Near-real-time Flood Mapping using Variants of Autoencoders with Synthetic Aperture Radar Images and other Ancillary Data

(合成開口レーダ画像と補助データを用いたオートエンコーダによる準リアルタイム洪水マッピング)

A dissertation submitted in partial fulfilment of the requirements for the degree of
Doctor of Engineering

by

Vaibhav Katiyar

September 2021

Graduate School of Sciences and Technology for Innovation,
Yamaguchi University

Vaibhav Katiyar

vkatiyar@yamaguchi-u.ac.jp

ORCID iD: 0000-0003-3453-3158

PREFACE

This dissertation is hereby submitted to the Graduate School of Sciences and Technology for Innovations (大学院創成科学研究科）of Yamaguchi University, Japan in partial fulfilment of the requirements for the degree of Doctor of Engineering. The dissertation is the author's original contribution and relevant references are provided in the dissertation itself, wherever the author has used the results or data obtained from previous studies.

The publications that have been published in the course of the research and which have contributed significantly in the chapters of this dissertation are as follows:

Papers

- *Katiyar, V., Tamkuan, N., Nagai, M (2021). Near-Real-Time Flood Mapping Using Off-the-Shelf Models with SAR Imagery and Deep Learning. Remote Sens. 2021, 13, 2334.*
  https://doi.org/10.3390/rs13122334
  This journal article focuses on the near real-time flood mapping using Sentinel-1, including the application of the trained model as an off-the-shelf model on the selected test site in Kerala India. This article is contributed significantly to chapter 3 of this dissertation.

Conference Papers

- *Katiyar, V., Nagai, M. (2020). Katiyar, V., Training data development strategy for applying deep learning in remote sensing applications. 1st Intercontinental Geoinformation Days (IGD).*
  This paper discusses the strategies of the training chips creation from big satellite images. Similar techniques are employed to create training data in this study wherever the customized dataset is used. Chapter 4 of this study mainly comes from this paper.

- *Tamkuan, N., & Nagai, M. (2020). Flood Area Detection Using Sar Images with Deep Neural Network During, 2020 Kyushu Flood Japan. In The 41st Asian Conference on Remote Sensing (ACRS 2020).*
  The content of this paper contributed mainly to chapter 5 of this dissertation, which mainly focuses on the near-real-time flood mapping using ALOS-2 data.

- *Katiyar, V., & Nagai, M. (2019). Automated extraction of water bodies from ALOS-2 images using U-net and rough training set. In 40th Asian Conference on Remote Sensing, (ACRS 2019).*
  This paper presented the work to detect water bodies from ALOS-2 using U-Net. It has contributed to a part of chapter 6.

ACKNOWLEDGEMENTS

First of all, I am grateful to the almighty God for the good health and wellbeing that were necessary to complete this doctoral dissertation with the title **Near real-time flood mapping using variants of autoencoders with Synthetic Aperture Radar (SAR) images and other ancillary data**. It is with great pleasure that I present my sincere thanks to all those who helped me to complete this work.

I wish to express my deep gratitude first and foremost to Prof. Masahiko Nagai, my academic supervisor, for his guidance, motivation, and useful suggestions to finish my doctoral program through hundreds of discussions and research meetings. His continuous support and unwavering patience throughout the study helped me to deal with all the ups and downs easily.

Besides my advisor, I would like to thank the rest of my committee members Prof. Hideaki Nakamura, Prof. Takahiro Osawa, Dr Keiji Imaoka and Dr Ariyo Kanno for their insightful comments, hard questions and encouragement.

I would also like to thank all the parties who have provided the data used in this research: JAXA and ESA for the SAR data, JAXA and NASA for the DEM data, JSI for validation data, Sentinel-Asia platform for various kinds of satellite data for better training-data creation.

My special thanks to Dr Nopphawan Tamkuan and Dr Ratnesh Kumar for active discussion on the various aspects of the study and their continuous assistance, feedback, and uplifting comments. I would like to include my friends Dr Seemanta Sharma Bhagabati and Ritukar Vashistha too who were even though staying far but keep calling me and encouraging me to do my best. I also take this opportunity to express my gratitude to Asahi Harimoto for his continuous support since my arrival in Ube until he graduated, also Hana Watari and other Japanese lab mates for their help in translating Japanese and organizing many lab parties. It has been an enjoyable and fun-filled journey during my doctoral program and the credit goes to my friends including Ann, Sam, Husniyah, Amanda, Dr. lally, Bien, Oka, Arpit, Pratibha, Mariko, Marin, Satomi, and many other international students at Yamaguchi University.

Finally, I wish to thank my beloved Mother Rani Katiyar, Father Ravindra Katiyar and Sister Dr Neelam Katiyar for their endless care, support, love, and prayers throughout my life. Also, thanks to all my near and dear ones who have blessed me always and keep me in their prayers.

TABLE OF CONTENTS

CHAPTER 5 NEAR-REAL-TIME FLOOD MAPPING USING ALOS-2 IMAGERIES, WITH A CASE STUDY ON KYUSHU FLOOD 2020 41

CHAPTER 6 EXTRACTION OF WATER BODIES FROM ALOS-2 IMAGES BY UTILIZING ROUGH TRAINING SET 51

CHAPTER 7 DISSERTATION FINDINGS, CONCLUSIONS, AND FUTURE RECOMMENDATIONS 58

LIST OF FIGURES

LIST OF TABLES

Chapter 1 **Introduction**

Background and scientific motivation

A flood is one of the most devastating types of natural disaster. As per the report, ‘Economic Losses, Poverty & Disasters 1998-2017’ from the United Nations Office of Disaster Risk reduction (UNDRR), 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 (Figure 1.1). The importance of surface water mapping can be further understood by studying UN Sustainable Development Goals (SDGs), in which as many as four goals directly mention surface water monitoring, including food security (target 2.4), water-related ecosystem management (targets 6.5 and 6.6), and the effect on land (target 15.3). However, the most relevant target concerning this study is target 11.5 under goal 11 (sustainable cities and communities). It states “By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the economic losses relative to the gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations” (UNEP, 2015). In this context, near-real-time (NRT) flood mapping becomes very necessary.

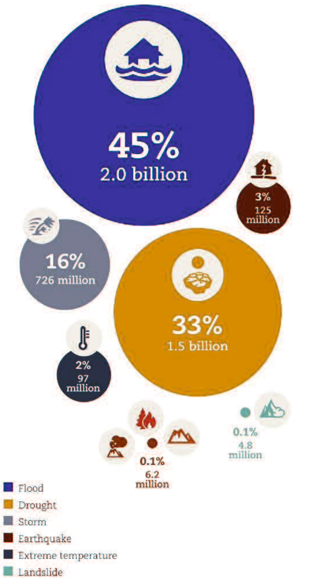


Figure 1.1 Number of people affected per disaster type during 19998-2017 (CRED report ‘Economic Losses, Poverty & Disasters 1998-2017’)

There are mainly two types of causes of floods, natural or manmade. Natural reasons are localized extreme rainfall, snowmelt, dike or dam break, glacial or lake outbursts, storm surges from a tropical cyclone and tsunami (Cloke et al., 2017), whereas, manmade reasons include deforestation, urbanization, elimination of floodplains and wetlands as well as river regularizations (Kundzewicz, 2008). Flooding is a more widespread phenomenon affecting thousands of kilometres and some of the majorly affected areas are lying in the developing countries of Asian and African continents which are lacking on-site hydrological monitoring and other physical systems. Moreover, the transboundary water further makes it a complex exercise to monitor the flood extent. Due to these reasons remote sensing is an important tool and with the improved spatial, temporal, and radiometric resolution of satellite images, it becomes the obvious choice for flood mapping (H. Yang et al., 2011). There are many works related to the extraction of surface water information, including floods, varying from different sensor types to different methods (C. Huang et al., 2018). Huang et al. (2018) mentioned that the number of related works with “surface water” or “flood inundation” and “remote sensing” has seen three- to seven-fold growth since the year 2000 in comparison to the previous decade. A similar trend was also observed in the case of the mapping surface water from SAR imagery (Schumann & Moller, 2015). Much of the surface water and flood extent extraction studies using optical images have mainly focused on using water indices such as the Modified Normalized Difference Water Index (MNDWI) (Xu, 2006), the Automated Water Extraction Index (Feyisa et al., 2014), and other rich spectral information (X. Yang et al., 2018). Though studies by Yang et al. with sentinel-2, Herndon et al. with Landsat-8 (Herndon et al., 2020) and Feng et al. with the GaoFen-2 and WorldView-2 images (Feng et al., 2019) have provided good results, optical images carry an inherent limitation; namely, dependence on solar radiation. This limits the optical satellites image acquisition time only in the day, inability to penetrate cloud cover and get affected by the adverse weather conditions which are often prevalent during high-impact flood events (Schumann et al., 2018). Some studies have used multimodal information, such as combining optical images and LiDAR with synthetic aperture radar (SAR) images (Assad, 2019) and combining other auxiliary information with SAR (Bioresita et al., 2019). Although multi-modal-based models provide improved results due to the complementary capability but using a variety of data make it difficult to provide an NRT flood map because of more time spent in data acquisition, receiving and processing. Therefore, only SAR images are being used in this study. Flood mapping with only SAR images is further categorized into two types: change detection using images captured before and during the flood, and water area detection using a single SAR image captured during the flood. Though change detection methods have the advantage of handling overprediction by providing additional information about water-like surfaces, an additional limitation that occurs with this method is finding an adequate reference image for flood detection (Hostache et al., 2012). During the disaster response phase, the time is “key” and in this situation, it is necessary to have as few constraints as possible to speed up processing, which is why this study adopted flood mapping based on a single SAR image. For flood mapping from a single SAR image, the most well-known technique is thresholding with certain algorithms, such as Otsu thresholding (Nobuyuki Otsu, 1979) and minimum error thresholding (Kittler & Illingworth, 1986), with a global or tile-based approach.

Because global thresholding has a constraint of bimodality of data, which creates a problem in the case of full image processing, tile-based thresholding, and the split-based approach (Martinis et al., 2009) were proposed. However, with these methods, the choice of tiles plays an important role and the wrong tile may impact the threshold significantly (Landuyt et al., 2019).

In recent times, we have seen lots of applications of deep learning in the remote sensing domain. DCNNs have achieved significantly higher accuracy in comparison to other image processing methods especially in cases like road segmentation (T. Li et al., 2019), building detection (W. Li et al., 2019), land cover classification (C. Zhang et al., 2019) etc. However, here we need to emphasize that these higher accuracies were achieved due to well established public datasets, provided through SpaceNet challenges (Etten et al., 2018), ISPRS labelling contest (Demir et al., 2018; Rottensteiner et al., 2014), DeepGlobe challenge (Demir et al., 2018) etc. These kinds of public datasets are not available in many other areas such as in more dynamic cases of natural disasters. Moreover, most of the datasets are available for optical high-resolution images. Synthetic Aperture Radar (SAR) datasets are very scarce, which created the need to develop our customized datasets. This study has used ALOS-2 level-1.5 image scenes of HH polarization to create our customized dataset. All the ALOS-2 scenes belong to the flood events of japan of different areas and times. Along with the above, the recently released dataset Sen1Floods11 (Bonafilia et al., 2020) which contains image chips based on Sentinel-1 was used in this study. This also helped us to explore transfer learning (Goodfellow et al., 2016) in detail between the Sentinel-1 and ALOS-2 SAR satellite data. To my knowledge, this area of applying transfer learning in between different SAR images has been completely untouched till now. However, this is a very important domain as SAR satellites are becoming ubiquitous in earth observation and along with government agencies many private space companies are also launching their own SAR satellite constellations.

Deep learning techniques had been applied successfully with SAR images for various purposes such as for land cover classification (Sirirattanapol et al., 2020; Q. Zhang et al., 2019), road extraction (Q. Zhang et al., 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, Dilation convolution-based model such as DeepLab (L. C. Chen et al., 2018), Attention-based models such as DANet (Fu et al., 2019), Encoder-decoder (autoencoder) based models such as HRNet (J. Wang et al., 2020), Unet (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017)etc. This study selected UNet and SegNet like architecture to segment the flooded region as these architectures are fulfilling the criteria of "low complexity and high accuracy" (Bahl et al., 2019) as well as their proven history of better segmentation in single-band medical images (Ronneberger et al. 2015) which can be seen on par the SAR images.

As satellite scenes are too big that is why we need to create image chips out of them (Han et al., 2017), which can be feed to DCNNs that can run efficiently on the GPUs memory. As per Ning et al. (2020), training a network with a higher number of image chips normally leads to greater accuracy (Ning et al., 2020). However, labelling SAR images are not only time consuming but also need an expert understanding of the SAR images, this makes it a very expensive task financially and time-wise (Bonafilia et al. 2020). Though data augmentation has proved a successful mechanism to increase the variability from limited data and in turn improves the performance of the Convolutional Neural Networks (Ding et al., 2016). There is still a need to find other ways to increase the availability of training data without significantly increasing the cost.

Formulation of Objectives

Based on the previously stated background and scientific motivation, the major objective for this study can be outlined as:

1. How and what kind of <u>data</u> to feed the DCNNs?
   Under this objective, the major focus is on the data preparation part. Starting from selecting the labelling techniques for annotation, to the methods of creation of chips from satellite images. Moreover, finding suitable band polarization combinations for the flood mapping or improving the accuracy of models.
2. How to efficiently <u>train the model</u> for a given input?
   This objective mainly focuses on the training part. From selecting the best model architecture to the different training scenarios. Lastly, using transfer learning making the training quicker while needing a smaller number of training samples.
3. How to <u>apply the model</u> to various satellite scenes?
   The last objective is to apply the trained model to real-life satellite images. Under this will be focusing on the methods to efficiently apply the trained models on much bigger

satellite images while managing the boundary pixels error. Also, will be testing for the generalization capability of the model by using transfer learning in different satellite images. This will be especially helpful in the case of the SAR satellites which are recently launched and training data for those satellites are not available to train the models from scratch.

Structure of the dissertation

This dissertation consists of seven chapters. Chapter 1 presents the background and scientific motivation, detailed objectives, and the structure of the dissertation. A short explanation of the basic theory behind SAR and unique properties of the water areas in SAR data and the basics of deep learning has been discussed in chapter 2. Chapter 3 proposes detailed methods for near-real-time flood mapping using Sentinel-1 training data and a model has been proposed to be used as an off-the self-model, the application of which has been done at the Kerala flood 2018. The detailed discussion of training data preparation and the best method for chip creation was discussed in chapter 4. In chapter 5, the near-real-time flood mapping with ALOS-2 images is discussed with the focus on the recently happened Kyushu flood in 2020. Whereas, chapter 6 discusses the utilization of similar methods for permanent water bodies extraction. Finally, the findings and the conclusions of this dissertation, along with ideas on the future role of deep learning for flood mapping, are presented in chapter 7.

Chapter 2 **SAR imaging and inundated regions response against sensor-related factors**

Concept of SAR Imaging

Imaging SAR generally works with varying frequencies, some of the most used are X-, C-, L- and P-bands. Synthetic Aperture Radar (SAR) was developed in response to the limitation of Real Aperture Radar (RAR) with the size of the antenna for achieving higher resolution. In the case of SAR, the aperture is synthesized through modified data recording and processing techniques (Figure 2.1). SAR is a side-looking sensor that looks towards the ground at the perpendicular of its motion (Figure 2.2). Due to this reason, the effective antenna length increases from near-range to far-range, resulting in the constant azimuth resolution irrespective of range.

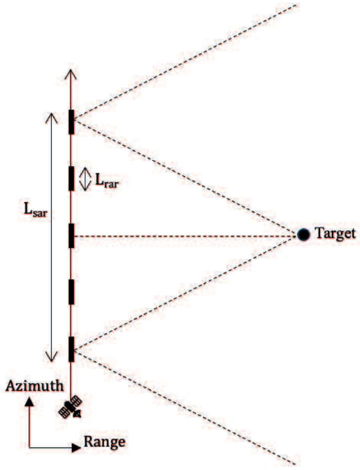


*Figure 2.1 Concept of synthetic aperture (*K.-S. Chen, 2016*, modified)*

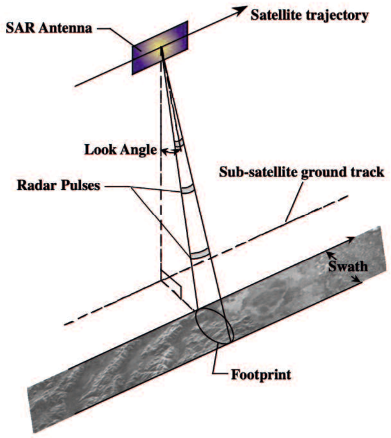


Figure 2.2 Schematic diagram of SAR satellite in orbit (Sandwell et al. 2019)

SAR signal response over different flooded regions

SAR signals are scattered against different kinds of surfaces and objects differently. In the case of inundated regions, this varies from the specular reflections to the diffused scattering. Some of the specific surfaces and the corresponding scattering responses are discussed in the following sections.

Over Ocean Surface

Backscattering over the ocean is mainly dependent upon the wind influenced waves. Therefore, wind speed and wind direction play a major role in deciding the range of backscattering from the sea. In the special case of the capillary waves and smaller gravity waves, making a regular periodic structure and whose distances are in the range of the wavelength of the SAR sensor, will

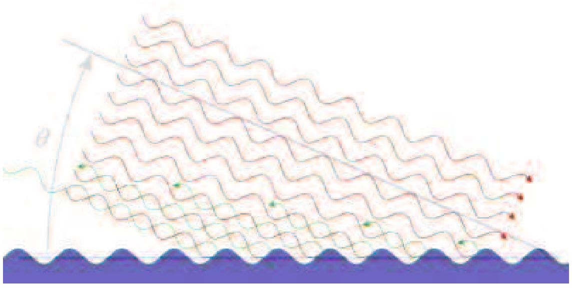


Figure 2.3 Bragg scattering and constructive interference (Wolff, n.d.)

lead to Bragg scattering (Figure 2.3). Here constructive interference may lead to significant echo signals, that will represent by the bright pixels in the image.

In the case of a shallow smooth surface which is mostly homogenous will results in specular reflections and this leads to no or negligible return of the backscattering towards the sensor. This result in the dark pixels in the SAR image. In some cases, diffused scattering is also to be seen at the wave breaks or because of rain-induced ripples etc.

Over the inland water bodies

The presence of calm wind over large water bodies or smaller water bodies that do not get affected much by wind are mainly shown the surface backscattering. Or in other ways, inland water bodies can be considered as perfectly smooth surfaces with high dielectric constant, which acts as a specular reflector. Due to very low signal returns these water bodies, in general, appear dark in the SAR image. This land-water contrast rises with a higher incidence angle. However, it will also result in larger shadow in the mountainous regions or urban areas in the case of high-resolution SAR sensors. Moreover, a higher incidence angle may limit the vegetation penetration capability and hinders the mapping of the inland water bodies such as rivers due to the presence of trees at its banks.

Over the flooded vegetation

Some of the major kind of scattering happens in case of flooded vegetation area are scattering happening over the canopy level, from canopy to trunk level, tree stems to surface level, scattering from the rugged floor as well as from inundated water surface (Figure 2.4).

Richards et al. (1987) and Ormsby et al. (1985) discussed how SAR sensor with different wavelength affects the type of objects the radar is sensitive to. The magnitude at which the backscattering range increases due to flooding under vegetation is analyzed with the help of L-and C- band satellite data (Ormsby et al., 1985; RICHARDS et al., 1987). Concerning vegetation, Ormsby et al. observed that the L-band SAR data help to separate the deciduous or coniferous vegetation from the shorter, partially submerged grasses and shrubs. The X- and C-band radar data are useful when the vegetation of the disaster region is relatively short.

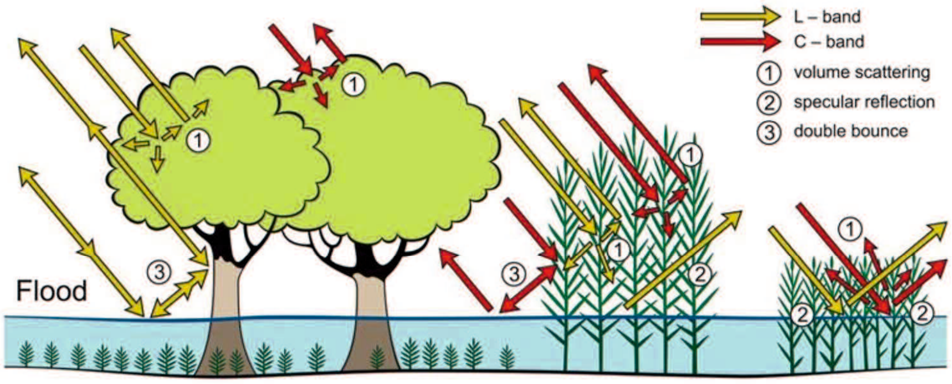


Figure 2.4 C and L-band SAR interactions with flooded vegetations (Evans et al. 2009, modified)

In general, over such flooded forest regions, high classification accuracy is obtained when using the high wavelength SAR signals during the leaf-off conditions. In the case of using the C-band SAR data, Lang et al.(2008) and Martinis et al. (2009) prove that C-band SAR data still act as a valuable tool in mapping the flood beneath the forest regions where the canopy is sparse and short (Lang et al., 2008; Twele et al., 2016).

