was able to grasp the complexity of SAR images.



SAR Image **Thresholding** **Our method**

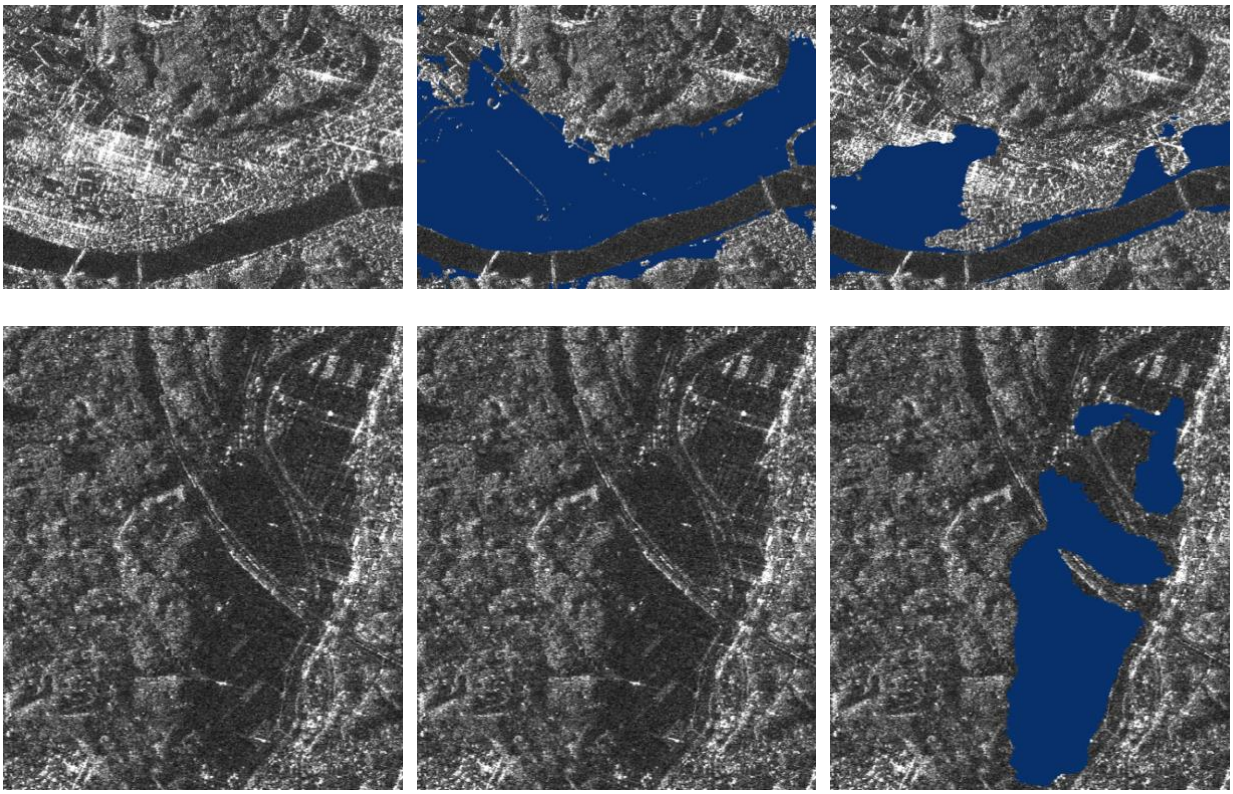
Figure 6: Part of results to highlight the advantages of thresholding methods. The shades of blue colour in our method are representing the confidence of network for classifying the area belongs to flood (Dark blue means the confidence is in range of (0.9-1.0) and lightest blue is in range (0.4-0.6) and the remaining one shades lie in between (0.6-0.9))

4.1 Limitation and future perspective:

Although the U-Net has performed really well in comparison to the thresholding method this study

has encountered few limitations too. Some of them can be seen in figure 7, in the first, the areas as per the ground truth available are the urban area which has been flooded and the last row is showing the areas that contained river with surrounding agriculture field. In the first row, our result has been able to detect some areas as flooded areas which may become possible due to the presence of some urban flooded areas in our training data too. This is a very challenging task as an urban flood can become very bright due to double bounce and it may also become bright in certain look angle so differentiating between both of them will be a big challenge. In the last row, our model is showing a clear presence of flood however GSI's map does not mention anything about this area. One possible reason can be the presence of flooded paddy fields which has been detected as a flooded region. Another reason can be that it was flooded and GSI may miss it, as one news article mentions that the Yamada River (the river presents in the last row area) has been flooded but did not mention the exact stretch of the flooded river so cross-verification was not possible.

