A DEEP NEURAL NETWORK SLOPE REDUCTION MODEL ON SENTINEL-1 IMAGES FOR WATER MASK EXTRACTION

Marios Mpakratsas[1], Anastasia Moumtzidou[1], Ilias Gialampoukidis[1], Stefanos Vrochidis[1], and Ioannis Kompatsiaris[1]

[1]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
{*mbakratsas,moumtzid,heliasgj,stefanos,ikom*}*@iti.gr*

September 6, 2019

KEY WORDS: Water bodies, Synthetic Aperture Radar, Surface Water Mapping, Deep Learning

ABSTRACT The monitoring of water bodies from space is one of the main challenges for flood risk assessment, food security and climate change monitoring applications. A common issue in Synthetic Aperture Radar (SAR) images water-bodies masks generation problem is falsely identifying as water, areas that are characterised by a complex morphology. Existing works on water bodies estimation either remove the steep slope areas or not, on a case-by-case manner. The deep learning era allows for automatic adaptation to Sentinel 1 images where the false removal of water, in cases of rivers and lake shores, can be avoided. This paper proposes a DNN model that generates water-bodies masks for Sentinel-1 satellite data by fusing the SAR backscatter coefficients and the Digital Elevation Model (DEM) data. Hence, disregarding the steep sloped areas can eradicate these false positives, in the cost of removing at the same time some actually inundated areas. Our proposed method uses as input the different polarisation bands of Sentinel-1 images and combines them with the corresponding DEM information to estimate the output class, filtering out the highly sloped areas, leading to improved delineation accuracy. The model involves a training stage using machine learning algorithms including neural networks. The algorithm decides upon the effectiveness of slope removal for the specific pixel or block of pixels. In order to validate these findings and to quantify the impact of high slope removal, we run experiments on three major Italian lakes and their surrounding territories. Specifically, when filtering out these water areas, the F-Score, capturing the correct identification of water areas, increased from 81.1% to 94.1% for Garda lake, from 75.33% to 92.77% for Maggiore lake and from 75.54% to 88.91% for Trasimeno lake.

1 INTRODUCTION

The mapping of water bodies from space by considering Earth Observation (EO) satellite images has paved the way to detecting flooded areas, to identifying aquatic pollution that may be caused from oil spills, plastics, or debris from natural disasters, to monitoring damaged areas and eventually to taking effective actions. Classifying accurately the pixels of a satellite image as an inundated area or not allows for creating maps which are then used by civil protection agencies and first responders.

By considering the high applicability of the Artificial Intelligence (AI) and their promising results in several fields, among them remote sensing, we decided on using the for developing a model that can efficiently and automatically identify water bodies. We consider a small Deep Neural network with 3 layers for water body identification.

For the water bodies extraction, processing of satellite images is required. Both optical data and radar data can be used with the latter demonstrating features that allows mapping of the earth surface during flood events. The histogram-based thresholding methods using SAR amplitude cannot always distinguish water from occluded regions caused by high terrain relief and flat land which has similar reflective characteristics with open water body, deeming it necessary to take the slope into consideration at the water bodies estimation process.

There are several publications e.g. Hong et al. (2015), Clement et al. (2018), Čotar et al. (2016), Lu et al. (2011), Acharya et al. (2018), that study the impact of elevation to the water bodies extraction issue, describing methodologies based on classic remote sensing techniques that reduce its negative effect. Nevertheless, to the best of our knowledge, we present for the first time in literature a neural network based method that includes the DEM in the calculation model, providing encouraging results about the positive impact of elevation information at the estimation of the water bodies, reducing the effect of the highly sloped areas that are erroneously recognized as water. For the training of the model data of 3 different lakes with different topographical features were used, creating a robust model that is able to map lakes and water reservoirs with varied morphology.