Over the flooded urban regions

Assessment of flooded urban regions using a SAR image is complicated due to multiple backscattering effects that are mainly dominated by the corner reflection (dihedral and trihedral scatterer) of urban buildings and other concrete infrastructures (Mason et al., 2010). As seen in Figure 2.5, the flooded street AD will be affected by layover (AB) and shadow (CD) due to the buildings, resulting in a very small part of the flooded street i.e., BC is visible. In this case, only

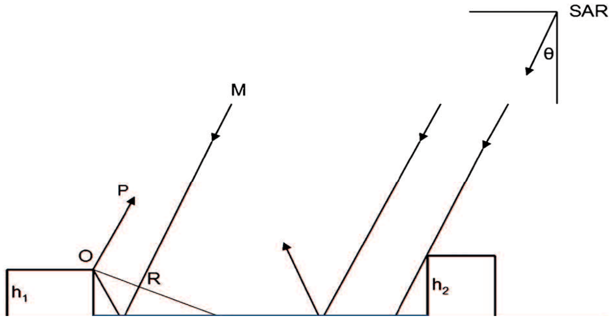


Figure 2.5 Layover (AB) and shadow (CD) regions in a flooded street (AD) between adjacent buildings of height h1 and h2. θ is the incidence angle (Mason et al. 2010, modified)

using “during flood” images, may not be sufficient which leads to many proposed hybrid methods such as Using LiDAR DSM data, SAR interferometry, change detection etc.

Effect of Radar polarization for mapping of the flooded regions

In SAR image-based flood area mapping application, selecting the right choice of the polarization image has to be given maximum importance as this plays a significant role in classifying the inundated regions with maximum possible accuracy. Hess et al. (1990) analyzed temporal SIR-C data of the Amazon river flood region and experienced both C-band and L-band co- and cross-polarization data and observed that HH polarization was most useful for distinguishing flooded from non-flooded vegetation and cross-polarized L-band data provided the best separation between woody and nonwoody vegetation (HESS et al., 1990).

Henry et al. (2006) show the basic statistics of HH, HV, and VV polarization images and its analysis exhibits that HH histogram is much wider and exhibits higher radiometric dynamics than the other two polarization images (Figure 2.6). The wide histogram of the HH polarization image helps for the better discrimination of thematic classes including possible differentiation of identifying the open water from recently inundated areas, which mostly requires high-resolution SAR images (Henry et al., 2006). Both Y. Wang et al., (1995) and P. Townsend, (2002) observed that the ratio of backscatter from a flooded forest to that from a non-flooded forest is higher at HH polarization than at VV polarization.

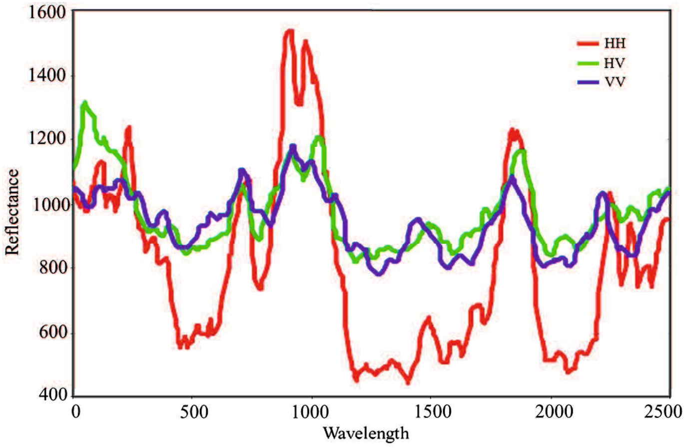


Figure 2.6 Radiometric profile of HH, HV, and VV polarization data of ENVISAT ASAR of Elbe River flooding (Henry et al. 2006)

Effect of incidence angle for mapping the different flooded regions

In flood area mapping applications, the incidence angle of the sensor at which the terrain information is captured plays a significant role as SAR signal response from steep to narrow incidence angle has a considerable impact on the radiometric and backscattering range of the SAR image. Studies by Richards et al. (1987), Hess et al. (1990), Wang et al. (1995) and Bourgeau-Chavez et al. (2001) indicate that the steeper small incidence angles are preferable to distinguish the flooded forest from non-flooded forest regions (Bourgeau-Chavez et al., 2001; HESS et al., 1990). The signals of steep incidence angles have more penetration capabilities through the canopy cover due to its shortest travel path, which in turn supports increased transmissivity at the crown level as well as permits more energy interaction at the ground-trunk level. In contrast, shallow incidence angle signals interact more at the canopy level, which in turn increase the volume scattering.

Chapter 3 **Near-Real-Time flood mapping using the off-the-shelf models by using a public dataset with Sentinel-1**

Study area and data used

Study area

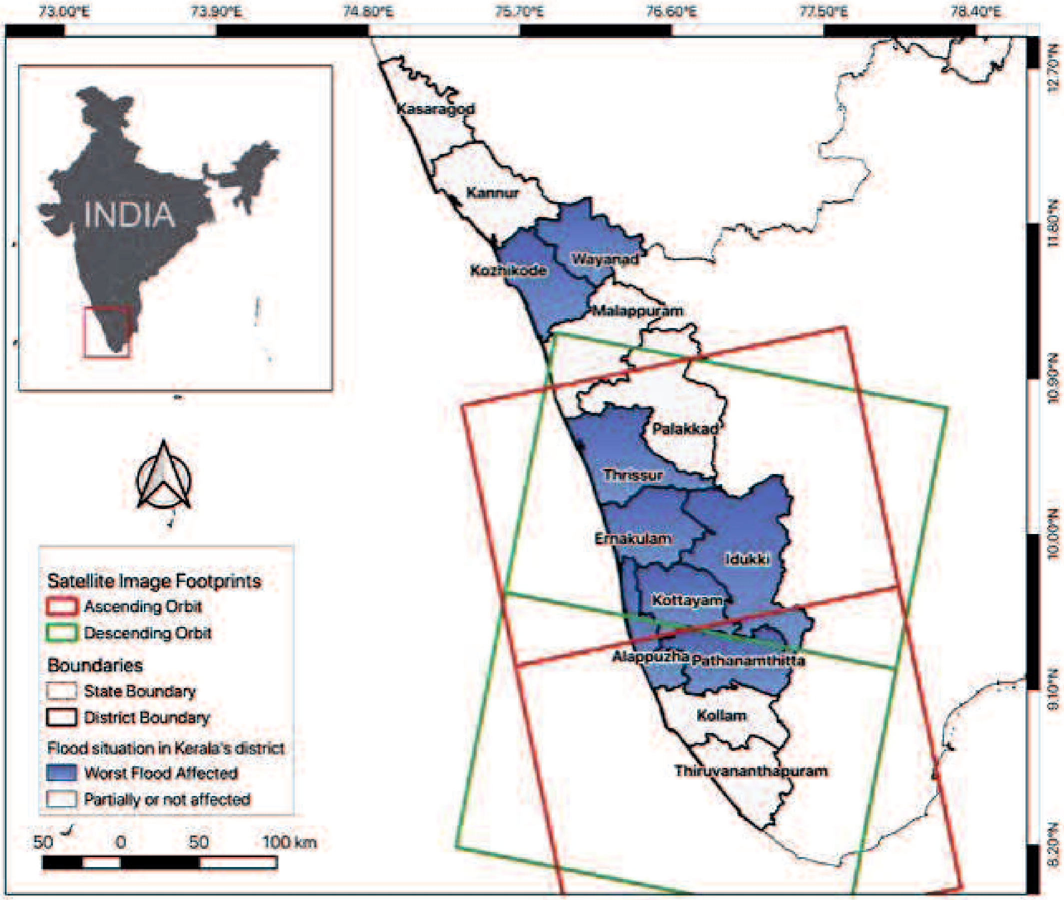


Figure 3.1 Kerala district map showing the areas worst affected by the 2018 flood.

The study area that was chosen for applying the off-the-self model is in the southern state of India named Kerala, as shown in Figure 3.1. In 2018, an especially devastating flood occurred in Kerala; this flood took more than 400 lives and affected millions more.

Table 3.1 Characteristics of the Sentinel-1 and Sentinel-2 images used for the test area

Satellite image name	Acquisition date (yyyy/mm/dd)	Flight direction	Processing level
S1A_IW_GRDH_1SDV_20180821T004109_20180821T004134_023337_0289D5_B2B2	2018/08/21	Descending	L1-GRD (IW)
S1A_IW_GRDH_1SDV_20180821T130602_20180821T130631_023345_028A0A_C728	2018/08/21	Ascending	L1-GRD (IW)
S1A_IW_GRDH_1SDV_20180821T004044_20180821T004109_023337_0289D5_D07A	2018/08/21	Descending	L1-GRD (IW)

S1A_IW_GRDH_1SDV_20180821T130631_20180821T130656_023345_028A0A_E124	2018/08/21	Ascending	L1-GRD (IW)
S2B_MSIL1C_20180822T050649_N0206_R019_T43PFL_20180822T085140	2018/08/22	Descending	Level 1C

Figure 3.1 highlights the worst-affected districts in Kerala; western districts faced much more severe flooding than eastern districts because western districts are topographically flat (coastal plains). As mentioned in Table 3.1, four Sentinel-1 images of the affected area on the same date of 21 August 2018 were selected for testing. Two images were acquired during the ascending flight direction and two were acquired during the descending flight direction. The closest Sentinel-2 image of the same area is available for 22 August 2018. However, most of the area in this image has clouds. So finally, only the area belonging mainly to the Alappuzha district was selected because this image has no or very few pixels affected by clouds in the Sentinel-2 image. This was done to validate the detection from the Sentinel-1 image. In general, reference flood mask is generated using aerial images (Y. Li et al., 2019; Martinis et al., 2009) or optical images such as Worldview (Twele et al., 2016), Sentinel-2 (Clement et al., 2018; Tiwari et al., 2020). Therefore, authors adopted using Sentinel-2 images, which previously were successfully utilized for flood mapping (Caballero et al., 2019), on their own and as a flood reference mask to validate the results. To make a reference water mask from the Sentinel-2 image, MNDWI, false-colour composite using bands B12, B8, and B4 and the true-colour composite using B4, B3, and B2 bands were used, along with this visual inspection was performed to maintain the accuracy of the mask.

Dataset details

This study used a public dataset, Sen1Floods11 that was released during the 2020 Computer Vision and Pattern Recognition Workshop (Bonafilia et al., 2020) and generated by Cloud to Street, a public benefit corporation. Details of the dataset are given below.

The dataset is divided into two parts, one containing data related to flood events and another for permanent bodies of surface water. The permanent water data include images from the Sentinel-1 satellite constellation and corresponding labels from the European Commission Joint Research Centre (JRC) global surface water dataset. We mainly used the flood events dataset in this study, which has two types of labels: weakly labelled and hand labelled. Weakly labelled here means that the labels have not been checked for quality, as they were generated through semi-automated algorithms that use certain thresholds to separate water and non-water areas. The weakly labelled data have two kinds of labels generated from Sentinel-1 and Sentinel-2 images, respectively. These labels are binarized images containing ones (for water pixels) and zeros (for non-water pixels). Sentinel-1 weak labels were prepared using the Otsu thresholding method over the focal mean-smoothed VH band. For creating the weak labels from the Sentinel-2 images expert-derived thresholds of 0.2 and 0.3 were applied over the Normalized Difference Vegetation Index and MNDWI bands, respectively. These weakly labelled data have not been quality controlled and over- or under-segmentation is possible. The hand-labelled data were created using information from overlapping tiles of both Sentinel-1 and Sentinel-2. The manual classification was performed using the Sentinel-1 VH band and two false-colour images of Sentinel-2 (RGB: B12, B8, B4 and B8, B11, B4) that highlight the water areas in the optical images. The resultant labels are more accurate and have three values in the output: 1 (water pixels), 0 (non-water pixels), and −1 (clouds or cloud shadows).

Overall, 4830 non-overlapping chips were available to us that belong to flood events of 11 countries. Of these, 4385 chips are weakly labelled with corresponding S1Weak (Sentinel-1) and S2Weak (Sentinel-2) labels, while 446 chips are hand labelled and have corresponding quality-controlled labels. Each chip size is 512 × 512 pixels. All chips have overlapping Sentinel-1 and Sentinel-2 images. Sentinel-1 chips were created using du-al-polarized Sentinel-1 ground range detected (GRD) images. As these images have been downloaded from the Google Earth Engine, each image was pre-processed using the Sentinel-1 Toolbox by the following steps: thermal noise removal, radiometric calibration, terrain correction using SRTM 30, and finally conversion of both bands' values into decibels via log scaling. In contrast, Sentinel-2 chips are from raw Sentinel-2 MSI Level-1C images having all 13 bands (B1 to B12). The 13 spectral bands represent the top of atmosphere reflectance, scaled by 10,000.

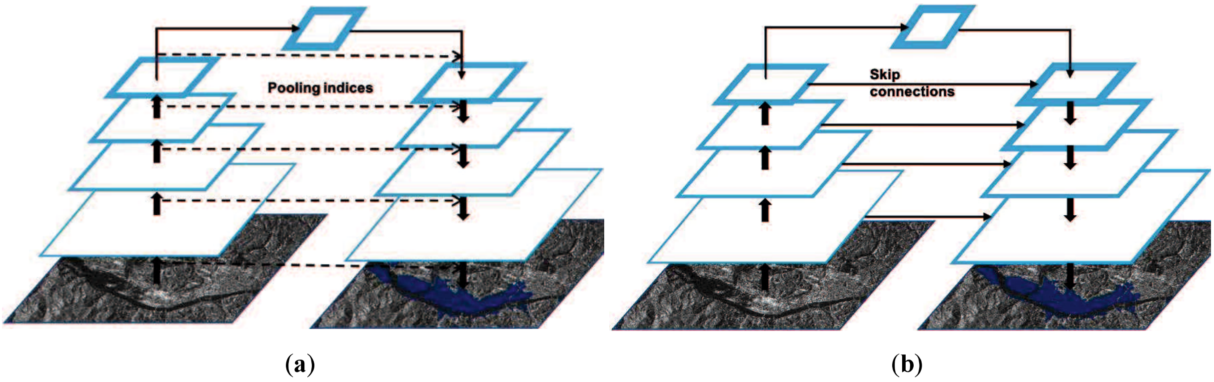
The hand-labelled data were split into three parts with a ratio of 60:20:20 into training, validation, and test sets. In contrast, all the weakly labelled data were used for training purposes only. In this way, the test set remained the same throughout the study, while training data could be changed according to our requirements, and we could do cross-comparison for different kinds of training data.

Methodology

Network's architecture

Segmentation networks are like autoencoders consisting of encoders that encode an image into the smaller latent space and decoders that revert to reach the same spatial size. In this research variants of autoencoders like networks, namely, SegNet-like and UNet-like architectures (Figure

3.2), were selected for extracting the water areas from Sentinel-1 chips. These networks, as shown in Figure 3.2, were selected because they are simple compared with other existing networks for segmentation such as HRNet, DANet, etc., and they also have shown great performance when the dataset is limited in size. Both networks can be divided into two parts, the contraction phase (encoder path), and the expansion phase (decoder path). Each block in the encoder path contains two convolution layers that have a kernel size of 3 × 3 and the 'same' padding along with the batch normalization (Goodfellow et al., 2016) and rectilinear unit (relu) activation layer. This is followed by a max-pooling layer with a size of 2 × 2 and a stride of 2. In this way, the convolution layer increases the number of features in the channel space (depth) while the max-pooling layer contracts the dimensions of the spatial feature space. Between both networks, the number of blocks in the encoder and decoder path remained the same, but the method for increasing the spatial size (up-sampling) in the decoder section was the main difference. Here, UNet uses up-convolution along with the skip connections to use the features from previous layers, while in SegNet, up-sampling in the decoder section uses pooling indices that are computed in the max-pooling step of the corresponding encoder blocks. Thus, in the case of the SegNet, only spatial information is transferred from the lower-level layers, while in the UNet, the low-level feature space is also transferred to the high-level feature space and concatenated with it at the corresponding levels. This passing of the low-level features to high-level becomes possible due to skip connection which can bypass the intermediate layers. The networks remain fixed across the training cases and in various band-combination inputs. This means only the shape of the input layer is modified while all intermediate and output layers remain constant.



(a) (b)

Figure 3.2 Network representations. (a) SegNet-like network. (b) UNet-like network.

Hyperparameters

For the entire study, the mini-batch size was selected as 16 and iterated over the whole dataset 200 times (epoch). The loss function used here was a custom loss function that used both Dice loss and binary cross-entropy (BCE) (Jadon, 2020; Sekou et al., 2019) in a weighted manner. While the Dice score mainly looks for the similarity of segmentation blob, BCE calculates pixel-wise variance. This means the dice score do capture the spatial information better than the BCE, which is why Dice loss was given a higher weight of 0.85, and BCE received a lower weight of 0.15. The Adam optimizer (Kingma & Ba, 2015) was used for training optimization, with an initial learning rate of 0.01. The learning rate is decayed for faster convergence and to avoid over-fitting. If there is no improvement (tolerance is set to 0.001) for continuously 10 epochs on a validation set, the learning rate is reduced by the factor of 0.8. The minimum value for learning has been fixed to 0.0001. The training was performed on a single NVIDIA Titan-V GPU. The whole model development and training were performed using the Tensorflow platform along with the Keras library in Python.

Training strategy

In total, three training cases were selected as shown in Figure 3.3: (1) training using more accurate hand labels (2) training using Sentinel-1 weak labels, and (3) training using Sentinel-2 weak labels.

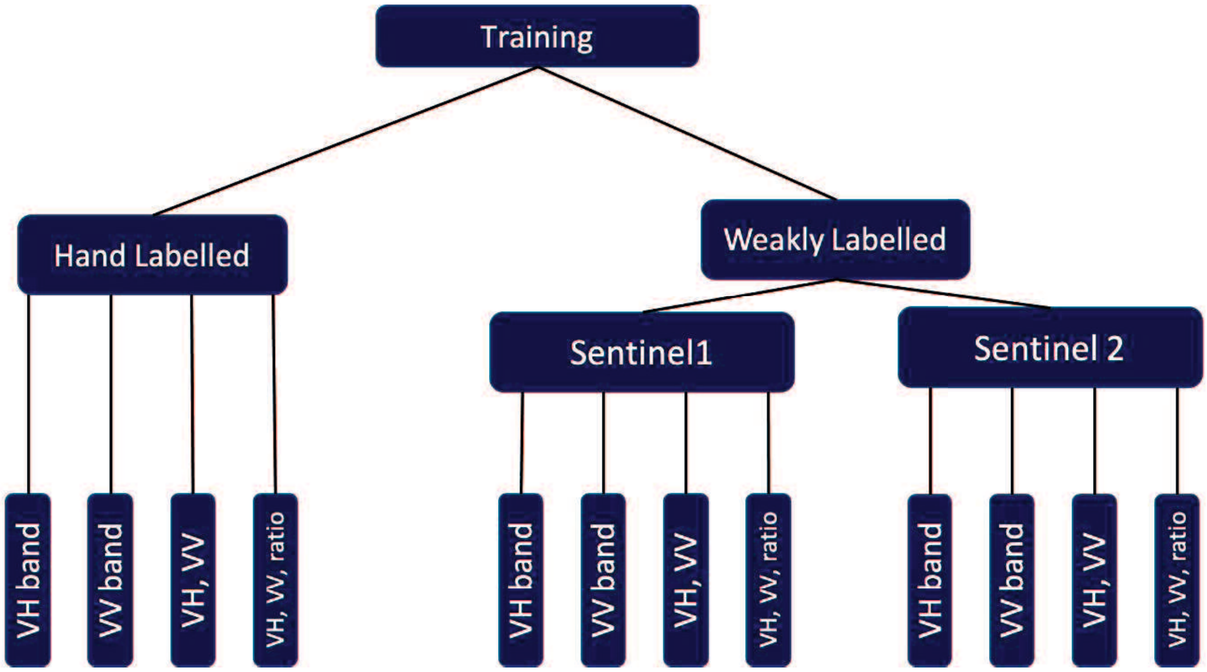


Figure 3.3 All the chosen training cases.

In each case, four SegNet-like and UNet-like networks were trained for the different band combinations: using both polarizations (VV, VH), using only cross-polarization (VH), using only co-polarization (VV), and using a ratio as the third band, making the input as VV, VH, and VH/VV. Here it should be noted that the VV and VH bands are already log scale, so the values of each pixel ranged between -50 and 1 dB. These inputs were normalized using min-max values so that the resultant values were between 0 and 1 before they were passed on for training. Also, for the calculation of the VH/VV ratio, we simply subtracted the log-scaled VH and VV bands due to the log properties:

$$\log\left(\frac{VH}{VV}\right) = \log(VH) - \log(VV)$$

Transfer learning

Transfer learning was also used to explore the option of making our model more adaptable and scalable. For this step, three cases were selected. In the first case, the whole network was retrained and pre-trained weights were used as starting weights rather than random weights, which are typically used during training from scratch. In the other two cases, we conducted training only during the contraction phase (encoders) while freezing the expansion phase (decoders) and vice versa. Transfer learning has various benefits, such as the ability to include more training data in the future to further tune the network, faster convergence due to pre-trained weights (L. Li et al., 2019), and the possibility of extending the trained model to be used with a new set of satellite images, such as from a different SAR satellite (Z. Huang et al., 2020).