For future study, it will be good to include some other data except SAR images due to the limitation of SAR images such as distorted geometry, shadow regions and uncertainty of urban flood areas visualization. DEM information can be combined to take care of the geometry and shadow regions. Multi-temporal SAR images and different polarization may also help in separating the flooded paddy fields with other flooded areas. One very important method enhancement is possible by using the pre-flood image along with the during-flood image for applying a change detection mechanism.



SAR Image | Ground truth of flood by GSI | Our method output

Figure 7: Limitation of the method

5. Conclusion

Disaster management needs quick results as response time plays important role in saving people lives by timely evacuation. For this purpose, our method using U-Net have done a good job as it

has been able to process the complete scene of the satellite within a few seconds while maintaining the good detection ability. The detected flooded regions are more connected and homogenous in comparison to the patchy result of the thresholding method. Overall the accuracy of the method has been close to 90% which can be considered as a good detection rate. There has been some wrong detection as well which can be seen by the F1 measure, however, there are few reasonable reasons for that such as limitations of the SAR images and lack of credible training data. Additional data along with the more credible ground-truth will surely improve the flood area extraction.

References and/or Selected Bibliography

References from Journals:

Bahl G., Daniel L., Moretti M. and F. Lafarge, 2019. "Low-Power Neural Networks for Semantic Segmentation of Satellite Images," IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 2469-2476, DOI: 10.1109/ICCVW.2019.00302.

Chen, L., Papandreou, G., Schroff, F., & Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. ArXiv, abs/1706.05587.

Fu, J., Liu, J., Tian, H., Fang, Z., & Lu, H., 2019. Dual Attention Network for Scene Segmentation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3141-3149.

Huang, B., Reichman, D., Collins, L. M., Bradbury, K., and Malof, J. M. 2018. Tiling and Stitching Segmentation Output for Remote Sensing: Basic Challenges and Recommendations. arXiv:1805.12219. Available online at http://arxiv.org/ abs/1805.12219

Katiyar, V., & Nagai, M. 2020. Automated extraction of water bodies from ALOS-2 images using U-net and rough training set. Paper presented at the 40th Asian Conference on Remote Sensing, ACRS 2019.

Lee, S., Chung, S.Y., & Park, R. 1990. A comparative performance study of several global thresholding techniques for segmentation. Comput. Vis. Graph. Image Process., 52, 171-190.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., & Terzopoulos, D., 2020. Image Segmentation Using Deep Learning: A Survey. *ArXiv, abs/2001.05566*.

Ohki M., Watanabe M., Natsuaki R. et al. 2019., "Flood Area Detection Using PALSAR-2 Amplitude and Coherence Data: The Case of the 2015 Heavy Rainfall in Japan", in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2288-2298, DOI: 10.1109/JSTARS.2019.2911596.

Ronneberger O., Fischer P., and Brox T., 2015. "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2015, pp. 234–241

Sirirattanapol, C., Tamkuan, N., Nagai, M., & Ito, M. 2020. Apply Deep Learning Techniques on Classification of Single-band SAR Satellite Images. In Geoinformatics for Sustainable Development in Asian Cities (pp. 1–11). https://doi.org/https://doi.org/10.1007/978-3-030-33900-5_1

Tiwari, Varun et al. "Flood inundation mapping- Kerala 2018; Harnessing the power of SAR,

automatic threshold detection method and Google Earth Engine." *PloS one* vol. 15,8 e0237324. 19 Aug. 2020, DOI:10.1371/journal.pone.0237324

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B., 2020. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE transactions on pattern analysis and machine intelligence*.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J., 2017. Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6230-6239.

References from Other Literature:

Wallemacq P., Below R., McLean D., 2018. UNISDR and CRED report: Economic Losses, Poverty & Disasters (1998 - 2017), pp. 06-10.

References from websites:

ALOS-2/ ALOS User Interface Gateway. Accessed August 2020. https://auig2.jaxa.jp/openam/UI/Login?

CWMD Kumamoto University,2020. 2020 年 7 月豪雨に伴う熊本県南部における災害調査速報 Retrieved Oct 11 2020, https://cwmd.kumamoto-u.ac.jp/wp/wp-content/uploads/2020/07/report_20200708.pdf, pp. 5-8.

Wallemacq P., Below R., McLean D., 2018. UNISDR and CRED report: Economic Losses, Poverty & Disasters (1998 - 2017), Retrieved Oct 10 2020, from https://www.cred.be/unisdr-and-cred-report-economic-losses-poverty-disasters-1998-2017