The paper is structured as follows. In Section 2 we examine the existing works that are related to the detection of steep areas and the application of methods to reduce its negative impact on the water bodies estimation performance using traditional remote sensing techniques. Section 3 describes the methodology, while Section 4 concerns the experiments and presents the results. Finally, Section 5 concludes and discusses future enhancements.

2 RELATED WORK

Several works have appeared in the literature that or occupied with the effect of elevation information at the water bodies extraction task.

The work at Hong et al. (2015) showed that using direct elevation information to filter occluded areas tend to erase actual water reservoirs that reside at high altitudes. Instead, the slope information showed better performance for this problem. But, still some false positives remain. An object-based classification method with an object size criterion was applied to remove the noises.

Within the methodology of Clement et al. (2018) a terrain filter is applied to remove areas where the topographical location suggests that flooding is unlikely, but where SAR image acquisition may result in misclassification. The Ordnance Survey 5 m Digital Terrain Model (DTM) was used to create Height above Nearest Drainage (HAND) and slope datasets. The slope aspect of the filter is required to remove areas of radar shadow, found when large vertical structures limit the ability of the SAR system to record data from the lee of the feature. The minimal radar response in these areas is similar to that of flat water. The HAND data set represents the topographic difference between a pixel and its hydrologically determined nearest water course (Rennóet al., 2008; Nobre et al., 2011). The addition of HAND reduces the impact of the slope filter in the lowlands by including features such as river banks, which would be otherwise removed. For this project a HAND threshold of 20 m, along with 3 slope, were combined to create the terrain filter.

Čotar et al. (2016) performs post-processing on generated water-bodies masks, based on the assumption that water bodies do not lie on steep terrain and high in the mountains, where snow covers the ground well into spring. Geometric errors are being removed, disregarding all the pixels classified as water on slopes steeper than 8 degrees, in areas higher than 1400 m, and in radar shadows. Also, radiometric errors where ratio (vv/vh) < 0.75 are removed, as well as the official airports that at SAR imagery resembles to water. Finally, a region growing technique is performed to retrieve miss-removed and rough water pixels.

Lu et al. (2011) combine the difference between NDVI and NDWI (NDVI–NDWI) with slope and near infrared (NIR) band. The NDVI–NDWI index is used to enhance the contrast between water bodies and the surrounding surface features; the topographic slope is used to eliminate the mountain shadow; and the NIR band is used to reduce the effects of artificial construction land. Thus, three masks are generated, with their intersection forming the final water bodies mask. For the thresholding of the slope map a hard threshold of 10 was used, due to the water bodies often occur on the areas with a slope lower than 10. The Otsu thresholding method, as introduced by Otsu (1979), is used to automatically estimate the thresholds for the other two maps.

Acharya et al. (2018) evaluated various water indices on a scenery with massive variation in altitude. Comparing visually and quantitatively, not a single method was able to extract surface water in the entire scene with better accuracy. The outcome of the comparison was that with segmenting the test scene with elevations above and below 665m as suggested value, and using NDVI and NDWI respectively for detecting water presented the best results. However, it was susceptible to shadows.

Lately, with the advance of technology, neural networks are being used to facilitate the execution of imagery related application.

Michail et al. (2018) demonstrated a method that is based on the combination of Mahalanobis Distance-based classification to create the flood mask. For the classification, four-dimensional classification features derived directly from the image pixels (namely R, G, B and Near-Infrared channel) are used. Afterwards, a morphological post-processing is performed for the mask correction, filtering very small non-flooded areas and eliminating small non-flooded areas inside flooded area, which are probably false negatives, by applying image dilation and erosion.

To automatically discriminate water and shadows from the background in a SAR image, Zhang et al. (2019) have incorporated a deep neural network model that tackles the noisy information, integrating several deep neural networks. To improve the performance of the Decoder, a high-level feature fusion approach has been followed, where two implementations of the Multi-level Features Extraction and Fusion (MFEF) network are generating separate score maps that are eventually fused with different weights to generate the final score map for water and shadow extraction.