Testing strategy

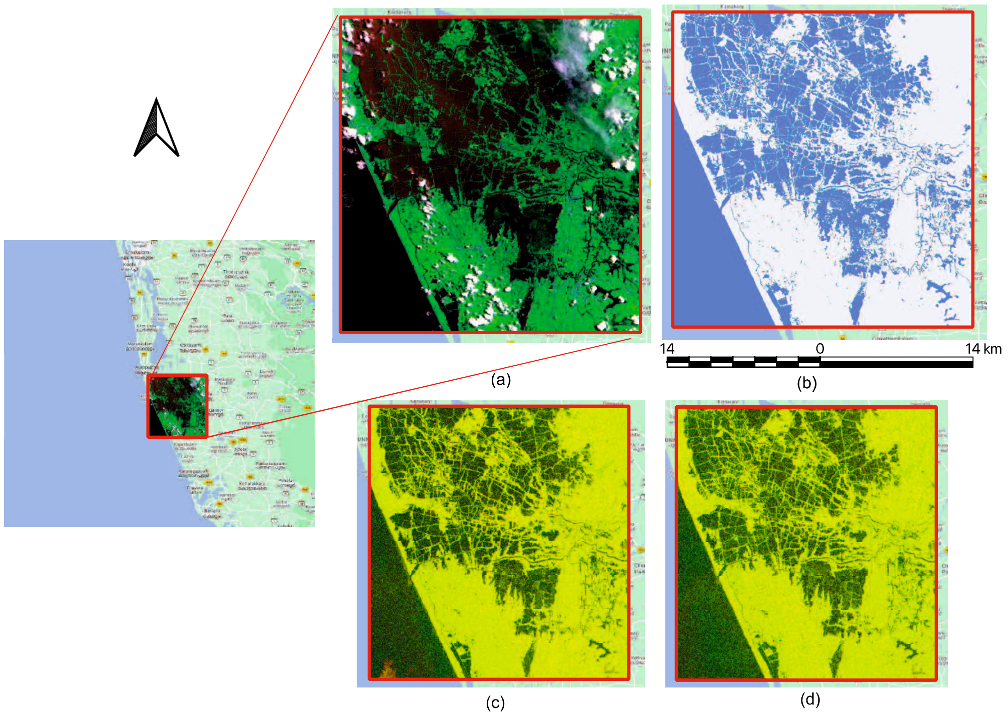
Testing on the Test Dataset

Three test cases were selected: all surface water detection, only permanent water detection (using corresponding JRC labels), and flooded water detection (difference between all water and permanent water). Because some of the image chips did not have any permanent water, they were removed from the test set of permanent water. In total, we had 90 test image chips for all water and flood water detection, and 54 chips for permanent water detection. All the trained networks, totalling 24 networks (SegNet and UNet), were tested over the given three test cases.

Testing as an off-the-shelf model on the whole image during the 2018 Kerala Floods

To verify the generalizability of the trained model for use as an off-the-shelf model during an emergency, a completely different flood event, the 2018 Kerala floods was selected. The first validation flood mask was prepared using the Sentinel-2 image. Although most of the Sentinel-2 image was covered by clouds, fortunately, the area most affected by the flooding had minimal cloud cover, so that area was selected, amounting to 794 km2 (Figure 3.4). After obtaining the desired area, the semi-automatic classification in QGIS was used over the Sentinel-2 bands B2, B3, B4, B8, B11, and B12 along with the MNDWI. In this step, certain regions of interest for water pixels were manually chosen across the selected area, after which classification was performed. However, this classification still had numerous errors and cloud obstructions. Thus, after the classification was complete, a manual inspection was performed to further improve the classified results. In the end, these accurate classified results were exported as a binary flood mask, and this mask performed the role of ground truth in the validation phase.

The Sentinel-1 images were first pre-processed using the European Space Agency's snappy package in Python to sequentially perform thermal noise removal, radiometric calibration, speckle filtering, and terrain correction. As the selected area lies where two satellites images from the same flight direction meet, both images were merged and gap-filled using QGIS. The same method was also applied to the two images from different flight directions. After this, two separate methods were employed for water area classification. The first is a thresholding method where a threshold was selected based on a combination of minimum distance and the Otsu method, which was implemented using scikit-image learning libraries. In the second method, our best-performing trained model (after transfer learning), generated in the previous step, was used on the pre-processed image to obtain a flood map as binarized output. The whole image with a size of 13,797×7352 pixels was processed within 1 minute and transformed into a binarized output. After this, the same area as that selected in the Sentinel-2 was clipped from the output for evaluation purposes.



(a) (b) (c) (d)

Figure 3.4 Details of the selected area from the test site. (a) Sentinel-2 composite image in false colour (B12-B8-B4). (b) Created validation flood mask. (c) and (d) Descending and ascending SAR images of the area of interest, respectively.

Accuracy evaluation measures

For measuring the performance of the different approach, this study uses Intersection over Union (Eq 1), F1 score (Eq 2), omission and commission error (Eq 3-4). Due to imbalanced classes (flooded pixels are very low in comparison to non-flooded pixels), IoU and F1 score is better metrices to choose from as they are focusing on foreground pixels (flooded pixels) than background (non-flooded pixels). In the equation TP, FP, FN, denotes to the ‘true positive’, ‘false positive’ and ‘false negative’ respectively.

$$IoU = \frac{Ground\ Truth\ \cap Predicted}{Ground\ Truth\ \cup Predicted} = \frac{TP}{TP + FP + FN} \qquad \text{Eq 1}$$

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \qquad \text{Eq 2}$$

$$\text{Omission error} = \frac{FN}{FN + TP} \qquad \text{Eq 3}$$

$$Commission\ error = \frac{FP}{FP + TP} \qquad \text{Eq 4}$$

Results and validation on the test set

Results

Figure 3.5 and Figure 3.6 show the different models' mean IoU (mIoU) over the whole test set for SegNet and UNet, respectively. The x-axis shows the training cases, while the y-axis represents the mIoU. The detailed quantitative results from each network along with the respective errors are presented in Table 3.2 and Table 3.3. Columns in the table represent the three detection test cases, namely, permanent water, flooded water, and all surface water. For each test case, the three evaluation criteria, mIoU, omission error (Om.), and commission error (Comm.), as used in (Bonafilia et al., 2020), is presented. In the rows of the tables, the three training cases, namely, training using Sentinel-1 weakly labelled, Sentinel-2 weakly labelled, and hand-labelled data and their corresponding results are given. Each training case further has four variations consisting of different combinations of SAR bands. Along with our results, the baseline results from (Bonafilia et al., 2020) are also shown for each training case, as well as Otsu thresholding results for better comparison.

In both types of networks, a common pattern can be seen. For the permanent water, the band combination of VV, VH and VH/VV ratio performed best in most of the cases, while in cases of flooded water and all surface water, the input with both polarizations VV and VH gave the best results. Note that when we used only the co-polarized band (VV), the network trained on weak labels performed worst, especially in the case of flooded water detection and all surface water detection, with a very high omission error. The cause can be understood by the property of SAR backscattering, which in the case of flooded vegetation or agriculture field may show very high backscattering in the co-polarized band due to double bounce (caused by the small wavelength of C-band SAR). A more detailed explanation is provided in the discussion section.

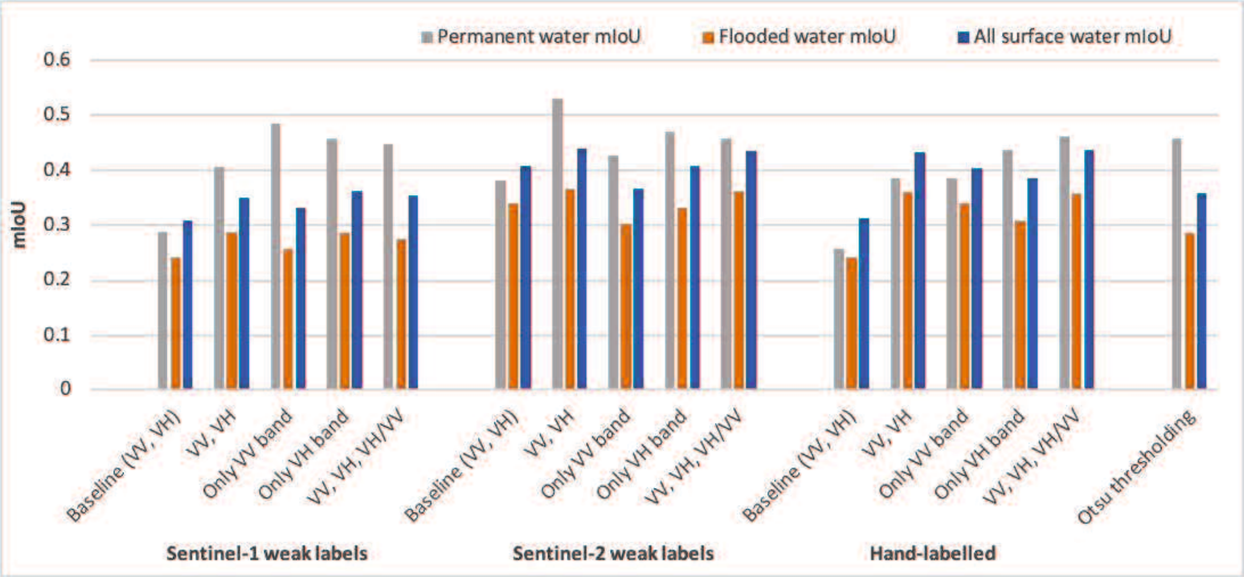


Figure 3.5 Test results using SegNet, mIoU with different labels, and the band combinations

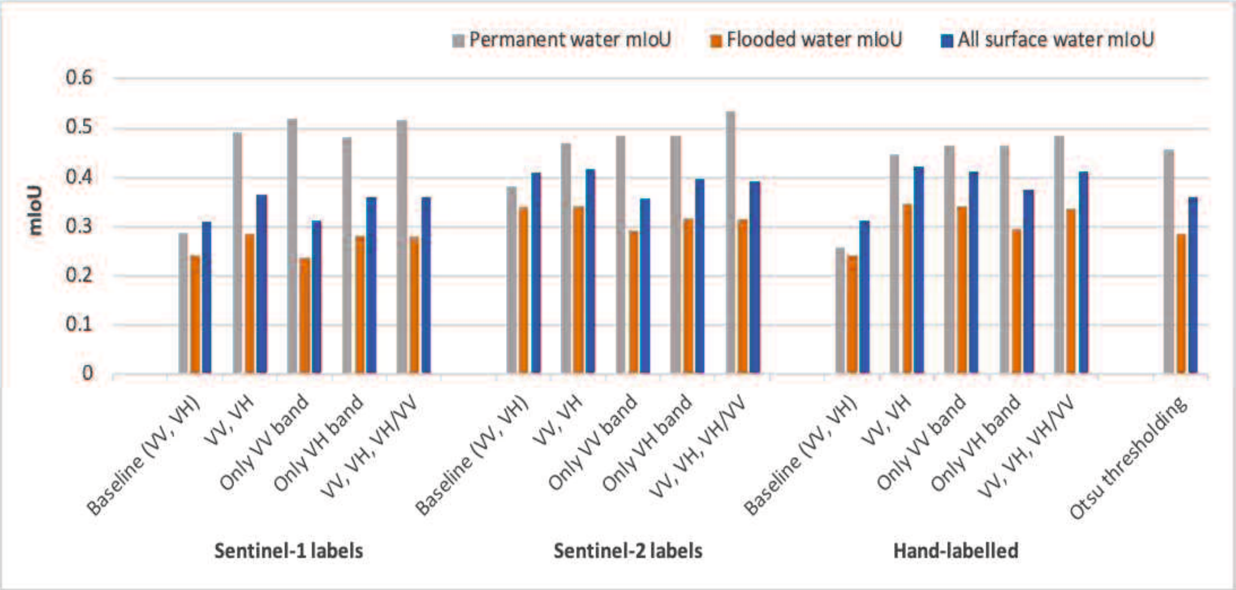


Figure 3.6 Test results using UNet, mIoU with different labels, and the band combinations

Table 3.2 Performance of different SegNet models over all the test cases

SegNet/Dataset and Band Used	Permanent Water			Flooded Water			All Surface Water		
	mIoU	Om.	Comm.	mIoU	Om.	Comm.	mIoU	Om.	Comm.
Sentinel-1 weak labels									
VV, VH	0.492	0.008	0.046	**0.286**	0.378	0.044	**0.364**	0.249	0.044
Only VV band	**0.519**	0.255	0.032	0.238	0.474	0.037	0.313	0.397	0.037
Only VH band	0.482	0.009	0.048	0.282	0.382	0.049	0.359	0.251	0.049
VV, VH, VH/VV	0.515	0.011	0.043	0.281	0.41	0.042	0.360	0.269	0.043

Sentinel-2 weak labels									
VV, VH	0.469	0.008	0.024	**0.342**	0.365	0.021	**0.417**	0.239	0.020
Only VV band	0.484	0.315	0.021	0.292	0.434	0.016	0.357	0.393	0.016
Only VH band	0.483	0.006	0.025	0.318	0.382	0.021	0.396	0.252	0.020
VV, VH, VH/VV	**0.534**	0.014	0.018	0.315	0.438	0.014	0.392	0.290	0.014
Hand labelling									
VV, VH	0.447	0.005	0.029	**0.347**	0.341	0.024	**0.421**	0.223	0.024
Only VV band	0.463	0.285	0.031	0.342	0.355	0.023	0.412	0.331	0.023
Only VH band	0.463	0.011	0.032	0.296	0.429	0.021	0.374	0.283	0.021
VV, VH, VH/VV	**0.484**	0.007	0.024	0.336	0.404	0.019	0.411	0.265	0.019
Benchmark from (Bonafilia et al., 2020)									
Otsu thresholding	0.457	0.054	0.085	0.285	0.151	0.085	0.359	0.143	0.085
Baselines from (Bonafilia et al., 2020)									
Sentinel-1 weak labels (VV, VH)	0.287	0.066	0.135	0.242	0.119	0.100	0.309	0.112	0.997
Sentinel-2 weak labels (VV, VH)	0.382	0.121	0.053	0.339	0.268	0.078	0.408	0.248	0.078
Hand labeling (VV, VH)	0.257	0.095	0.152	0.242	0.135	0.106	0.313	0.130	0.106

Table 3.3 Performance of different UNet models over all the test cases

UNet/Dataset and Band Used	**Permanent Water**			**Flooded Water**			**All Surface Water**		
	mIoU	**Om.**	**Comm.**	**mIoU**	**Om.**	**Comm.**	**mIoU**	**Om.**	**Comm.**
Sentinel-1 weak labels									
VV, VH	0.406	0.006	0.052	**0.288**	**0.352**	0.050	0.349	**0.231**	0.050
Only VV band	**0.485**	0.285	0.035	0.257	0.445	0.038	0.332	0.389	0.038
Only VH band	0.457	0.008	**0.021**	0.285	0.390	0.043	**0.362**	0.256	0.043
VV, VH, VH/VV	0.446	**0.006**	0.039	0.275	0.396	**0.037**	0.353	0.259	**0.037**
Sentinel-2 weak labels									
VV, VH	**0.529**	0.009	**0.017**	**0.366**	0.358	**0.014**	**0.439**	0.236	**0.014**
Only VV band	0.427	0.293	0.021	0.303	0.402	0.019	0.367	0.364	0.019
Only VH band	0.469	**0.004**	0.025	0.332	0.362	0.021	0.407	0.236	0.021
VV, VH, VH/VV	0.458	0.004	0.029	0.362	**0.313**	0.024	0.434	**0.205**	0.024
Hand labelling									
VV, VH	0.386	0.005	0.042	**0.361**	**0.274**	0.035	**0.432**	**0.181**	0.035
Only VV band	0.386	0.289	0.038	0.339	0.315	0.029	0.404	0.306	0.029
Only VH band	0.436	0.005	0.035	0.309	0.363	0.029	0.386	0.236	0.029
VV, VH, VH/VV	**0.462**	**0.003**	**0.027**	0.359	0.309	**0.024**	0.436	0.202	**0.024**
Benchmark from (Bonafilia et al., 2020)									
Otsu thresholding	0.457	0.054	0.085	0.285	0.151	0.085	0.359	0.142	0.085
Baseline from (Bonafilia et al., 2020)									
Sentinel-1 weak labels (VV, VH)	0.287	0.066	0.135	0.242	0.119	0.100	0.309	0.1124	0.997
Sentinel-2 weak labels (VV, VH)	0.382	0.120	0.053	0.339	0.268	0.078	0.408	0.2482	0.078
Hand labeling (VV, VH)	0.257	0.094	0.152	0.242	0.135	0.105	0.312	0.1297	0.105

In contrast with the results in Bonafilia et al. (2020), where the best results came from Otsu thresholding, our results clearly show that both SegNet and UNet convincingly surpass the Otsu thresholding, as well as the baseline results, in all training cases. While in the case of permanent water our models show as much as 50% enhancement over the baseline, for all surface water our model also shows an improvement of 20% over the Otsu thresholding.

In contrast with the results in Bonafilia et al., where the best results came from Otsu thresholding for permanent water, our results clearly show that both SegNet and UNet convincingly surpass the benchmark data by Otsu thresholding, as well as the baseline results, in all training cases. However, other results for the flooded water and all-surface water are in sync with Bonafilia et al., as the best detection in the case of SegNet, as well as UNet, comes from the models trained using Sentinel-2 weak labelled dataset. Moreover, in the case of flooded water, our models show as much as 50% enhancement over the baseline and for all surface water also our model shows an improvement of around 40% with the UNet model trained with the hand labelling dataset.

Overall, the UNet-like networks outperformed the SegNet-like networks in detecting the flooded water and all surface water, which is the target in the study. One of the reasons may be the use of skip connections, which propagate the shallow layer features to the deeper layers, helping to create a better feature set for pixel-level classification. For this reason, subsequent processing was done using UNet only. This means that features from the encoder layers played a more important role in processing the SAR images, and this was further proved when transfer learning was used. In other words, encoder retraining gives better results than does decoder retraining.

The weak labelling technique has the advantage of creating a larger set of training samples in an automated way in a shorter time and less manpower than the hand labelling. A larger number of training samples helps in finding greater insight. However, hand-labelled data have consistency and include cases that could not be captured by weak labelling techniques. Therefore, transfer learning was employed to take advantage of both situations, namely, more samples for generalization and accurate labels for tuning. As our focus in the study is flood mapping, the model that was trained using Sentinel-2 weak labels with both polarization bands (VV and VH) was selected for transfer learning because it performed best among all other band combinations, for “flooded water” and “all surface water” detection. Then, transfer learning was employed on it using hand-labelled data and retraining it for the three cases, namely, retraining the whole model, retraining only the expansion phase, and retraining only the contraction phase, with the pre-trained weights (Figure 3.7). The results are presented in

Table 3.4. Overall, the model retrained on the encoder part showed the best result and that was used for real-time flood area detection at the chosen test site.

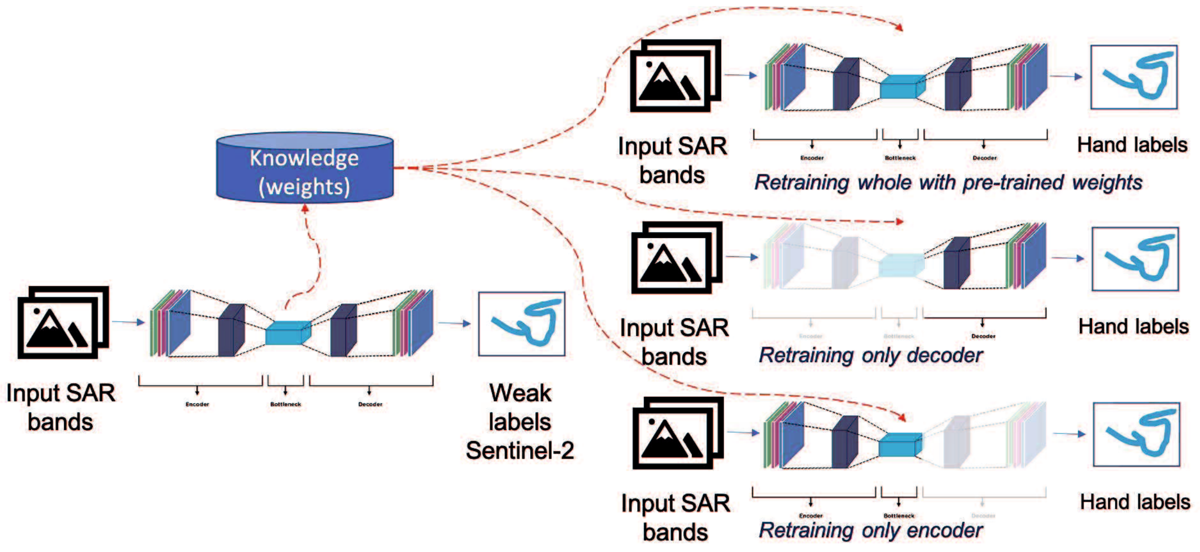


Figure 3.7 Transfer learning used cases in the study.

Table 3.4. Results of the transfer learning were performed using hand-labelled data on the UNet model trained on Sentinel-2 weakly labelled data with both polarizations (bold text represents the best result).

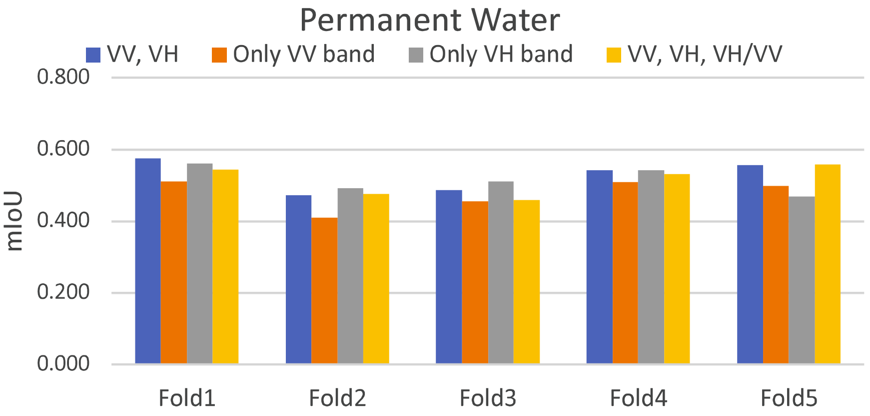
Transfer Learning/Dataset	**Permanent Water**			**Flooded Water**			**All Surface Water**		
	mIoU	**Om.**	**Comm.**	**mIoU**	**Om.**	**Comm.**	**mIoU**	**Om.**	**Comm.**
Hand labelling									
Whole model	0.530	0.0051	0.0264	0.409	0.3494	0.0207	0.483	0.2287	0.0207
Whole decoder	0.531	0.0054	0.0324	0.366	0.3745	0.0238	0.443	0.2451	0.0238
Whole encoder	**0.532**	**0.0041**	**0.0243**	**0.420**	**0.3086**	**0.0204**	**0.494**	**0.2042**	**0.0204**
Otsu thresholding (OT)	0.457	0.054	0.0849	0.285	0.151	0.0849	0.3591	0.1427	0.0849
% improvement over OT	+16.4	−92.4	−71.37	+47.3	+104.3	−75.9	+37.6	+43.1	−75.9

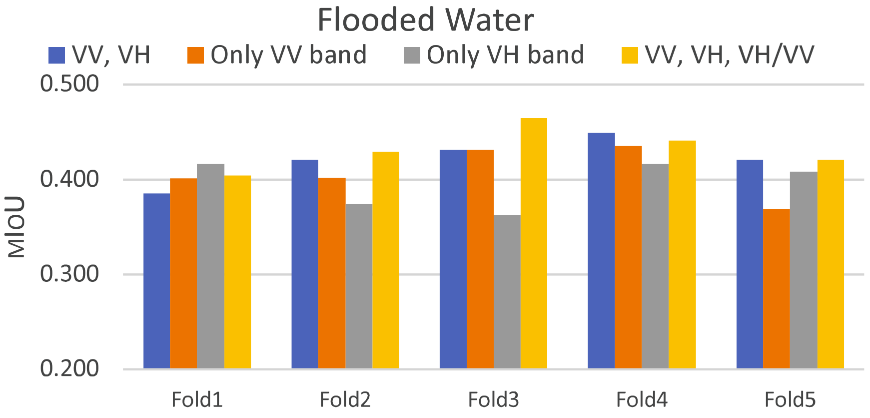
K-fold cross-validation

To investigate the generalization capability of the model as well as to ensure that model is not overfitting on the given data, K-fold cross-validation was used. For the implementation we have used hand-labeled data. First dataset has been divided into five equal parts or folds, and then four model for each band combination has been trained by leaving one part for testing and using four parts for training. This way we have trained five sets of models by leaving different part as a test set in every time, to cover the whole dataset. The result of the models against the permanent water, flooded water and all-surface water is shown in Figure 3.8. The average of all five models with each band combination is mentioned in Figure 3.6 along with the standard deviation. Results in Table 3.5 suggest that our models are consistent throughout different folds with standard deviation ranging between 2–4%.