Our method uses both the elevation information combined with the backscatter coefficients to predict the existence of water taking into the consideration the elevation information of the areas, allowing better discrimination of the dark water surfaces and adjacent dark radar shadow areas that may occur in terrain with steep slopes. The model has been trained with data of different areas with different morphological and topographical characteristics deeming it more robust to a variety of under investigation areas.

3 METHODOLOGY

The method proposed is based on Machine Learning (ML) techniques that involves the study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. ML algorithms build mathematical models based on sample data, known as ”training data”, in order to make predictions or decisions without being explicitly programmed to perform the task.

In order to classify the various areas of lake images to the classes “inundated” and ”non-inundated”, we built two DNN models from scratch. The first model takes as input two features, the VV and VH values in decibel per pixel, whereas the second model additionally takes as input the DEM values.

Our model is a fully connected neural network in which all neurons connect to all neurons in the next layer. It involves 3 layers. The tuples of pixel values for each under investigation area are inserted to the first (input) layer, bringing the initial data into the system for further processing. Two hidden layers follows. The first one consists of 12 neurons, whereas the second one of 8 neurons. Eventually, each of the initial tuples of values is classified at the output as ”water” or ”non-water”.

As far as the input of the model is concerned, it is produced by generating tuples of various characteristics for each pixel. Thus, given that the input images are satellite images, that consists of several polarisation bands, we evaluate triples of VV and VH band paired with the elevation values.

We consider the following values to train the model and to eventually perform the water bodies extraction:

- Single VV polarisation band.
- Single VH polarisation band.
- DEM: Shuttle Radar Topography Mission data at 30m Global 1 arc second V003 elaborated by NASA and NGA was used.

We use the following combinations of the aforementioned information files to create a representation for each of the too models:

- VV-VH: The first NN model only uses the backscatter coefficients of the two polarisation bands.
- VV-VH-Dem: The second NN model uses the backscatter coefficients of the two polarisation bands and the corresponding DEM information.

Figure 1 depicts the framework of our Deep Neural Network that takes as input the processed VV and VH bands as well as the corresponding DEM values and outputs the predicted class for each pixel.

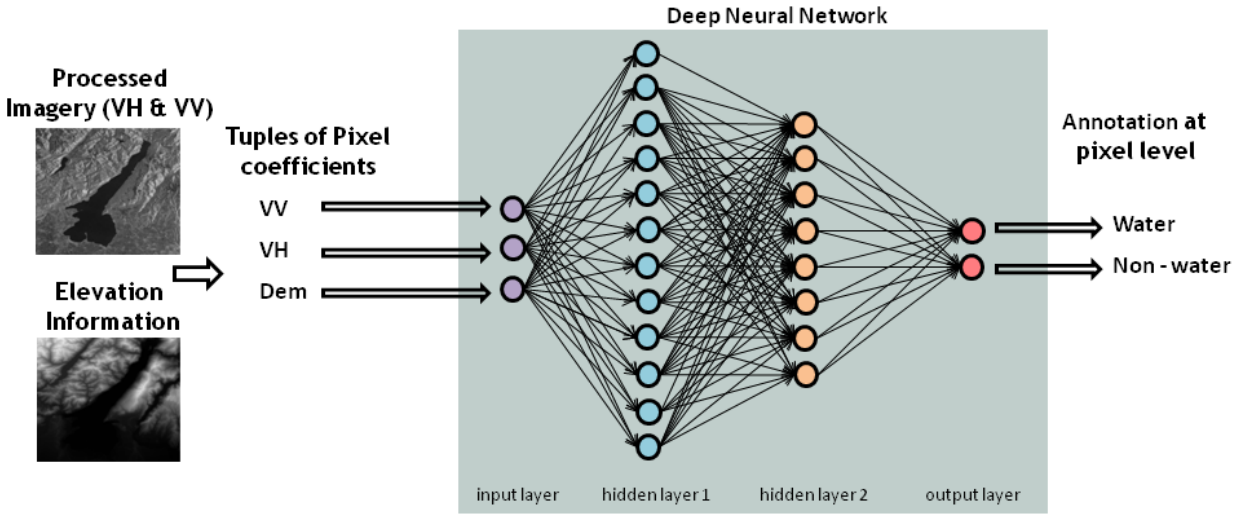

Figure 1: Our Deep Neural Network framework.