Table 3.5 Five-fold validation results on the hand labelled data.

	Permanent Water		**Flooded Water**		**All surface water**	
Band used	**Average mIoU**	**Std. Dev.**	**Average mIoU**	**Std. Dev.**	**Average mIoU**	**Std. Dev.**
VV, VH	0.524	0.040	0.421	0.021	0.473	0.026
Only VV band	0.474	0.039	0.407	0.024	0.454	0.025
Only VH band	0.514	0.033	0.395	0.023	0.451	0.025
VV, VH, VH/VV	0.511	0.039	0.432	0.020	0.484	0.024





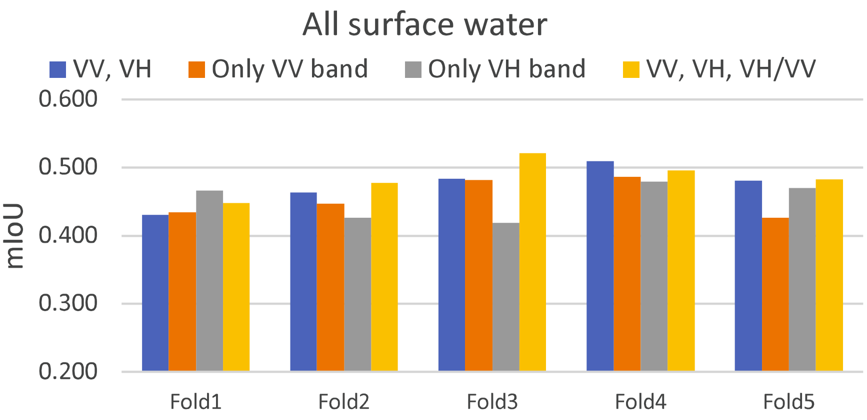


Figure 3.8 Model's performance with different cases of waters (permanent water, flooded water and all-surface water)

Quantitative results on the error-free test set

As there were few wrong annotations which were presents due to the manual mistakes as well as some annotation includes very small water bodies (only few pixels wide) which were simply not possible to detect from the side looking Sentinel-1 images. So, to check the effect of these problematic test dataset, we have removed them and used same model to check the performance on this modified test set which now includes total 71 image chips in place of 90. The result of which is presented in the Table 3.6

Table 3.6 Quantitative results on error-free test set

Transfer learning/ dataset	**Permanent water**			**Flooded water**			**All surface water**		
	mIoU	**Om.**	**Comm.**	**mIoU**	**Om.**	**Comm.**	**mIoU**	**Om.**	**Comm.**
Hand labelling									
Whole encoder TL (error-free test set)	**0.661**	0.0032	0.0341	**0.511**	0.255	0.0453	**0.612**	0.166	0.0452
Whole encoder TL (Old test set)	0.532	0.0041	0.0243	0.420	0.307	0.0204	0.494	0.204	0.0204
Otsu thresholding (OT)	0.457	0.0542	0.0849	0.285	0.151	0.0849	0.359	0.143	0.0849

Results on the test site

The model resulting from the transfer learning performed notably well, and it was used on the test site in both ascending and descending flight directions. As shown in Table 3.7, it gave a better result than did the thresholding method. Moreover, the omission error was reduced significantly from around 16% to 6%, which is a very important criterion in emergency mapping, where omission error should be as low as possible. This means that false negatives should be fewer, even when some false positives may creep in. False negatives are a problem because leaving a flood-affected area off the map may lead to bad decision making – such as failing to evacuate or people travelling into the flooded regions.

Figure 3.9 shows the merged SAR images of the ascending flight direction and corresponding combined result of the surface water detection by Deep learning (our method) and Otsu thresholding. In the detection result, the white and black pixels are representing that both methods have classified the same either water or non-water, respectively. Contrarily to the red and cyan pixels illustrate that both methods have been classified differently. Cyan pixels imply that our method has classified the pixels as water whereas the thresholding method classified it as non-water and just the opposite in the case of Red pixels. In general, thresholding suffers from the noise in the output, as is visible in the combined results in terms of salt and pepper noise, as well as in the yellow and green insets. Owing to such kind of noise, a post-processing step, such as morphological erosion-dilation or minimal mapping unit application, is required after thresholding (Landuyt et al., 2019). The yellow rectangle displays a partially flooded agricultural area that was detected successfully by the deep learning model (in cyan color). In addition, the area shown by the green rectangle, which contains a few oxbow lakes on its far-right side, was successfully segmented by our model. In contrast, the blue rectangle shows the area around Kochi Port, which is one of the largest ports in India and docks multiple large vessels. This area produced some of the brightest pixels, and our method was not able to detect water in that area, while the thresholding method was able to achieve better results (red pixels). One of the reasons that the water was not detected by our method is that deep learning models learn the contextual information through spatial feature mapping, and it is a rare phenomenon to have water pixels covered by brighter pixels (in this case from ships). One way to detect such kinds of rare events is by including a few similar patterns in the training set or using some other ancillary data.

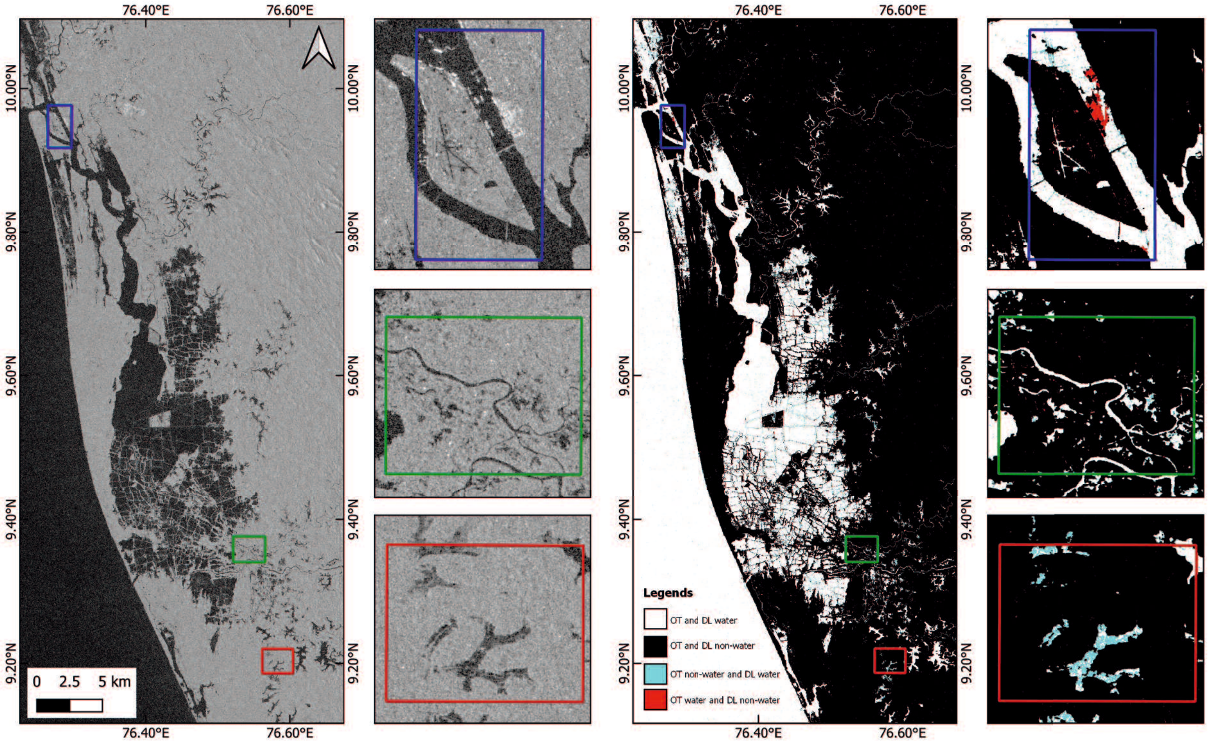


Figure 3.9 SAR (VH band) image and the corresponding combined result of Otsu thresholding (OT) and deep learning (DL) in a single image. White and black pixels in the resulting image represent pixels detected by both algorithms as water and non-water, respectively. Cyan and red pixels show areas where the algorithms differed, with cyan representing areas classified as water by DL but as non-water by OT and red representing areas classified as water by OT but as non-water by DL.

Table 3.7 Evaluation of the different methods on the test site (2018 Kerala flood)

Method	**Images**	**IoU**	**F1 score**	**Om. error**	**Comm. error**
Minimum and Otsu thresholding	Merged images of ascending flight direction	0.8394	0.9127	0.1328	0.0367
Our model (after transfer learning)	Merged images of ascending flight direction	**0.8849**	**0.9389**	**0.0587**	**0.0635**
Minimum and Otsu thresholding	Merged images of descending flight direction	0.8214	0.9019	0.1535	0.0347
Our model (after transfer learning)	Merged images of descending flight direction	**0.8776**	**0.9348**	**0.0661**	**0.0642**

Discussion and prospects

The results presented in section 3 allow us to make the following observations:

1) When the labels are weak, models trained on the co-polarization VV band performed poorly in comparison to models trained on the cross-polarization VH band. One of the reasons can be the high sensitivity of co-polarization towards rough water surfaces, for example, due to wind, as described by Manjushree et al. (Manjusree et al., 2012) and Clement et al. (2018). However, for

hand-labelled data, VV performs better than VH, especially for flooded areas. Figure 3.10 shows the results from the models trained on different band combinations. Because the training set here was hand-labelled, VV performed mostly better than VH bands except in rows 6 and 7. One of the interesting outcomes was that the three bands combined (VV, VH, and their ratio) gave the best results, except for the first row in Figure 3.10. This combination provided very good improvement in some of the difficult test cases, as in rows 5–7. This was particularly interesting as no new information is provided in the third band, it is just the ratio of already present input bands.

2) Models trained on Sentinel-2 weakly labelled data gave better results in comparison to Sentinel-1 weakly labelled data, which is consistent with the results of Bonafilia et al. (Bonafilia et al., 2020). Moreover, the models' trained onhand-labelled data approximately matches the accuracy of the models trained with Sentinel-2 data and sometimes even beat them despite limited samples, which goes against the results of Bonafilia (Bonafilia et al., 2020), who concluded that hand-labelled data are not necessary for training fully convolutional neural networks to detect flooding. We have demonstrated that models trained with hand-labelled data perform better throughout, as shown in Tables 2 and 3. Figure 3.11 shows a few examples of the improvement achieved by hand-labelled data. However, sometimes models trained with hand-labelled data give over-detection, as can be seen in the red circled areas in the first and last rows of the figure.

3) Successful implementation of transfer learning proves two things: First, there is no substitute for more accurate labels (hand-labelled data), as can be seen by the improved results. Second, it is a good approach to generate many training samples automatically, and a model trained on more samples gives better generalization. Further, we can use transfer learning to tune the model for our given test set. However, another interesting result is that, for finding surface water in SAR images, general features play a larger role than do specific features. As explained by Yosinski et al. (2014), layers close to the input, encoder blocks in our case, are responsible for general feature extraction, and deep layers are responsible for obtaining specific features (Yosinski et al., 2014). In our experiments, freezing the expansion phase and retraining the contraction phase gave the most favourable result. This can be further explored with different architectures; if the same behaviour persists, then we may use many shallow layer networks, making an ensemble to detect water areas from SAR images without wasting too many resources. The enhancement in water area detection using transfer learning is presented in Figure 3.12. Some of the examples, such as rows 1, 2, and 5, show tremendous improvement.

4) If we look only at mIoU in the test dataset, then its value, which was less than 0.5, does not present a good picture of the surface water detection. However, if we see some examples of the test set true labels along with the detected mask, such as in Figure 3.11, where we can see that the detection is quite accurate, especially by the model trained on hand-labelled data. Similar accuracy is seen in Figure 3.12, which shows the results of transfer learning models. Some of the reasons for low mIoU can be understood in Figure 3.13. In rows 1 and 2 of Figure 3.13, where a very narrow stream has been labelled, this stream is either not visible due to mountainous terrain (first row) or trees growing along with it (second row), and it becomes difficult to identify any significant water pixels in the SAR image. Another issue is having very small water bodies containing very few pixels scattered over the whole image (row 3). In this case, even though the omission error will be lower, the IoU will be near zero, affecting the mIoU of the whole test dataset. Moreover, a few incorrect labels are present in the test dataset. Some examples of this are shown in rows 5 and 6, where the red ellipses show the locations of incorrect labels. In these situations, even though our model is performing quite well, the IoU becomes very low or in some cases goes to zero, such as in the last row. Whereas, according to the given label, there are no water bodies, so the intersection will be zero and the union will be the detected water body pixels, which will result in an IoU of zero. Moreover, there are also many possible scenarios where, due to the special properties of the SAR, the detection is not accurate, such as in the case of row 4 in Figure 10. This area was flooded in a field with sparse vegetation, as can be seen in the true-colour image in the last row of Figure 3.14. This creates a double bounce from the specular surface of the water and vegetation in the co-polarized band (VV). This anomaly is the reason that the model is not able to identify it as a flooded field. Similar examples are shown in the first row of Figure 3.14, where sand deposits in the river have high backscatter in the VV band, though in general, flat sandy terrain works as a specular surface. One possible reason for the high backscatter is a wavy pattern in the sand due to wave action, which may also result in a double bounce.

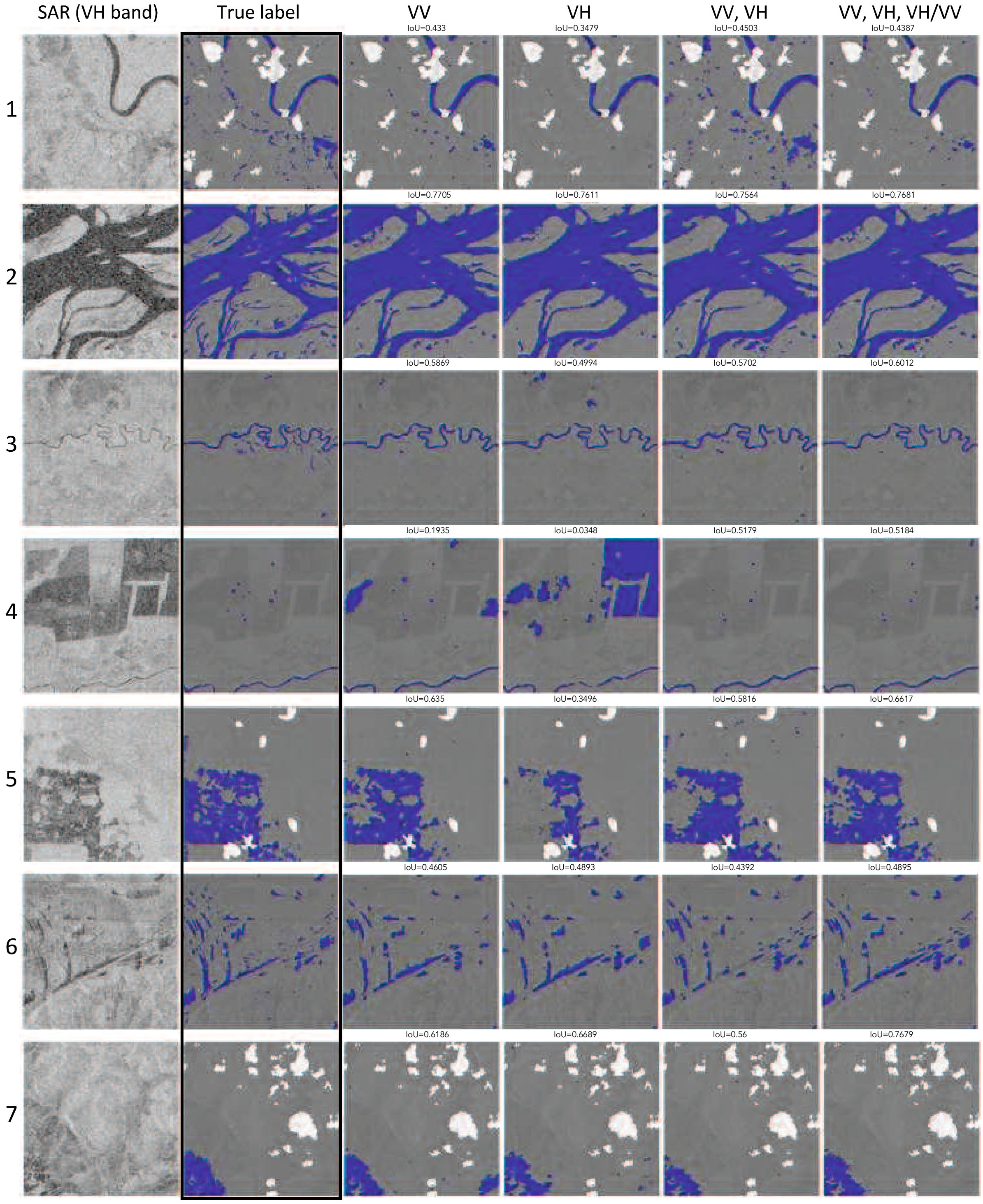


Figure 3.10 Results from the models trained using different band combinations with the hand-labelled dataset

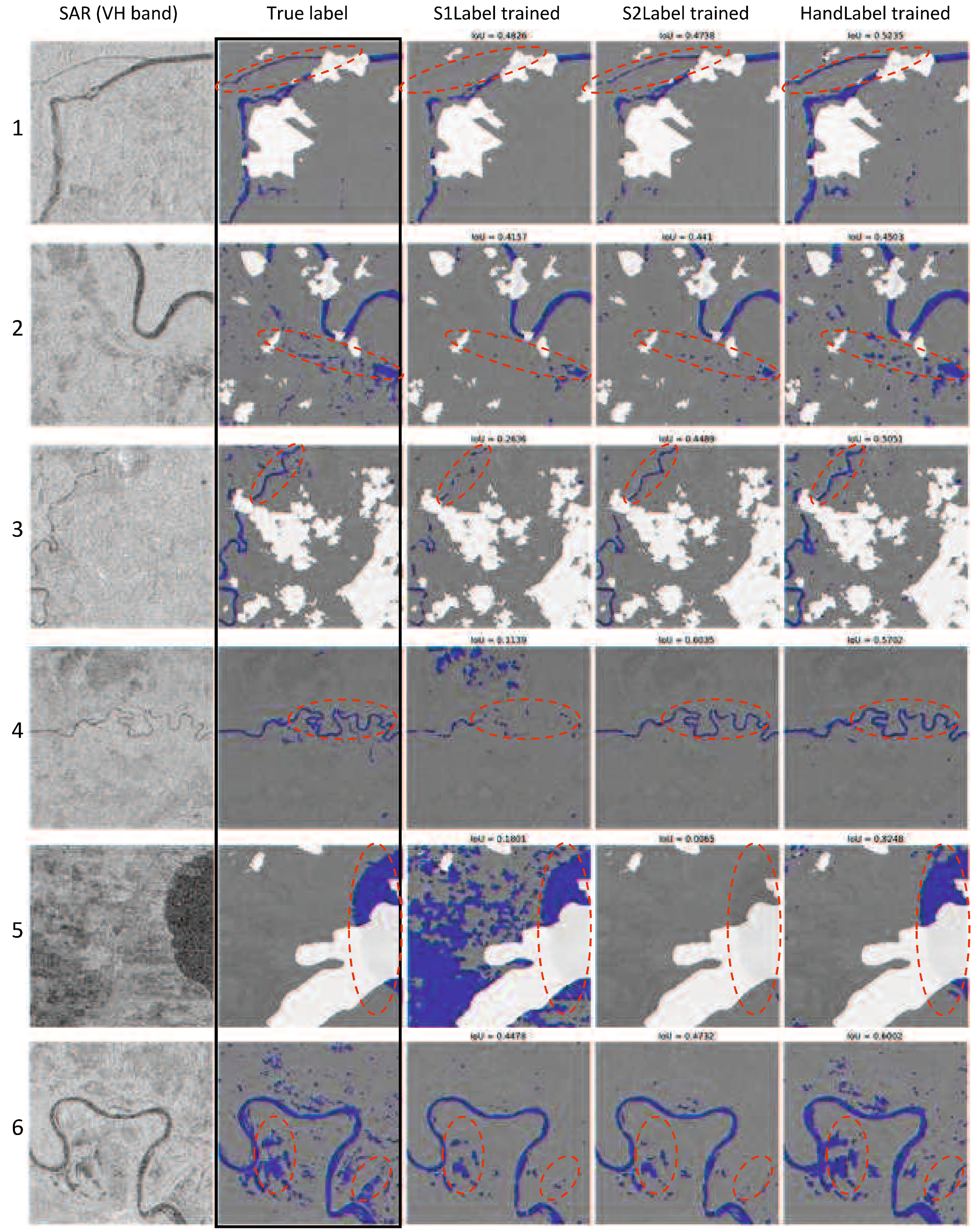


Figure 3.11 Results of the models trained with bands VV and VH combined over all three training sets