3.1 Pre-processing

The current chapter describes the pre-processing steps that are required in order to transform a Sentinel-1 GRD-IW product to a format that is suitable for further analysis. For this purpose the Sentinel Application Platform (SNAP), an open source common architecture for ESA Toolboxes ideal for the exploitation of Earth Observation data is used. Five major steps are executed on both the VH and VV bands:

- **Subset**: the initial product is cropped so it contains only the lake we want to observe and its close surrounding areas. Some balance between the inundated and non-inundated areas is desired.
- **Radiometric calibration**: Fixes the uncertainty in the radiometric resolution of satellite sensor.
- **Speckle noise removal**: Helps removing the pepper and salt like pattern noise that is caused by the interference of electromagnetic waves. "Lee Sigma" filter of Lee (1981) with a 5×5 filter size is used to filter the intensity data.

  As noted by Jong-Sen Lee et al. (2009), this step is essential in almost any analysis of radar images, due to the speckle noise aggravation of the interpretation process. The term noise itself is not strictly correct, because the effect appears due to the coherence of the transmitted pulse, where all of the waves emitted at the same time have the same frequency and phase, and does not reduce the quality of the image.
- **Terrain correction**: Projects the pixels onto a map system (WGS84 was selected) and re-sampled to a 10m spatial resolution. Also, topographic corrections with a Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) is performed. Corrects the distortions over the areas of the terrain.
- **Linear to Decibel (dB)**: The dynamic range of the backscatter intensity of the transmitted radar signal values is usually a few orders of magnitudes. Thus, these values are converted from linear scale to logarithmic scale leading to an easier to manipulate histogram, also making water and dry areas more distinctive.

3.2 Histogram Thresholding

Thresholding is one of the most frequently used techniques to distinguish water areas from land in SAR imagery (e.g. Townsend and Walsh (1998)). As a technique, it is straightforward, with fast computational times and reliable results for general cases. Thus, we use histoghram thresholding on the pre-processed VV and VH bands of the Section 3.1 as a baseline method to discriminate the two classes.

3.3 Deep Learning Slope Reduction Model

For the classification of areas as "inundated" or "non-inundated" a mix of values describing each pixel is required. Thus, triplets of the VV, VH and DEM values are fed as input to the model. Before this step, training the model is performed with a subset of triplets of values of the 3 lakes, paired with the annotation data.

For the implementation we used TensorFlow [*] and Keras [†] which is an open source neural network Python package for developing our models. Keras is used as it simplifies significantly the training of new CNN networks by modifying easily the network structure and the pre-trained weights, by allowing freezing the weights in the imported network and training the weights in the newly added layers. Therefore, it allows essentially easy importing existing knowledge to the new data such as satellite images.

4 EXPERIMENTS

This section contains a description of the dataset used, the settings and eventually the results of the experiments.

4.1 Data Set Description

The dataset is based on three satellite images, each one depicting a different lake in Italy and its surrounding area. The dataset additionally includes the corresponding elevation information. The annotation file delineating lakes and water reservoirs is provided by Alto Adriatico Water Authority (AAWA), in the context of EOPEN project, which is partially funded by the European Commission, under the contract number H2020-776019 EOPEN (2017). To align the annotation file with the under investigation satellite images, Bing Maps high resolution images were used to make corrections to the annotated data. For both the baseline method and the deep leaning approach the three images were pre-processed as described in subsection 3.1. For the baseline method, both VV and polarisation bands of the 3

[*] https://www.tensorflow.org/

[†] https://keras.io/

depictions of the lakes were used as input to estimate the threshold and calculate the water mask. Then the best result for each lake is kept for reference.

For the deep learning approach two different cases occur:

- **Only Polarisation Bands**: In the first case the input for our model are the values of VV and VH bands for each pixel.
- **Polarisation bands paired with DEM data**: In the second case the input for our model are the values of VV, VH bands and the DEM values for each pixel.

For the deep learning dataset, to speedup the training, a random fraction of the initial data was used. The amount of the selected inundated pixels equals the non-inundated pixels creating a balanced dataset, in order to avoid favoring any of the two classes, as most learners will exhibit bias towards the majority class as quoted by Johnson and Khoshgoftaar (2019).

4.2 Settings

To evaluate the performance of the baseline approach, an examination of the automatic thresholding technique was performed on both VV and VH band polarisations. Pixels with values less that the threshold value are marked as water pixels, whereas pixels with values greater or equal the threshold are marked as dry. Since the histogram thresholding technique is performed on different satellite images and areas, the outcome thresholds vary. For each lake the best result between VV and VH is kept. Regarding our approach, we have experimented with the following parameters:

- learning rate, i.e. 0.001, 0.01 and 0.1
- optimizer, i.e. Adam and SGD

The batch parameter is set to 10 and the epochs parameter is set to 25. To evaluate the performance of the different networks, we considered precision, recall, accuracy and F-Score as the evaluation metrics.

4.3 Results

The results of our analysis and the comparison with the aforementioned baseline method are shown in Table 1. It should be noted that the metrics provided refer to the polarisations that demonstrated the best performance.

Regarding the automatically detected thresholds they all fell within the optimum ranges for the classification of flood water as specified by Manjusree et al. (2012), where the backscattering coefficient of water, using Sentinel 1A data with VV polarisations, varies from −6 to −15 dB, and for VH polarisations, it varies from −15 to −24 dB.

Table 1: The baseline results for three major Italian lakes, using histogram thresholding on the processed satellite images.

Lake	Threshold (band)	Precision	Recall	Accuracy	Fscore
Maggiore	-22.0 dB (vh)	0,9623	0,6189	0,7973	0,7533
Garda	-21.7 dB (vh)	0,9555	0,7045	0,8358	0,8110
Trasimeno	-13.9 dB (vv)	0,8807	0,6613	0,7858	0,7554

As far as the proposed approach is concerned, a series of experiments were run that included all possible combinations among the parameters mentioned in the Section 4.2 and the overall top result was presented. For the VV-VH neural network the best results were obtained by using Adam optimizer with learning rate of 0.001 and can be seen in Table 2.

Table 2: These are the results for three major Italian lakes, based on our deep learning method with the VV and VH values.

Lake	Optimizer	Learning rate	Precision	Recall	Accuracy	Fscore
Maggiore	Adam	0.001	0,9137	0,9088	0,9115	0,9112
Garda	Adam	0.001	0,9001	0,8490	0,8774	0,8738
Trasimeno	Adam	0.001	0,8064	0,9580	0,8640	0,8757

Similarly, for the VV-VH-DEM neural network the best results were obtained by using Adam optimizer with learning rate of 0.001 and can be seen in Table 3.

Table 3: These are the results for three major Italian lakes, based on our deep learning method with the VV,VH and DEM values.

Lake	Optimizer	Learning rate	Precision	Recall	Accuracy	Fscore
Maggiore	Adam	0.001	0,9829	0,8784	0,9315	0,9277
Garda	Adam	0.001	0,9457	0,9363	0,9413	0,9410
Trasimeno	Adam	0.001	0,9367	0,8461	0,8945	0,8891

In here we can see the direct impact of DEM to our neural network model. For the Maggiore lake the F-Score increased from 91.12% to 92.77%, when moving from VV-VH to VV-VH-DEM, comparing to 75.33% of the baseline. For the Garda lake the F-Score increased from 87.38% to 94.10%, comparing to 81.10% of the baseline. Finally, for the Trasimeno lake the F-Score increased from 87.57% to 88.91%, comparing to 75.54% of the baseline.