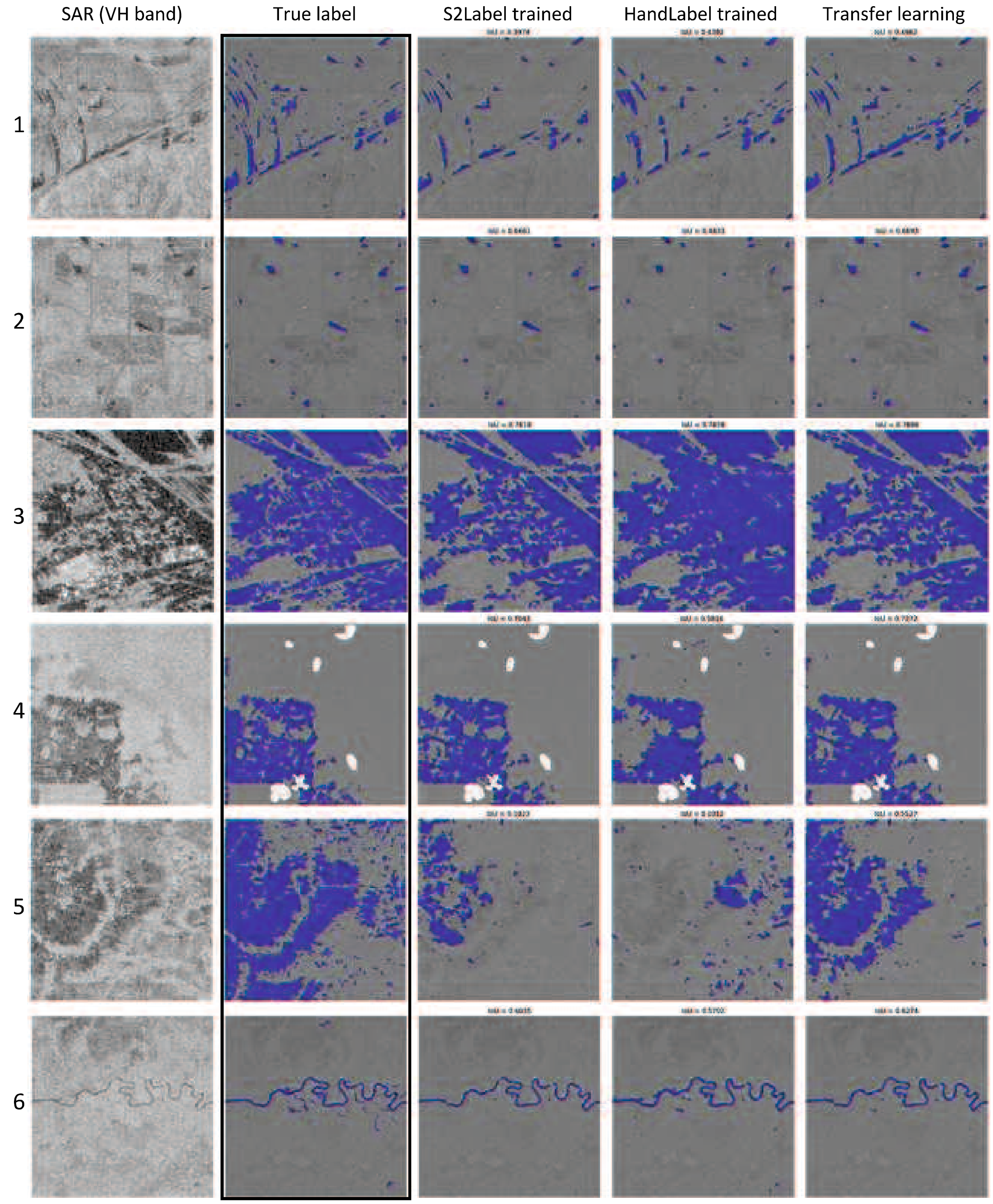


Figure 3.12 The improved result, achieved by using transfer learning. Results are from the model trained using the VV and VH bands combined and the Sentinel-2 weakly labelled dataset and retrained using the hand-labelled dataset.

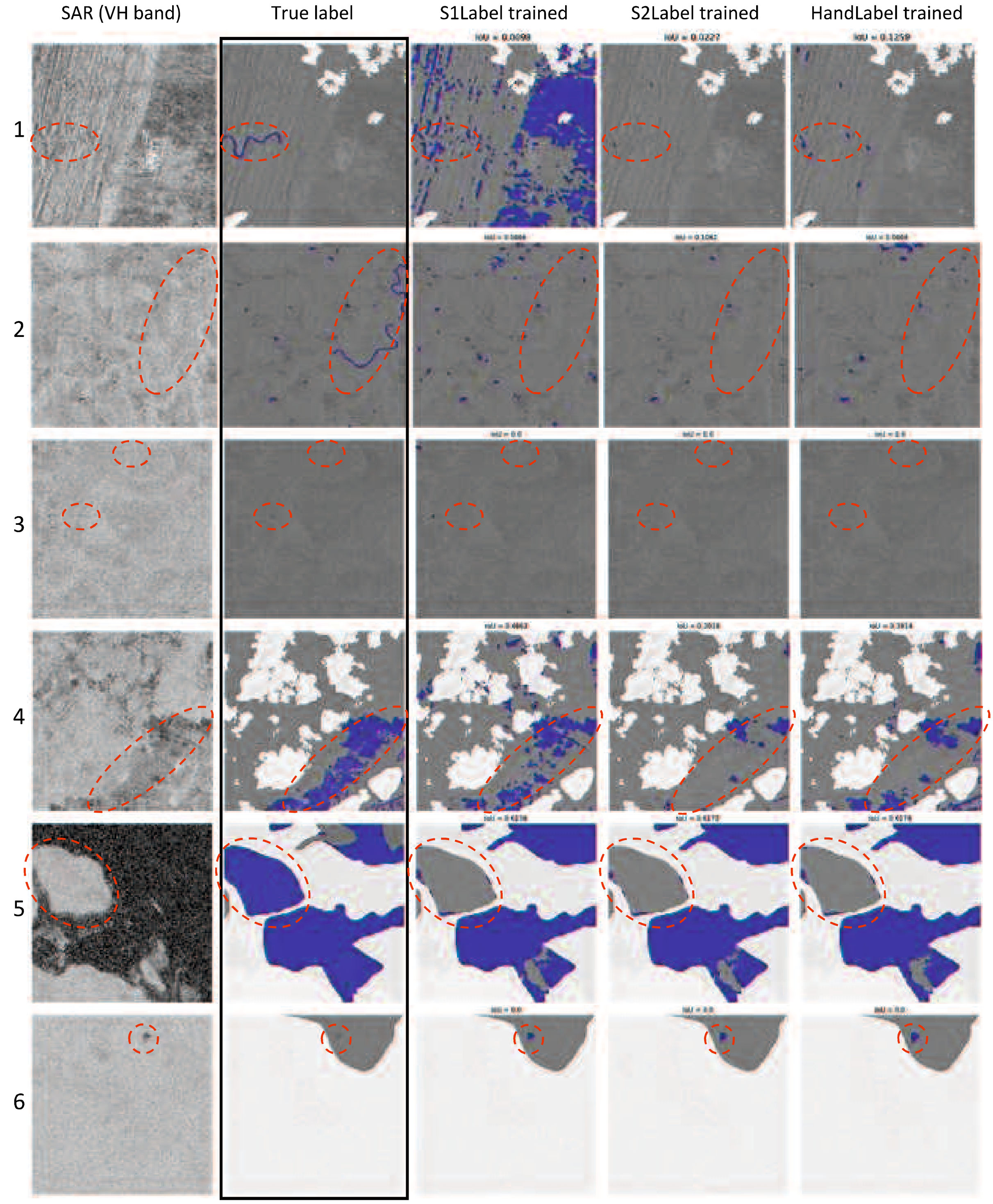


Figure 3.13 Major errors (circled by red dotted lines) in results obtained from the model trained using the VV and VH bands with the hand-labelled dataset

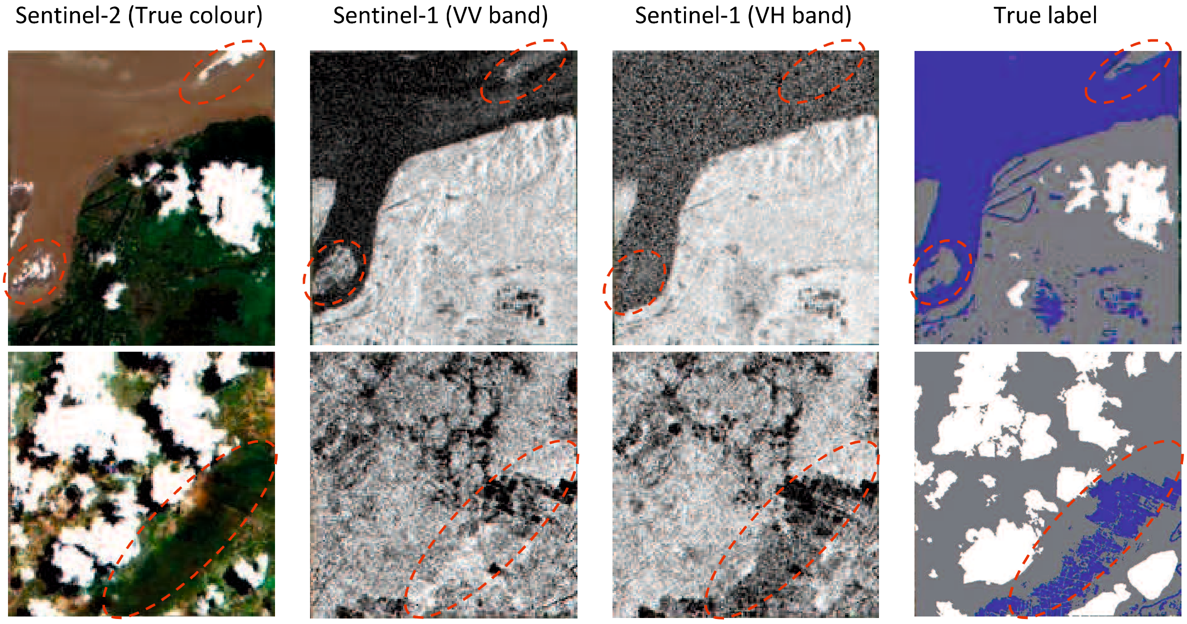


Figure 3.14 Some unique cases, where the classification behaviour is quite different with different polarizations (areas circled by red dotted ellipses). Row 1 is the case of river sand, and row 2 represents shallow flooding in agriculture fields with sparse vegetation.

Summary of the chapter

This chapter has presented the methodology for near-real-time flood mapping. It has been implemented using the Sen1Floods11 dataset that is available publicly. Results have shown a notable improvement over the baseline. Later, the trained model was used as an off-the-shelf model on a completely different site for validation purposes and the model has shown improved performance in comparison to the Otsu thresholding method which is one of the major algorithms used for the SAR image classification. The results of this chapter are presented in the Journal paper Katiyar et al., (2021).

Chapter 4 **Training dataset preparation strategy**

Introduction

As satellite scenes are too big that is why we need to create image chips out of them (Han et al., 2017), afterwards these chips are supplied to the DCNNs that can run efficiently on the GPUs memory. As per Ning et al., (2020), training a network with a higher number of image chips normally leads to greater accuracy. Also, data augmentation has proved a successful mechanism to increase the variability from limited data and in turn, improves the performance of the networks.

Our main objectives in this chapter are, studying different strategies to make image chips for training and then cross-comparing them on two very popular segmentation networks, U-Net and SegNet.

Training dataset preparation

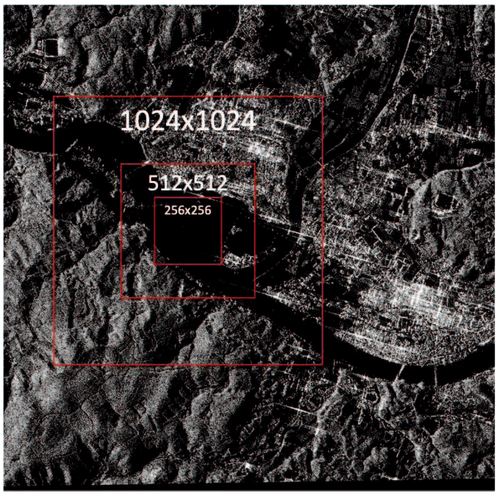


Figure 4.1 Importance of the chip size.

The selection of the size of the chips depends mainly on the two-constraints memory constraint and batch-size constraint. As we aim to detect land-water boundaries for better extraction of surface water, contextual information plays a greater role. Many a time the backscattering value of the pixels may be the same but the spatial context plays a role in deciding the pixel's class. This is the reason that smaller areas of shadow though have similar backscatter values as water, were remain out of the class of surface water, in final detection. In such a case, the context is the one thing that helps in differentiating them, so that is why it is preferred to have a larger patch size for more contextual information. However, we also need to take care of the GPU memory as bigger chips mean smaller batch size and larger training time. Given the above-mentioned reasons, this study selected the 512x512 size of the chip. For better understanding, we can look the Figure 4.1, in which it is very clear how chip size can affect contextual information. 256x256 size square is within the water areas only so it can not provide some significant information for land-water boundary extraction. On the other hand, 1024x1024 is having a lot of contextual information such as the contiguous area to the river has been flooded.

Another point to consider is that the flooded region is much smaller in comparison to the non-flooded region, this creates an imbalance in the dataset. To reduce this imbalance, the study has selected only those image chips which have at least 10% of total pixels belonging to a flooded area, called in this study as valid chips-

$$\frac{N_{FloodedPixels}}{N_{TotalPixels}} * 100 \geq 10$$

10% pixels from ALOS-2 image with 3m spatial resolution means approx. 8-hectare area, over which very few surface water bodies occur in that area. This helps to remove the noise from smaller lakes/ponds. Following methods have been used to extract image chips- 1) Sliding Window 2) Randomized sampling.

Sliding window

Under this method, a sliding window has been used to slide over the scene and create the chips with the different overlap of successive steps. Four sets of overlap have been used namely, no overlap, 30% of overlap, 50% overlap and 70% overlap as shown in Figure 4.2.

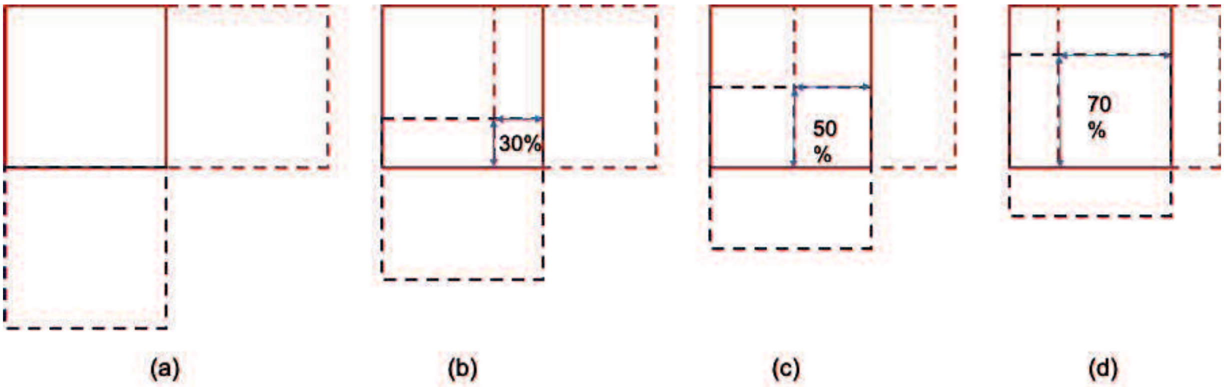


Figure 4.2 Overlap in successive steps in the case of a sliding window. (a), (b), (c) and (d) are showing the different scenario- 0%, 30%, 50% and 70% overlap for the chips

The different overlap will result in a different number of the total chips, and a greater number of chips may give a better result. So, in this study, we have decided to select an equal number of chips for each scenario. To decide that how many chips should be selected, we have used 70% overlap, as in this scenario the maximum number of chips will be created to cover the whole scene (Figure 4.3), the total valid image-chips have been found around 200 (this number depends upon the size of the image as well as the abundance of foreground pixels). For this reason, in all scenarios we have created 200 image chips, if the total number of valid chips is less than this limit such as in the no-overlap scenario, then we just duplicated the valid chips to satisfy the condition.

Randomized sampling/ random cropping

In this method, the 200 random patches have been selected using a 'sampling' method. The validity of these patches was calculated and recursively sampling has been done till the time the total number of valid image chips reach the mark of 200 or more (shown by Red square in Figure 4.3). In the end, only 200 of valid-chips has been saved for the training step (Yellow square in Figure 4.3).

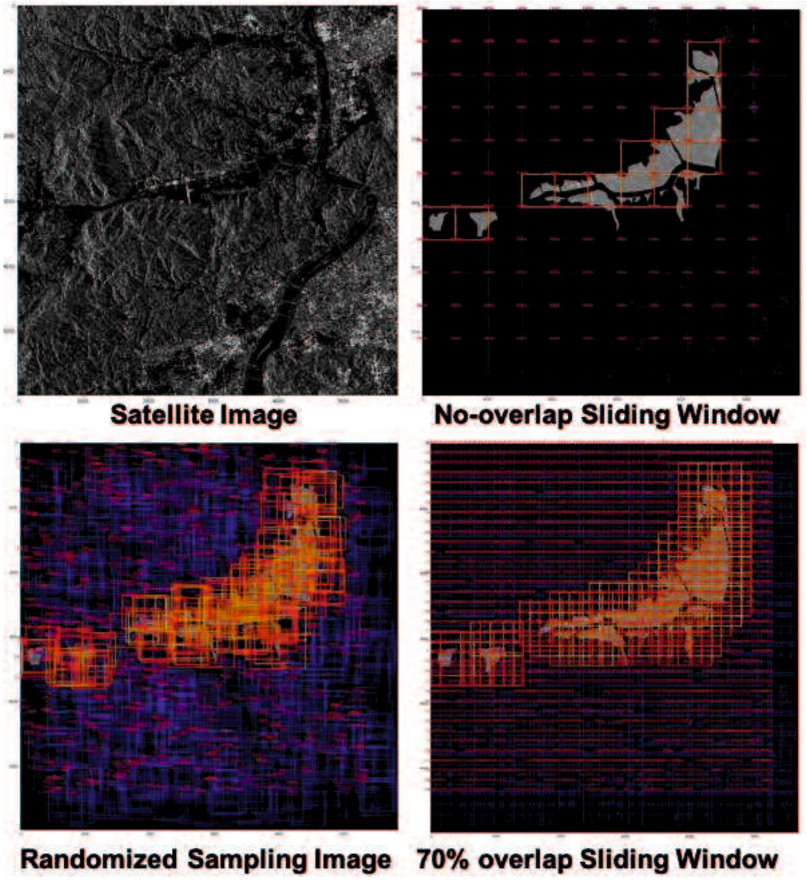


Figure 4.3 Training data creation. Image chips were created concerning the flood mask as can be seen in all the images above leaving satellite images aside. Blue colour bounding boxes (BBs) in the images are invalid due to not meeting the condition of 10% water pixels and Red colour BBs are valid while yellow BBs are the selected chips from total valid ones.

Results and Discussion

In the step of the creation of the training data, maximum time was spent in the randomized sampling method and in the case of the sliding window method, time is decreasing with decreasing overlap. So, the fastest method for image-chip creation was a sliding window with no overlap.

After training data preparation, U-Net and SegNet were trained on each training set i.e., five different training sets. Each training took between 15-20 mins for finishing 50 epochs. Here it needs to be focused on, that, hyper tuning of the network has not been done and rather than saving the best model, the model has been saved after 50 epochs. As the main aim was to do the cross-method comparison for different methods of training data preparation.

Training accuracy and binary cross-entropy loss during the training have also been plotted (Figure 4.4). As per the plot, the no-overlap scenario was converging fastest, while the randomized sampling one was more versatile and showing fewer sudden peaks with increasing epochs.

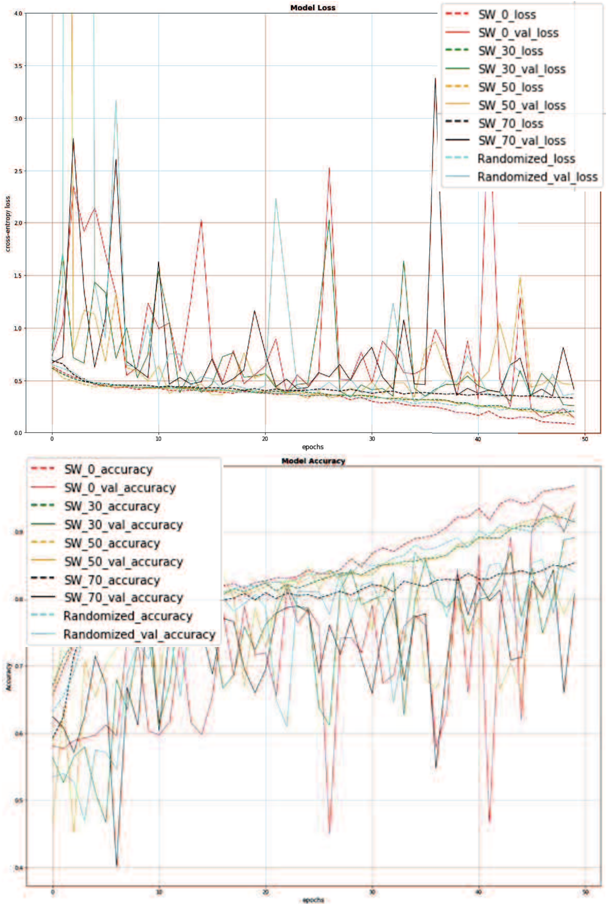


Figure 4.4 Cross-entropy loss and accuracy plot during the SegNet training.