Figure 2 depicts the processed VV band for the three lakes. Figure 3 depicts the processed VH band for the three lakes. These were used as input to our neural network and to the baseline approach.

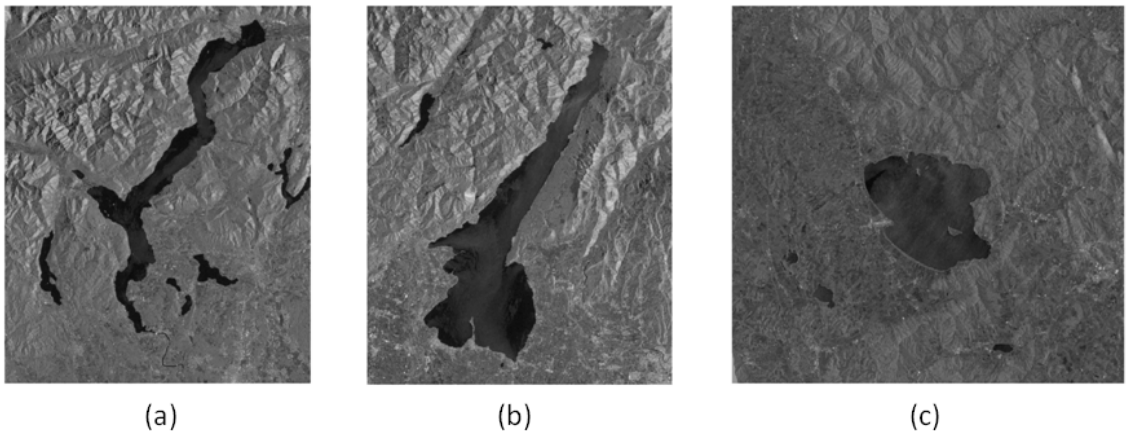

Figure 2: Processed VV band of lakes (a) Maggiore, (b) Garda, (c) Trasimeno.

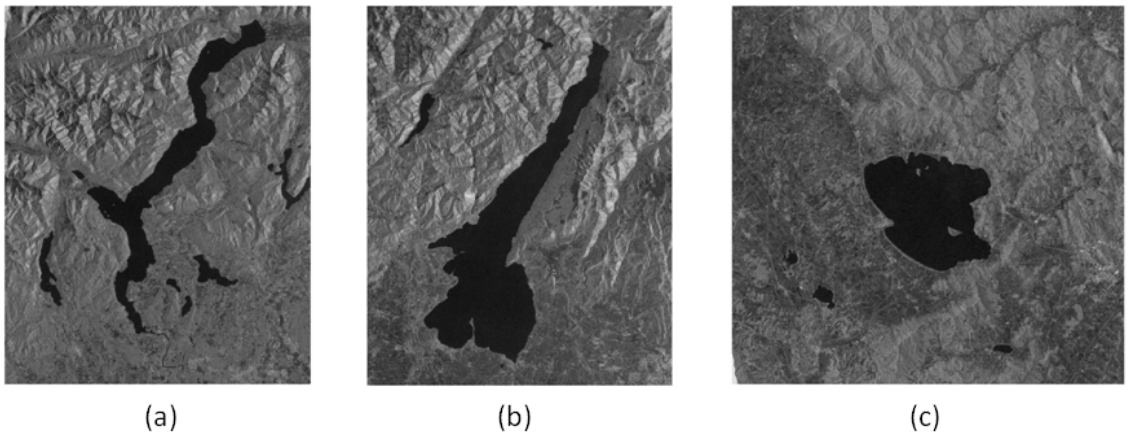

Figure 3: Processed VH band of lakes (a) Maggiore, (b) Garda, (c) Trasimeno.