The study has found that 50% overlap has the poorest scores throughout, one of the possible reasons can be that it has learned more on the negative samples, this can be seen in Figure 4.5, U-Net and SegNet cross method comparisons. In the figure, 50% overlap scenario is showing the highest accuracy which is just a measure of total correct pixels predicted and as non-flooded pixels are much higher in the scene, so predicting most of the pixels as non-flooded area leads to increase the accuracy. However, F1 Score and Jaccard score are the lowest showing the failure of predicting the right class for foreground i.e. flooded area.

Overall Randomized sampling shows best or approx. equal score around all the parameters. This seems logical too as all the methods in the study can also be seen as image-chips creation along with data augmentation in the manner of 'translation' (moving image to X and Y direction). Randomized sampling can be seen as translation of the image with an arbitrary factor within (0,512), whereas other methods and scenarios have a fixed-step translation.

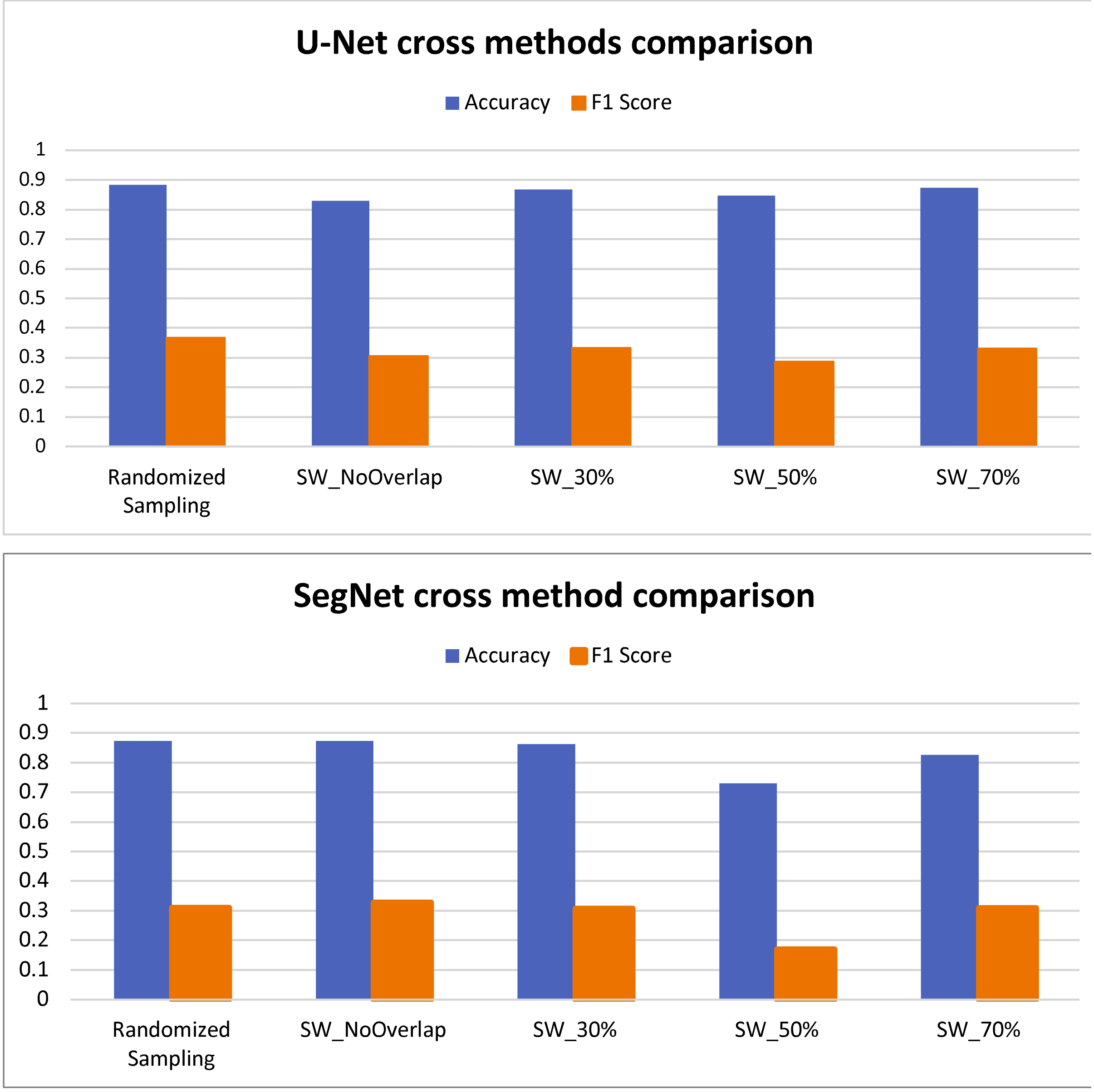


Figure 4.5 Charts of the performance of U-Net and SegNet with the chips created by all the previously mentioned methods

As per the time concerned for image-chips creation, it was the fastest in the case of the sliding window with no overlap. It was the case due to a smaller number of chips possible with this condition and then for making the desired number of chips, we have just duplicated the valid chips. This duplication is the reason that Figure 4.4 shows sharp convergence of SW_O (in Red colour) but with multiple bigger peaks at each interval. This behaviour may be better monitored when training for a greater number of epochs.

Although we have found that randomized sampling is best suited in this case, still it needs a greater number of test cases, for checking that either this observation holds in other cases or not. Moreover, we need to emphasize here that these strategies are not an alternative to more data. As in this case, data information remains the same and the only augmentation happens which has its limitations.

Chapter 5 **Near-Real-Time flood mapping using ALOS-2 imageries, with a case study on Kyushu Flood 2020**

Study area and data used

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 a levee breach in 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 the 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 5.1) 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.

This study has used ALOS-2 images during the flood. These images have been accessed through the Sentinel Asia portal, which is an international cooperation project with a focus on 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 images 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.

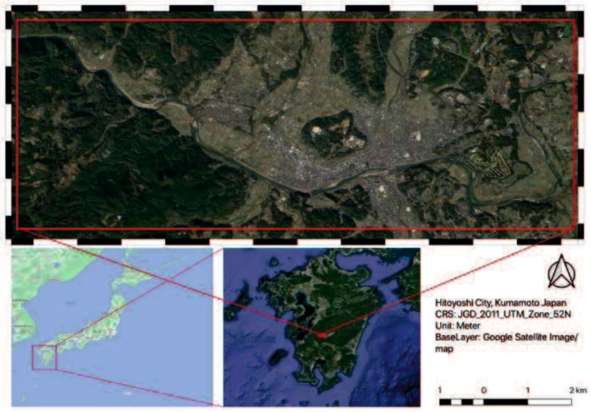


Figure 5.1 Study Area (Hitoyoshi city Kumamoto Japan)

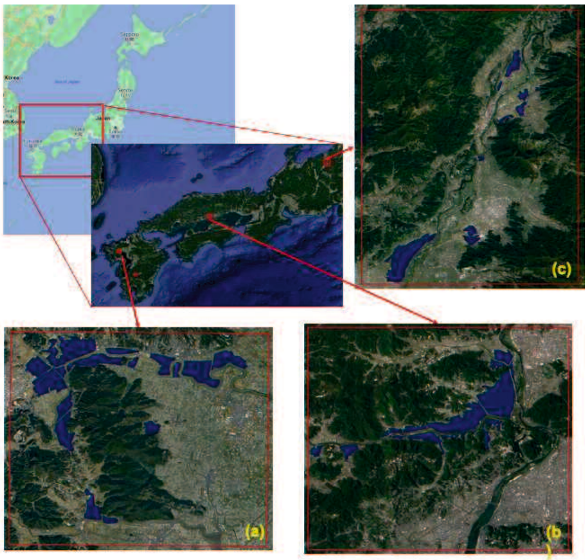


Figure 5.2 Areas used for training. (a) Saga city (b) Kurashiki city, Kumamoto (c) Nagano city. The blue colour in the area images is showing the flooded area

Training

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 the 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 5.2. Many different scenes which belong to the different areas have the advantage of capturing more spatial and temporal 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 with the help of a 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.

For testing, a 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.

Methodology

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 fulfil the condition that at least 10% of its pixels belong to the flooded area has been classified as valid tiles (Figure 5.3). 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.

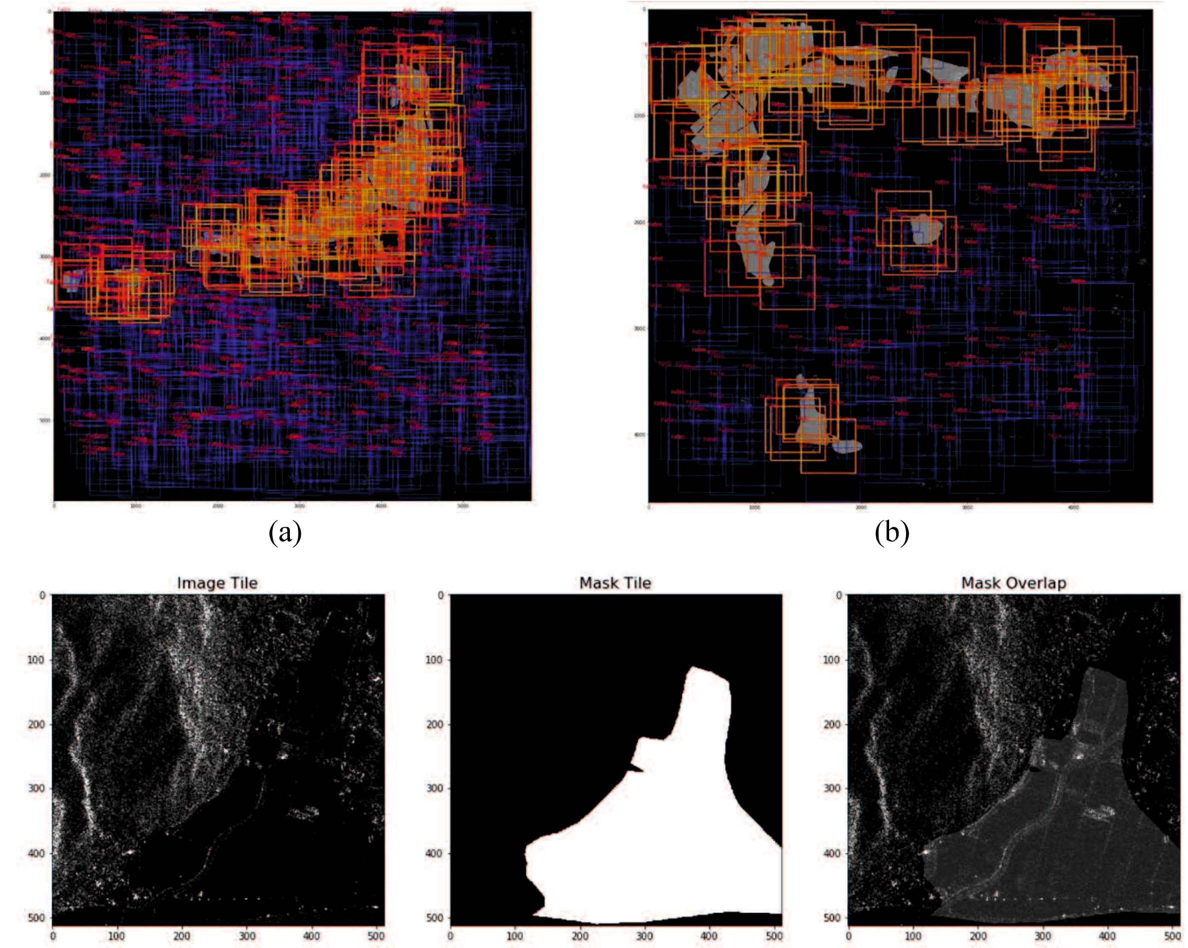


Figure 5.3 (a) and (b) are showing the training data creation using flood masks. Blue rectangles show not valid tiles(<10% flood area pixels), Red rectangles are valid and Yellow rectangles are the ones chosen out of red ones. (c) is showing the image and mask tiles, created through the above-shown procedure.

(c)

Network and Architecture

Due to the limited number of training tiles and binary classes (flood and non-flood class), we have selected U-Net (Figure 5.4) 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 that affects the larger area with diverse landscapes so larger RF will help in capturing local context details, which 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 of 0.0001. The adaptive learning rate was 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.

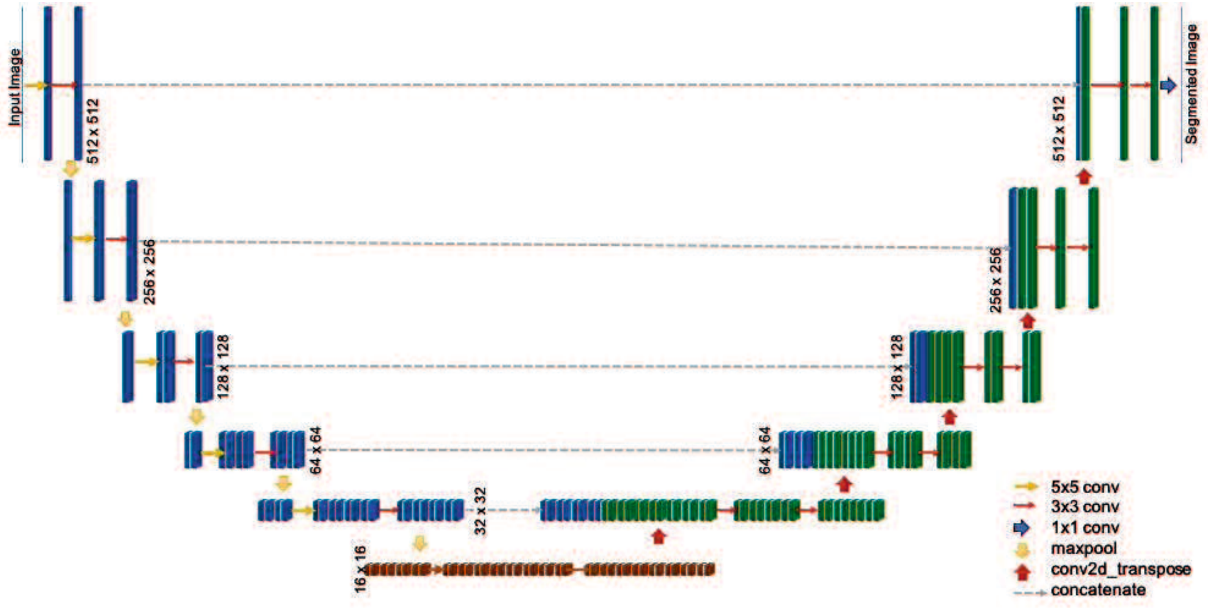


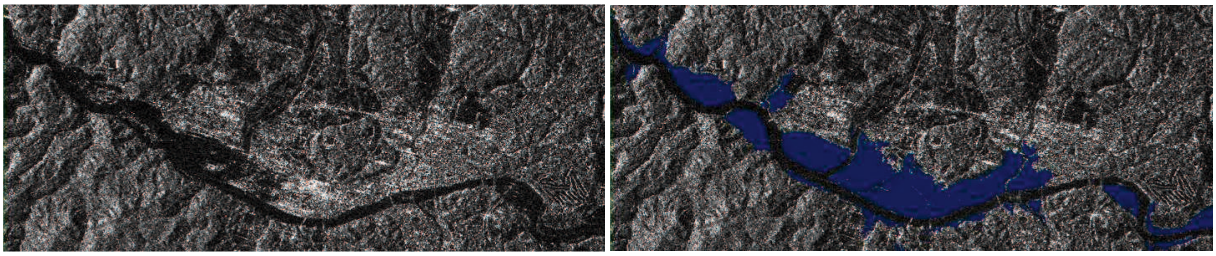
Figure 5.4 U-Net architecture

As the total number of training tiles are less for training a deep neural network, this study has used the 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 the validation split is happening at runtime. The network has been trained for 100 epochs with a batch size of 16 on two Nvidia Titan-V GPU. U-Net does not have many complex calculations and binary cross-entropy is a much simpler loss function to optimize, this is why total training has taken less than one hour.

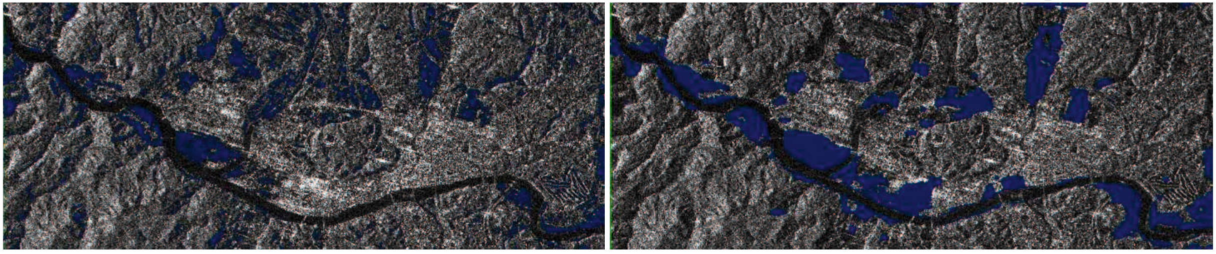
Test results

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, which 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.

Our results show that U-Net can extract the flooded area in a more homogenous way in comparison to the thresholding method as can be seen in Figure 5.5. The accuracy of the extracted flood areas, in comparison to the inundation map provided by GSI Japan comes to 89.57%, which is an improvement over the thresholding method (86%). Moreover, 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, a few areas 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. More detailed results are presented in the Figure 5.6

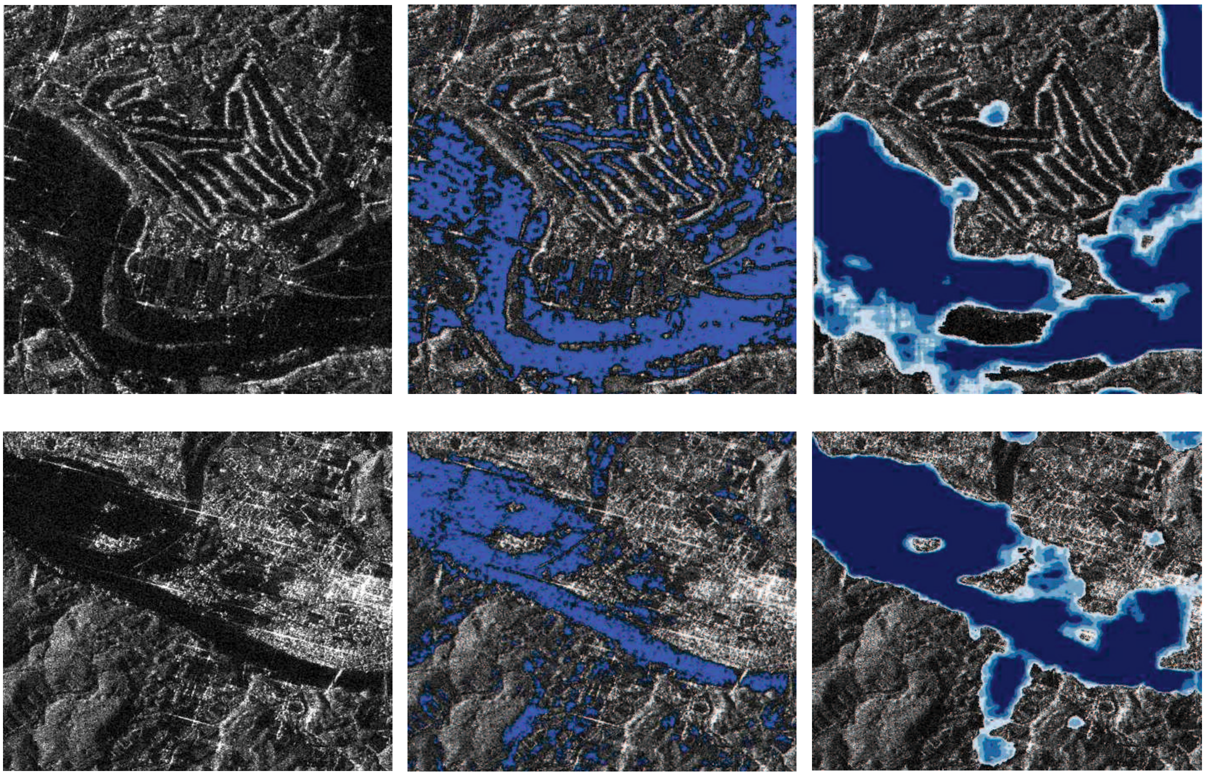


ALOS-2 Image of Hitoyoshi city
Inundation estimation map from GSI



Result by the thresholding method
Result by our method

Figure 5.5 ALOS-2 image of the study area along with the ground truth and results by thresholding as well as of our method (flooded-area is shown in Blue color)



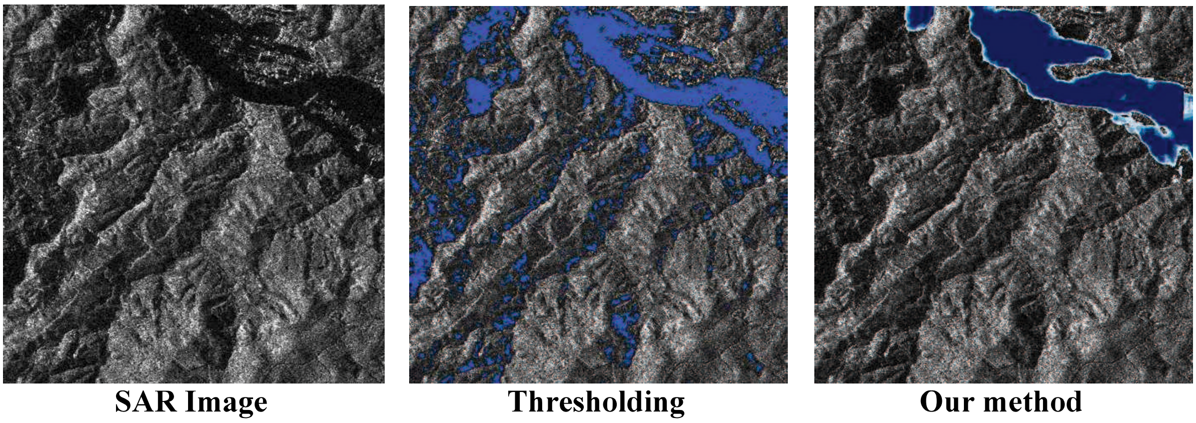


Figure 5.6 Zoomed areas of the results to highlight the advantages over thresholding methods. The shades of blue color in our method are representing the confidence of the network for classifying the area that belongs to flood (Dark blue means the confidence is in range of (0.9-1.0) and lightest blue is in range (0.4-0.6) and the remaining shades lie in between (0.6-0.9))

Transfer Learning from Sentinel-1 to ALOS-2 using flood area in the target training data

Preparing a large amount of training data for each kind of satellite system separately is not a practical approach. Building a larger training dataset for different SAR satellites have the following constraints-

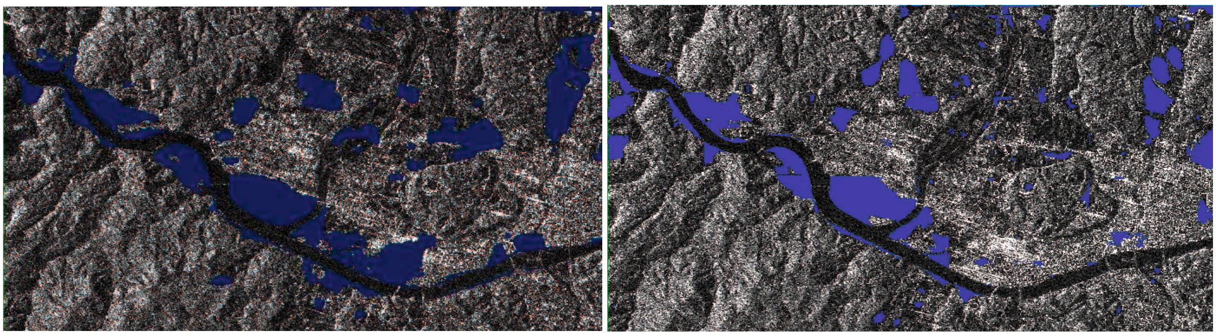
1) The interpretation of SAR images needs expert knowledge, so annotation is a time-consuming and resource-intensive procedure.