Figure 4 depicts the generated water masks for the three lakes using histogram thresholding, where we can see some qualitative results of the baseline approach.

Figure 5 depicts the generated water masks for the three lakes using our DNN trained with VV band, VH band and DEM file, where we can see some qualitative results of our approach.

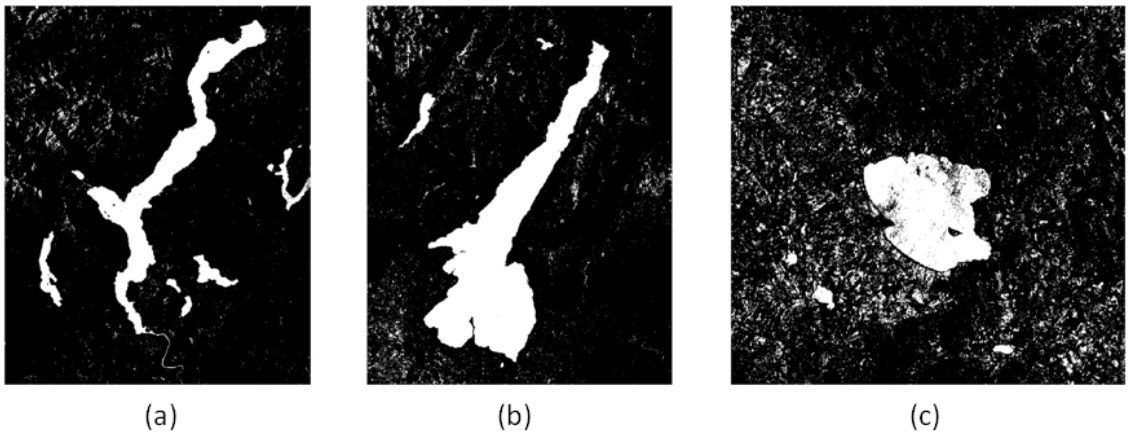

Figure 4: Water masks of lakes (a) Maggiore, (b) Garda, (c) Trasimeno using histogram thresholding.

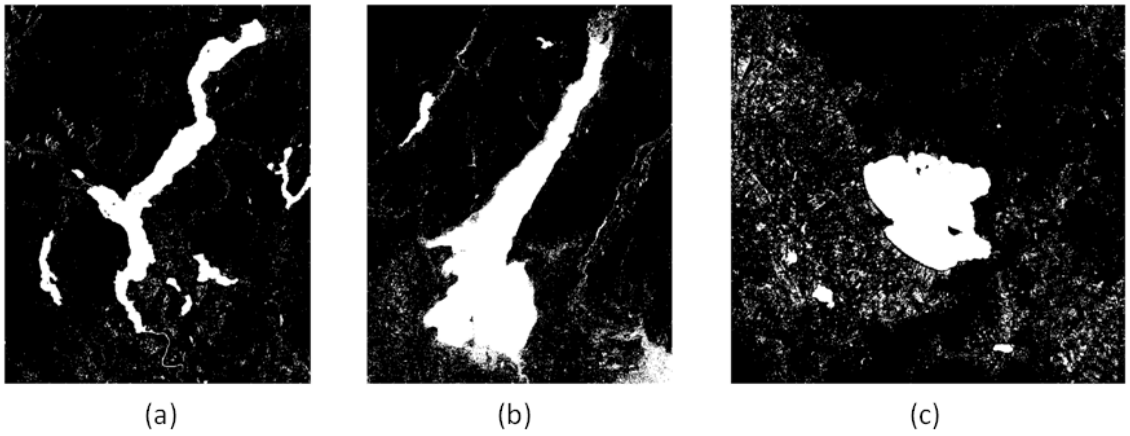

Figure 5: Water masks of lakes (a) Maggiore, (b) Garda, (c) Trasimeno using our DNN with VV, VH and DEM.