2) Most of the SAR satellites are commercial, so buying enough scenes just for training data preparation for each satellite will be very expensive and not very practical.

3) Many new satellites are getting launched and these do not have enough flood events in their archived data to prepare a large training dataset.

Due to these reasons, in the study, we have experimented with transfer learning especially using the weights of SAR trained models. Here we have used sentinel-1 trained models to work with the ALOS-2. As we have proved in the earlier chapters that feature from the encoder section of UNet plays a more important role than the decoder section so we have only retrained the encoder section during the transfer learning with a much smaller target dataset made using ALOS-2 images (number of chips used are the half of the chips used with the training from scratch). This retraining has been done for 100 epochs.

The qualitative results are shown in Figure 5.7 and quantitative results are presented in Table 5.1. The results show that by using transfer learning among the SAR images, even a smaller dataset (just 240 chips) is able to give comparative results with the network which were trained on the much bigger dataset (480 chips) and for a longer time.



Detected flood area with the network trained from scratch

Transfer learning from sentinel-1 to ALOS-2

Figure 5.7 Comparisons of the results of the different trained networks- from scratch and using transfer learning

Table 5.1 Results of the transfer learning from Sentinel-1 to ALOS-2

Method	Accuracy	F1 Score
Thresholding Method	84%	0.21
Transfer Learning	88.49%	0.38
Trained from scratch	89.57%	0.43

Overall discussions and conclusions

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. Figure 5.6 compares the result of our method with the thresholding method. The first row of Figure 5.6 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 detected some areas as flooded areas but still a lot of flooded urban areas has been left out. Another advantage of the method has also been seen in the case of shadows in hilly regions (last row of figure Figure 5.6) 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 the 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 regions 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 fixed geometry as we have used ascending and descending images without any priority and when applied on the test region it has given a very good result, this shows that network was able to grasp the complexity of SAR images.

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. 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 (detecting the undetectable such as differentiating the intentionally flooded paddy field versus flooded agriculture area due to disaster) and lack of credible training data. Additional data along with the more credible ground truth will surely improve the flood area extraction.

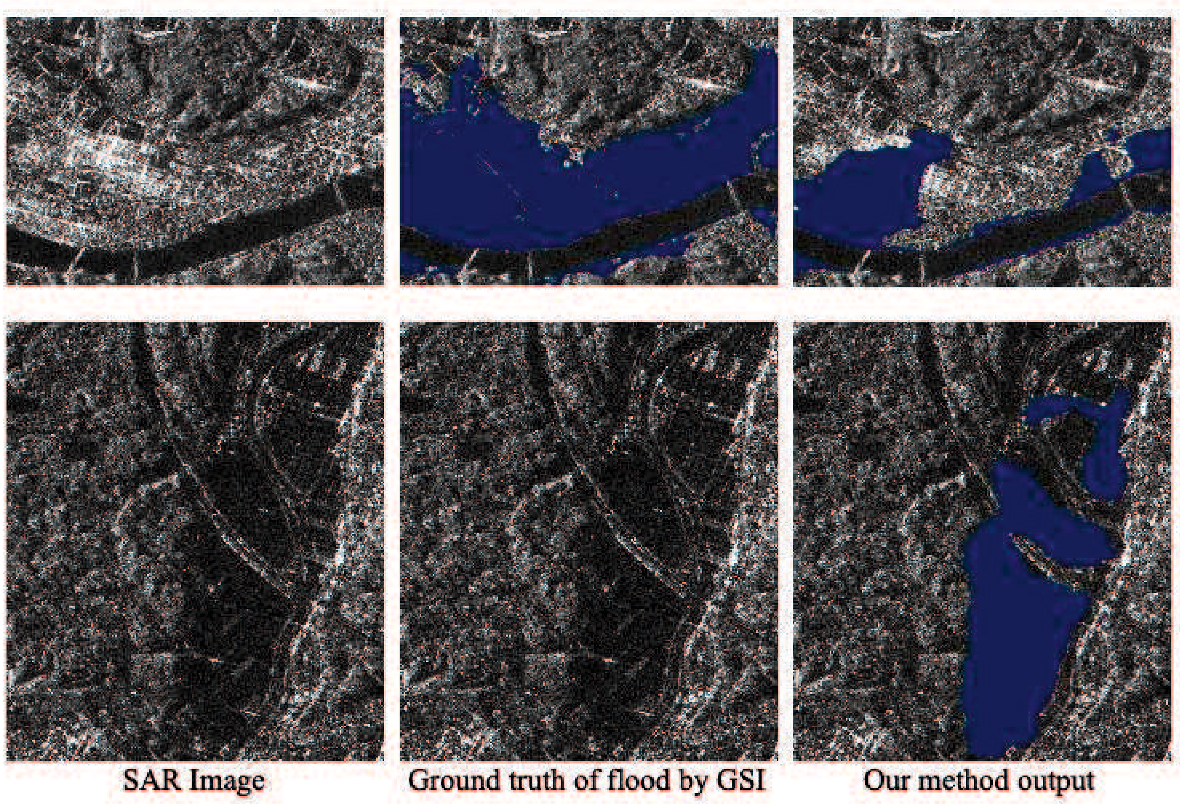


Figure 5.8 Limitations due to SAR special properties. The top row is showing the urban flood region and the bottom row shows the situation of either paddy field or unmapped flooded area in the ground truth.

The limitation of the training data has motivated us to apply transfer learning among different satellites. Our experiments of transfer learning from Sentinel-1 to ALOS-2 shows encouraging results. This can be further explored with a larger number of satellites and varied study areas to gauge the applicability of transfer learning for flood regions among different wavelength satellite systems.

Chapter 6 **Extraction of water bodies from ALOS-2 images by utilizing rough training set**

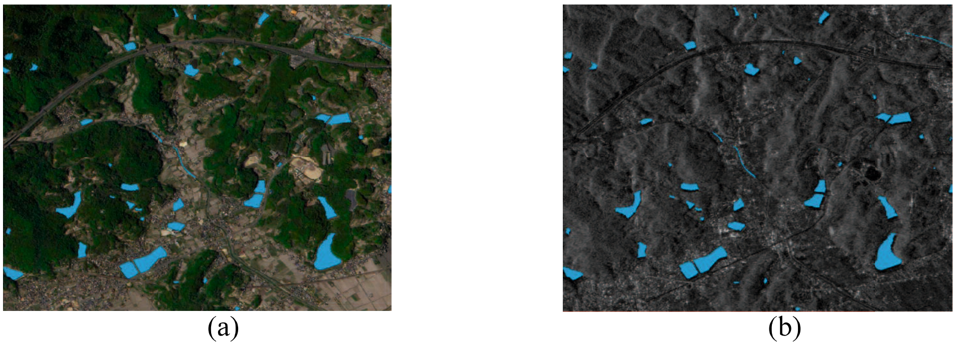
Introduction

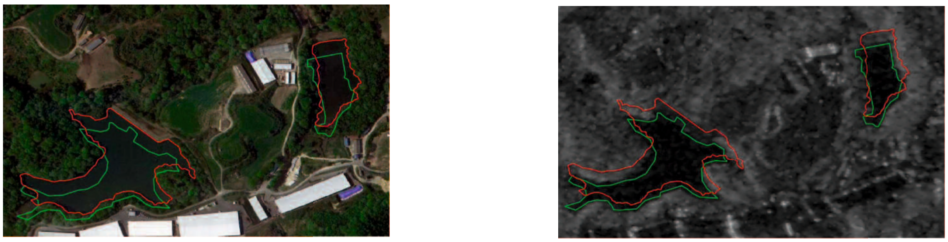
Automated extraction of water bodies in the geospatial domain has various applications such as urban planning, hazard mapping, change detection, etc. This has been specifically highlighted in Sustainable Development Goals 6 (SDG-6) i.e. 'Ensure availability and sustainable management of water and sanitation for all". In the sub-goal 6.6 it has mentioned as 'by 2020, protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers and lakes' (UNSKP, 2015). Implementation of these objectives is required frequent monitoring of these sites to avoid any encroachment and proper maintenance. However, regular manual surveys and monitoring are not practical as they will be more costly and human-resource intensive. Due to these reasons along with the improvement in the technology such as satellite imagery with higher special, spectral and temporal resolution and development of the machine learning techniques- an automated system of water bodies extraction has been propounded. Though most of them are mainly focusing on using optical images (W. Huang et al., 2018) such as low & medium spatial resolution satellite images (Jawak et al., 2015; Mishra & Prasad, 2015) and High & very high-resolution satellite images (Feng et al., 2019; L. Li et al., 2019). While they have produced some very good results but still the limitation of the optical image is always present that is affected by cloud and weather conditions. Due to this reason Huang, 2018 have recommended using SAR images along with optical images which can help not only in penetrating the cloud but also the vegetation. Chandran et al., (2018) are using sentinel-1 (SAR) satellite for extraction of the water bodies, however, they are training a separate network for finding the shadow to remove it from the final detected water bodies.

Study area and data preparation

As there was no training data available for ALOS-2 satellite images so training data for the water body has been prepared in-house. A single polarized ALOS-2 image (HH) of April 2019 has been chosen which has a resolution of 10 m with an observation width of 70 km. Two Zones which has been observed as having a greater number of water bodies has been selected with the dimension of 15 x 20 Km. For creating the training data faster, the study had used google earth high-resolution optical images with the QGIS digitization tool. The output of this step is saved as a vector file containing polygons of the shape of water bodies. Later on, this vector layer has been converted as a raster layer with binary value, 1 for water-bodies and 0 for others, converting it as a mask layer. Even though ALOS-2 PALSAR-2 L1.1 data has been orthorectified using ESA's SNAP (Sentinel Application Platform) toolbox, It was still not very well overlapped, one of the reasons was the usage of 30-Meter (low resolution) SRTM DEM (Shuttle Radar Topography Mission, Digital Evaluation Model) as well as certain limitation in horizontal positional accuracy of google earth (Goudarzi & Landry, 2017). Along with it, Mountainous terrain and SAR side-looking geometry were making things more complex for perfect overlapping. So this is why to create training data with better quality, manual shifting of the polygons have to be done, this is a complex task as small water bodies may not be visibly clear in the SAR image. Moreover, the boundary of the water bodies may also be blurred which in turn reduce the quality of training data. Due to those reasons, the study utilizes the same polygon which has been drawn on google earth images without shifting them as can been seen in Figure 6.1(a) and (b). The unevenness of the real ground truth can be seen in Figure 1(c) and (d) where the red colour polygon shows the boundary of the water body in google earth and the green colour polygon shows the boundary of the same water body in SAR image.

After this, iterative random clipping has been used to get 256x256 size of SAR image for the training set, totalling 20,000 image tiles. However, high unbalanced data have detrimental effects on the classification or segmentation algorithms (Buda et al., 2018). To reduce this problem, the study has used a threshold approach in the final phase of training data preparation i.e. selecting a tile as 'final' only when it has at least 5 % pixels under water bodies. For calculating this we have used a ratio of 'total number of pixels' with the 'total pixels that belong to water-bodies from the water-bodies mask layer.


(a) (b)



(c) (d)

Figure 6.1 Training data preparation. (a) and (c) are the clipped google earth images, blue color in them shows digitized water bodies area and (b) and (d) are the clipped ALOS-2 images after converting into backscatter (sigma zero - db) format using SNAP.

Methodology

This study has adopted U-Net architecture (Akeret, 2017; Akeret, 2018) with some modifications to handle the satellite images, this network is running on the two Titan-V GPUs with 12 GB of memory each. The language and supporting libraries used are Python and its supporting packages along with TensorFlow and Keras.

Network and architecture

U-Net is running with the 7 convolution layers in the encoder part and 6 convolutions in the decoder part. Due to the limited number of training samples there was a high possibility of overfitting and to avoid this, certain drop-out layers (regularization) has been inserted with probability (p) = (0.8, 0.5) for the different hidden layers of the network. Leaky ReLU activation function has been used in the hidden layers and sigmoid in the output layer due to binary classes.

Training

Total 20,000 images of the size 256x256 have been used for the training of the network with the validation split of 30 per cent. This means at one time 14,000 images have been used for training and the remaining 6000 has been used for validation purposes dynamically. Just a little over 100 epoch model shows convergence with the batch size 25 (Figure 6.2).

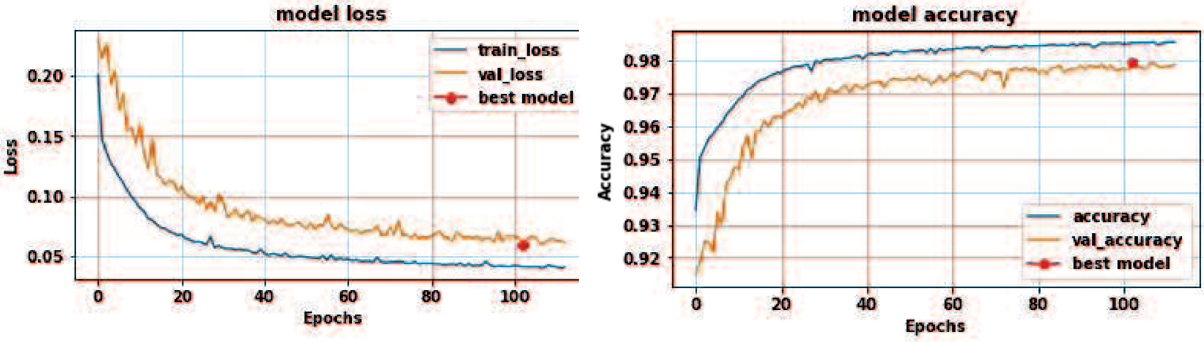


Figure 6.2 Model loss and accuracy curve of training and the location of the best model

Testing

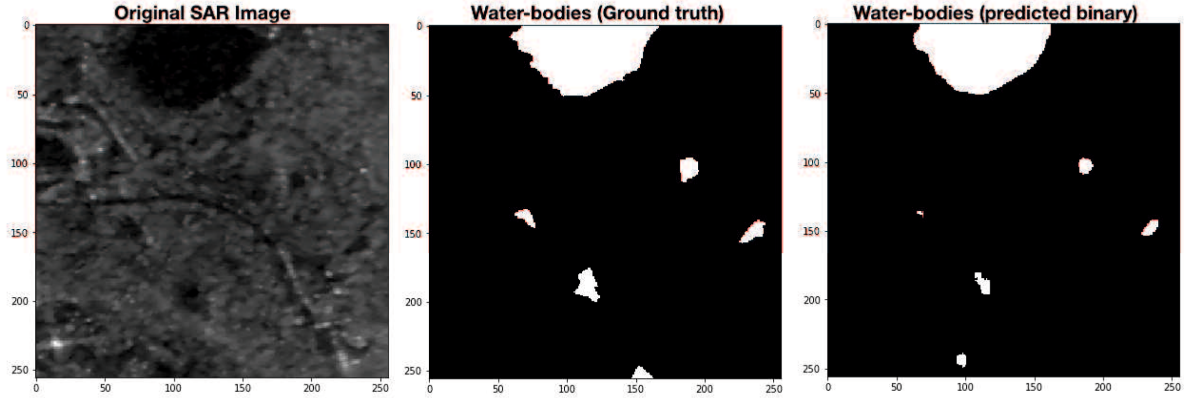
For testing and evaluating the model 1500 images of the same size (256x256) has been used which have not been seen by the model before. This set has been randomly divided into three parts to test the model. The predicted probability image has been converted to binary prediction using a threshold of greater than 50%. For evaluating the model, Intersection over Union (IoU) matrix has been chosen which can be defined as (Eq1)-

$$\boxed{IoU = \frac{(Groundtruth \cap Predicted)}{(Groundtruth \cup Predicted)}} \qquad \text{(Eq1)}$$

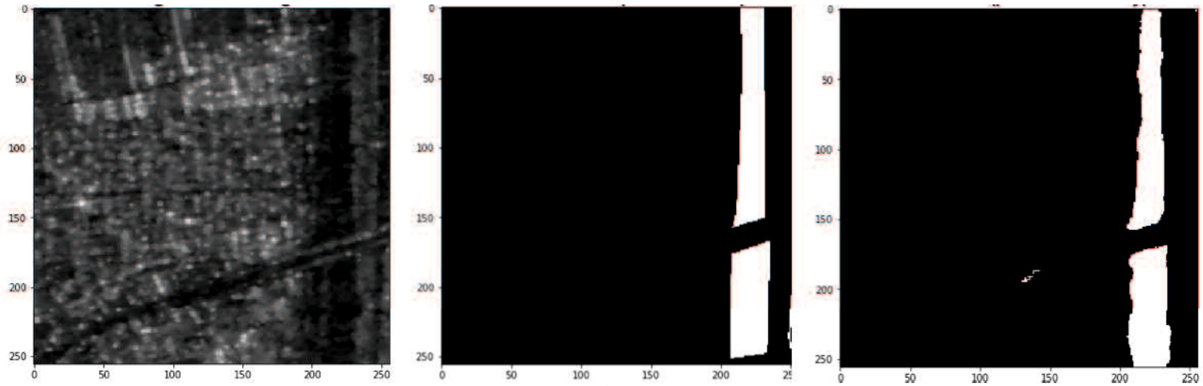
IoU for the parts of the test set has been falling in the interval of 0.69 to 0.76 with the loss hovering in the interval of 0.22 to 0.28. This IoU seems at the lower side, however, we need to realize that IoU is best measured when the bounding box is pretty good. In this study, the ground truth has been rough so even when the model will predict the exact object still IoU will not be close to 1.0.

Results

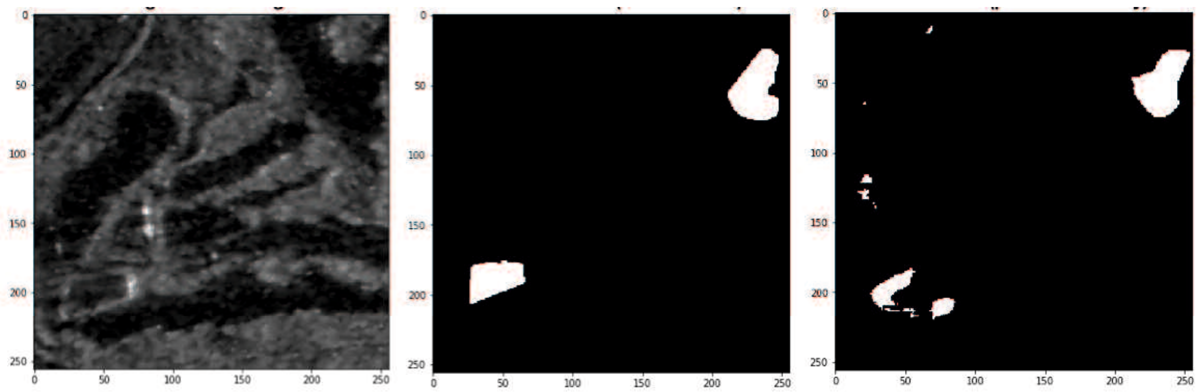
Some of the predicted output of the images by the trained model has been shown in Figure 6.3. If we see Figure 6.3(a) where all the water bodies even though they are small has been identified by the model though the extent of them seems a little bit reduced. In Figure 6.3(b) the rivulet with the over-bridge has been captured properly. Figure 6.3(c) shows a challenging task as small water bodies were in the golf areas which itself looks very similar to the water areas in the SAR image. However, the model has been able to distinguish between water bodies and golf fields beautifully. Even the water body at the bottom left corner has not been detected clearly due to the presence of power lines which create more backscattering. Though the still model has wrongly predicted few areas as water. On the other hand, in Figure 6.3(d) model has predicted the completely wrong result, the baseball field is very flat and it shows a very strong probability to classify it as a water area. While there are other sports fields and with a similar flat surface, the model does not recognize them as water bodies. In the case of baseball field specifically, it shows stronger bias towards the water, it may be a case of shape which model learn due to the presence of many lakes in the training set. Still, this point needed to be explored further in the study.



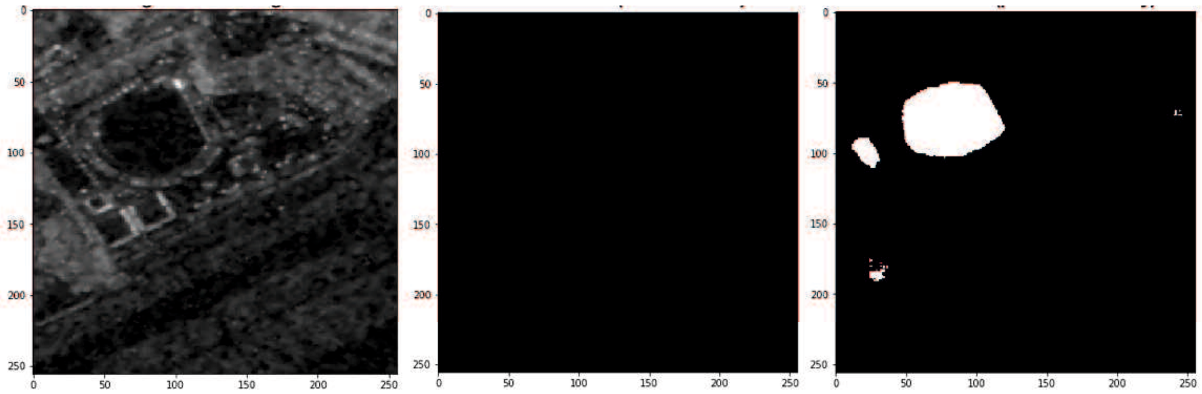
(a) Small water bodies



(b) Rivulet



(c) Water bodies in between golf course

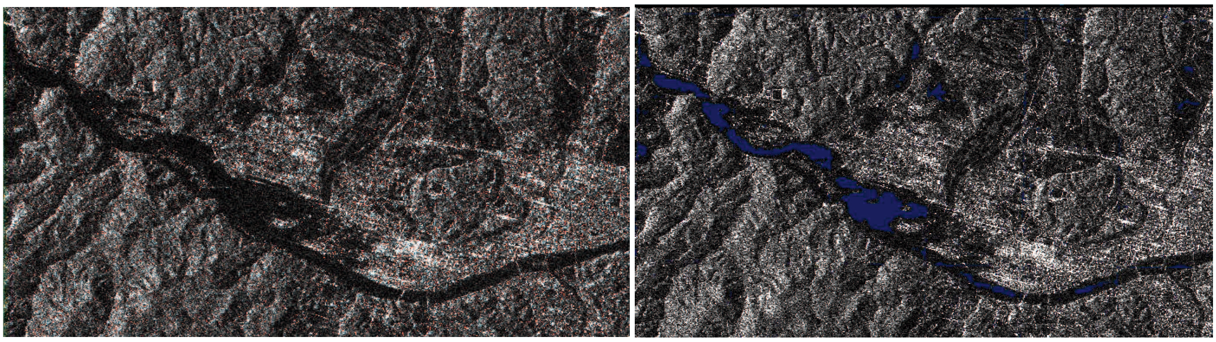


(d) Baseball stadium

Figure 6.3 Left side image shows SAR image, the middle image shows the ground truth and the right side is showing predicted water bodies in binary form (after thresholding)

Transfer Learning from Sentinel-1 to ALOS-2 using water bodies (permanent water) area in the target training data

As mentioned in chapter 5, in the cases of newly launched satellite, there is possibility that we do not have archive data which includes any flood event. In that scenario we cannot do transfer learning which were utilized in the chapter 5. For handling this kind of situation author have tried to utilize the water bodies data (or permanent water data) with the transfer learning. However, model performed very poorly as shown in the Figure 6.4. This suggest that only water bodies data for flood detection is not a suitable measure to map the flooded region. This result can be understood in the light of the diversity of the backscattering value included in the flooded region in comparison to the permanent water bodies. Flooded region includes much diverse pixels as has been shown in the previous chapter too. Another thing to be noted here that this result is agreement with the Bonafilia et al., (2020) where they have found that network trained with only permanent water bodies, performed worst for flood detection.



Detected flood area with the network trained from scratch

Transfer learning from sentinel-1 to ALOS-2

Figure 6.4 Comparisons of the results of the differently trained networks- from scratch and using transfer learning

Conclusion

While SAR data is already becoming ubiquitous, and many new SAR satellites are planned to be launched. Such kinds of models can be very helpful to process SAR images automatically and with less amount of human resources and expenditure. The above discussion shows that a simple model with even a rough training set can provide a good accuracy over SAR images. Therefore, the availability of a better annotated and larger dataset can help in better utilization of SAR images for many areas including disasters such as flood monitoring, urban development and water ecosystem management etc.

Chapter 7 **Dissertation findings, conclusions, and future recommendations**

Dissertation findings and conclusions

In this study, we have presented that if available more polarization will improved detection of flooded areas as seen in the case of Sentinel-1 using both polarizations (VV and VH). Also, adding a third band as a ratio of two polarizations has been shown to refine the results further in few scenarios. We also proved that there is no way to avoid hand labelling completely, but it can be used in combination with weak labels for developing a more generalized model. This way we can take advantage of both situations: more samples from weak labelling for better generalization and accurate samples from hand labelling for fine-tuning during transfer learning. Also, transfer learning showed that the same models can be enhanced with access to more training data in the future to further improve the model. During the implementation, we have found that in the specific case of Flood mapping from SAR image, it is the general features that are playing a greater role than many complex features, so much complex network may not be beneficial for this problem. In this way, existing datasets can be used for NRT flood mapping. As this technique uses only a single image, namely, only an image during flooding, it is much easier to implement a generalized model in any affected area without having the constraint of searching archived data and appropriate reference images. We have demonstrated a notable improvement over thresholding techniques with the off-the-shelf model, which can process a whole satellite image in less than 1 minute with a very low omission error. Thus, our models can be implemented as a prompt emergency response and information disburser for first responder organizations.

Another notable finding was the successful implementation of transfer learning among the different SAR satellites. The network trained on the Sentinel-1 data has been used with ALOS-2 data by freezing its decoder section and only retraining the encoder part of the network. It shows that even with a much smaller training set and training time we can achieve approximately a similar level of flooded area detection as we can get by training the network from scratch with double the dataset and a greater number of epochs. However, using only permanent water bodies data with ALOS-2 for training is not suitable to detect the flooded region and this conclusion has followed the similar interpretation given by Bonafilia et al, 2020 in their study on the Sentinel-1. This means before applying to any other satellite data we need to have some training data for flooded regions from that satellite. Though the requirement of the amount of training data is greatly reduced because of the transfer learning.

Future recommendations

As we have presented in the study, NRT flood mapping is the need of the hour due to frequent flood-related disasters and short response time. However, for calculating the damage by the flood some other parameters are also needed such as how long and how deep an area has been flooded. Also, to make the system more user friendly the constraint on its implementation need to be reduced. These points are further elaborated below:

Accuracy focused

- Easily accessible ancillary data, such as height above the nearest drainage (HAND), can also be added for more refined detection. This will not only help in reducing the false positives but can also help in creating a flood depth map. However, it will perform better with more accurate DEM.
- The training dataset should be further enhanced. Some recommendations to improve the datasets are:
    - Inclusion of specialized classes of the flooded areas as the type of floods, such as open flood, flooded vegetation, and urban flood. As each kind of flooded region has its unique properties, having separate classes will assist the detection of flooded regions with higher accuracy.
    - Balanced and error-free input data. This means enough samples must be provided for different kinds of flooded regions spatially and geographically. This will reduce the chances of anomaly type of regions that were giving erroneous detection and reduces the accuracy of the overall model.
    - Ensuring that the test set is error-free. The test set should be quality controlled so that efficacy of the different deep learning models can be better measured and compared.

Implementation focused

- In future, the same system can be tested on Google Collaboratory and other cloud platforms to make it more versatile for use without the constraint of in-house GPUs.
- Integration with the platform such as Google Earth Engine (GEE). As GEE provides freely accessible data, so it will remove the time loss due to downloading of data. This will make NRT further efficient.

References

Assad, S. E. A. A. el. (2019). *Flood Detection with a Deep Learning Approach Using Optical and SAR Satellite Data*. https://www.ipi.uni-hannover.de/fileadmin/ipi/abschlussarbeiten/master/2019/2019_Thesis_El_Emrani.pdf

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495.

Bahl, G., Daniel, L., Moretti, M., & Lafarge, F. (2019). Low-power neural networks for semantic segmentation of satellite images. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, 2469–2476. https://doi.org/10.1109/ICCVW.2019.00302

Bioresita, F., Puissant, A., Stumpf, A., & Malet, J. P. (2019). Fusion of Sentinel-1 and Sentinel-2 image time series for permanent and temporary surface water mapping. *International Journal of Remote Sensing*, *40*(23), 9026–9049. https://doi.org/10.1080/01431161.2019.1624869

Bonafilia, D., Tellman, B., Anderson, T., & Issenberg, E. (2020). Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2020-June*, 835–845. https://doi.org/10.1109/CVPRW50498.2020.00113

Bourgeau-Chavez, L. L., Kasischke, E. S., Brunzell, S. M., Mudd, J. P., Smith, K. B., & Frick, A. L. (2001). Analysis of space-borne SAR data for wetland mapping in Virginia riparian ecosystems. *International Journal of Remote Sensing*, *22*(18), 3665–3687. https://doi.org/10.1080/01431160010029174

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Caballero, I., Ruiz, J., & Navarro, G. (2019). Sentinel-2 satellites provide near-real time evaluation of catastrophic floods in the West Mediterranean. *Water (Switzerland)*, *11*(12). https://doi.org/10.3390/w11122499

Chandran, B., Remote, N., & Centre, S. (2018). *Water body extraction from SAR images. December*, 0–12.

Chen, K.-S. (2016). *Principles of Synthetic Aperture Radar Imaging*. CRC Press.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Clement, M. A., Kilsby, C. G., & Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, *11*(2), 152–168. https://doi.org/10.1111/jfr3.12303

Cloke, H., di Baldassarre, G., Landeg, O., Pappenberger, F., & Ramos, M.-H. (2017). Hydrological risk: floods. In *Science for disaster risk management 2017: knowing better and losing less* (pp. 198–238). https://hal.archives-ouvertes.fr/hal-02165196

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., & Raska, R. (2018). DeepGlobe 2018: A challenge to parse the earth through satellite images.

IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, *2018-June*, 172–181. https://doi.org/10.1109/CVPRW.2018.00031

Ding, J., Chen, B., Liu, H., & Huang, M. (2016). Convolutional Neural Network with Data Augmentation for SAR Target Recognition. *IEEE Geoscience and Remote Sensing Letters*, *13*(3), 364–368. https://doi.org/10.1109/LGRS.2015.2513754

Feng, W., Sui, H., Huang, W., Xu, C., & An, K. (2019). Water Body Extraction from Very High-Resolution Remote Sensing Imagery Using Deep U-Net and a Superpixel-Based Conditional Random Field Model. *IEEE Geoscience and Remote Sensing Letters*, *16*(4), 618–622. https://doi.org/10.1109/LGRS.2018.2879492

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23–35. https://doi.org/10.1016/j.rse.2013.08.029

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 3141–3149. https://doi.org/10.1109/CVPR.2019.00326

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. In *MIT Press*. http://www.deeplearningbook.org

Goudarzi, M. A., & Landry, R. J. (2017). Assessing horizontal positional accuracy of Google Earth imagery in the city of Montreal, Canada. In *Geodesy and Cartography* (Vol. 43, Issue 2, pp. 56–65). https://doi.org/10.3846/20296991.2017.1330767

Han, S., Fafard, A., Kerekes, J., Gartley, M., Ientilucci, E., Savakis, A., Law, C., Parhan, J., Turek, M., Fieldhouse, K., & Rovito, T. (2017). Efficient generation of image chips for training deep learning algorithms. *Automatic Target Recognition XXVII*, *10202*, 1020203. https://doi.org/10.1117/12.2261702

Henry, J. -B., Chastanet, P., Fellah, K., & Desnos, Y. -L. (2006). Envisat multi-polarized ASAR data for flood mapping. *International Journal of Remote Sensing*, *27*(10), 1921–1929. https://doi.org/10.1080/01431160500486724

Herndon, K., Muench, R., Cherrington, E., & Griffin, R. (2020). An assessment of surface water detection methods for water resource management in the Nigerien Sahel. *Sensors (Switzerland)*, *20*(2), 1–14. https://doi.org/10.3390/s20020431

HESS, L. L., MELACK, J. M., & SIMONETT, D. S. (1990). Radar detection of flooding beneath the forest canopy: a review. *International Journal of Remote Sensing*, *11*(7), 1313–1325. https://doi.org/10.1080/01431169008955095

Hostache, R., Matgen, P., & Wagner, W. (2012). Change detection approaches for flood extent mapping: How to select the most adequate reference image from online archives? *International Journal of Applied Earth Observation and Geoinformation*, *19*(1), 205–213. https://doi.org/10.1016/j.jag.2012.05.003

Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Reviews of Geophysics*, *56*(2), 333–360. https://doi.org/10.1029/2018RG000598

Huang, W., DeVries, B., Huang, C., Lang, M. W., Jones, J. W., Creed, I. F., & Carroll, M. L. (2018). Automated extraction of surface water extent from Sentinel-1 data. *Remote Sensing*, *10*(5), 1–18. https://doi.org/10.3390/rs10050797

Huang, Z., Pan, Z., & Lei, B. (2020). What, Where, and How to Transfer in SAR Target Recognition Based on Deep CNNs. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(4), 2324–2336. https://doi.org/10.1109/TGRS.2019.2947634

Jadon, S. (2020). A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*. https://doi.org/10.1109/CIBCB48159.2020.9277638

Jawak, S. D., Kulkarni, K., & Luis, A. J. (2015). A Review on Extraction of Lakes from Remotely Sensed Optical Satellite Data with a Special Focus on Cryospheric Lakes. *Advances in Remote Sensing*, *04*(03), 196–213. https://doi.org/10.4236/ars.2015.43016

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). https://doi.org/10.3390/rs13122334

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognition*, *19*(1), 41–47. https://doi.org/10.1016/0031-3203(86)90030-0

Kundzewicz, Z. (2008). Flood risk and vulnerability in the changing climate. *Annals of Warsaw University of Life Sciences - SGGW*, *39*(1), 21–31. https://doi.org/10.2478/v10060-008-0002-9

Landuyt, L., van Wesemael, A., Schumann, G. J. P., Hostache, R., Verhoest, N. E. C., & van Coillie, F. M. B. (2019). Flood Mapping Based on Synthetic Aperture Radar: An Assessment of Established Approaches. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), 722–739. https://doi.org/10.1109/TGRS.2018.2860054

Lang, M. W., Townsend, P. A., & Kasischke, E. S. (2008). Influence of incidence angle on detecting flooded forests using C-HH synthetic aperture radar data. *Remote Sensing of Environment*, *112*(10), 3898–3907. https://doi.org/10.1016/j.rse.2008.06.013

Li, L., Yan, Z., Shen, Q., Cheng, G., Gao, L., & Zhang, B. (2019). Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks. *Remote Sensing*, *11*(10). https://doi.org/10.3390/rs11101162

Li, T., Comer, M., & Zerubia, J. (2019). Feature Extraction and Tracking of CNN Segmentations for Improved Road Detection from Satellite Imagery. *Proceedings - International Conference on Image Processing, ICIP*, *2019-Septe*, 2641–2645. https://doi.org/10.1109/ICIP.2019.8803355

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L. (2019). Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing*, *11*(4). https://doi.org/10.3390/rs11040403

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). https://doi.org/10.3390/rs11192231

Manjusree, P., Prasanna Kumar, L., Bhatt, C. M., Rao, G. S., & Bhanumurthy, V. (2012). Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. *International Journal of Disaster Risk Science*, *3*(2), 113–122. https://doi.org/10.1007/s13753-012-0011-5

Martinis, S., Twele, A., & Voigt, S. (2009). Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data.

Natural Hazards and Earth System Science, *9*(2), 303–314. https://doi.org/10.5194/nhess-9-303-2009

Mason, D. C., Speck, R., Devereux, B., Schumann, G. J. P., Neal, J. C., & Bates, P. D. (2010). Flood detection in Urban areas using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(2), 882–894. https://doi.org/10.1109/TGRS.2009.2029236

Mishra, K., & Prasad, P. R. C. (2015). Automatic Extraction of Water Bodies from Landsat Imagery Using Perceptron Model. *Journal of Computational Environmental Sciences*, *2015*, 1–9. https://doi.org/10.1155/2015/903465

Ning, H., Li, Z., Wang, C., & Yang, L. (2020). Choosing an appropriate training set size when using existing data to train neural networks for land cover segmentation. *Annals of GIS*, *26*(4), 329–342. https://doi.org/10.1080/19475683.2020.1803402

Nobuyuki Otsu. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern*, *9*(1), 62–66.

Ormsby, J. P., Blanchard, B. J., & Blanchard, A. J. (1985). Detection of Lowland Flooding Using Active Microwave Systems. *Photogrammetric Engineering and Remote Sensing*, *51*(3), 317–328.

RICHARDS, J. A., WOODGATE, P. W., & SKIDMORE, A. K. (1987). An explanation of enhanced radar backscattering from flooded forests. *International Journal of Remote Sensing*, *8*(7), 1093–1100. https://doi.org/10.1080/01431168708954756

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Rottensteiner, F., Sohn, G., Gerke, M., & Wegner, J. D. (2014). *ISPRS Semantic Labeling Contest*. *September*, 4. https://doi.org/10.13140/2.1.3570.9445

Schumann, G. J. P., Brakenridge, G. R., Kettner, A. J., Kashif, R., & Niebuhr, E. (2018). Assisting flood disaster response with earth observation data and products: A critical assessment. *Remote Sensing*, *10*(8), 1–19. https://doi.org/10.3390/rs10081230

Schumann, G. J. P., & Moller, D. K. (2015). Microwave remote sensing of flood inundation. *Physics and Chemistry of the Earth*, *83–84*, 84–95. https://doi.org/10.1016/j.pce.2015.05.002

Sekou, T. B., Hidane, M., Olivier, J., & Cardot, H. (2019). *From Patch to Image Segmentation using Fully Convolutional Networks -- Application to Retinal Images*. http://arxiv.org/abs/1904.03892

Sirirattanapol, C., Tamkuan, N., Nagai, M., & Ito, M. (2020). Apply Deep Learning Techniques on Classification of Single-Band SAR Satellite Images. In S. Monprapussorn, Z. Lin, A. Sitthi, & P. Wetchayont (Eds.), *Geoinformatics for Sustainable Development in Asian Cities* (pp. 1–11). Springer International Publishing.

Tiwari, V., Kumar, V., Matin, M. A., Thapa, A., Ellenburg, W. L., Gupta, N., & Thapa, S. (2020). Flood inundation mapping-Kerala 2018; Harnessing the power of SAR, automatic threshold detection method and Google Earth Engine. *PLoS ONE*, *15*(8 August), 1–17. https://doi.org/10.1371/journal.pone.0237324

Townsend, P. A. (2002). Relationships between forest structure and the detection of flood inundation in forested wetlands using C-band SAR. *International Journal of Remote Sensing*, *23*(3), 443–460. https://doi.org/10.1080/01431160010014738

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://doi.org/10.1080/01431161.2016.1192304

UNEP. (2015). *Goal 11: Sustainable cities and communities*. United Nations Development Programme. https://www.unep.org/explore-topics/sustainable-development-goals/why-do-sustainable-development-goals-matter/goal-11

Etten, A. van, Lindenbaum, D., & Bacastow, T. (2018). *SpaceNet: A Remote Sensing Dataset and Challenge Series*. http://arxiv.org/abs/1807.01232

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*, *8828*(AUGUST 2019), 1–1. https://doi.org/10.1109/tpami.2020.2983686

Wang, Y., Hess, L. L., Filoso, S., & Melack, J. M. (1995). Understanding the radar backscattering from flooded and nonflooded Amazonian forests: Results from canopy backscatter modeling. *Remote Sensing of Environment*, *54*(3), 324–332. https://doi.org/10.1016/0034-4257(95)00140-9

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Yang, H., Wang, Z., Zhao, H., & Guo, Y. (2011). Water body extraction methods study based on RS and GIS. *Procedia Environmental Sciences*, *10*(PART C), 2619–2624. https://doi.org/10.1016/j.proenv.2011.09.407

Yang, X., Qin, Q., Grussenmeyer, P., & Koehl, M. (2018). Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery. *Remote Sensing of Environment*, *219*(September 2017), 259–270. https://doi.org/10.1016/j.rse.2018.09.016

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, *4*(January), 3320–3328.

Zhang, C., Wei, S., Ji, S., & Lu, M. (2019). Detecting large-scale urban land cover changes from very high resolution remote sensing images using CNN-based classification. *ISPRS International Journal of Geo-Information*, *8*(4). https://doi.org/10.3390/ijgi8040189

Zhang, Q., Kong, Q., Zhang, C., You, S., Wei, H., Sun, R., & Li, L. (2019). A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network. *European Journal of Remote Sensing*, *52*(1), 572–582. https://doi.org/10.1080/22797254.2019.1694447