5 CONCLUSION AND FUTURE WORK

SAR imagery is characterized of high noise. The thresholding approaches are usually applied on a single polarisation band and are able to provide satisfactory results but only for smoother water surfaces. Filtering the steep sloped areas can improve the delineation of the detected water masks. However, there are cases it worsens the results, cause it may remove actually flooded areas lying on steeper river banks.

Combining both VV and VH polarisation bands with the corresponding elevation information in a deep neural network increases the performance of the water extraction task, reducing the false positives that are caused of the steep mountain slopes. Our proposed model is able to balance the amplitude and elevation characteristics of satellite imagery and achieve accuracy and F-score evaluation measures up to 94.1%.

In conclusion, the filtering out of highly-sloped districts can have very positive impact on the SAR radar images. Nevertheless, there are cases that global slope filtering worsens the precision of the generated water-bodies masks. Thus, a model that incorporates the elevation information among with the amplitudes is deemed necessary.

Future work includes training a Deep Convolutional Neural Network (DCNN), like VGG-Net or ResNet, with full-training of the DCNN model using images produced by combining features extracted of satellite images. In this case pretrained weights are not being used but the model is trained from scratch. Alternatively, a pretrained DCNN can be used with satellite imagery as input to fine-tune existing networks that are trained on a large dataset like the ImageNet.

References

Acharya, T., Subedi, A., and Lee, D. (2018). Evaluation of water indices for surface water extraction in a landsat 8 scene of nepal. *Sensors*, 18(8):2580.

Clement, M., Kilsby, C., and Moore, P. (2018). Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, 11(2):152–168.

Čotar, K., Oštir, K., and Kokalj, Ž. (2016). Radar satellite imagery and automatic detection of water bodies. *Geodetski glasnik*, 50(47):5–15.

EOPEN (2017). opEn interOperable Platform for unified access and analysis of Earth observatioN data. https://eopen-project.eu/. Online; accessed 06 September 2019.

Hong, S., Jang, H., Kim, N., and Sohn, H.-G. (2015). Water area extraction using radarsat sar imagery combined with landsat imagery and terrain information. *Sensors*, 15(3):6652–6667.

Johnson, J. M. and Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27.

Jong-Sen Lee, Jong-Sen Lee, Jen-Hung Wen, Ainsworth, T. L., Kun-Shan Chen, and Chen, A. J. (2009). Improved sigma filter for speckle filtering of sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):202–213.

Lee, J.-S. (1981). Refined filtering of image noise using local statistics. *Computer graphics and image processing*, 15(4):380–389.

Lu, S., Wu, B., Yan, N., and Wang, H. (2011). Water body mapping method with hj-1a/b satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 13(3):428–434.

Manjusree, P., Kumar, L. P., Bhatt, C. M., Rao, G. S., and 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.

Michail, E., Moumtzidou, A., Gialampoukidis, I., Avgerinakis, K., Scarpino, M. G., Vrochidis, S., Vingione, G., Kompatsiaris, I., Labbassi, K., Menenti, M., and Elghandour, F.-E. (2018). Testing a flood mask correction method of optical satellite imagery over irrigated agricultural areas. In *2nd Mapping Water Bodies from Space Conference (MWBS2018)*.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

Townsend, P. A. and Walsh, S. J. (1998). Modeling floodplain inundation using an integrated gis with radar and optical remote sensing. *Geomorphology*, 21(3):295 – 312. Application of remote sensing and GIS in geomorphology.

Zhang, P., Chen, L., Li, Z., Xing, J., Xing, X., and Yuan, Z. (2019). Automatic extraction of water and shadow from sar images based on a multi-resolution dense encoder and decoder network. *Sensors*, 19(16):3576.