

Flood detection with Sentinel-2 satellite images in crisis management systems

Anastasia Moumtzidou

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
moumtzid@iti.gr

Marios Bakratsas

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
mbakratsas@iti.gr

Stelios Andreadis

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
andreadisst@iti.gr

Anastasios Karakostas

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
akarakos@iti.gr

Ilias Gialampoukidis

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
heliasgj@iti.gr

Stefanos Vrochidis

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
stefanos@iti.gr

Ioannis Kompatsiaris

Centre for Research & Technology Hellas
Information Technologies Institute
Thessaloniki, Greece
ikom@iti.gr

ABSTRACT

The increasing amount of falling rain may cause several problems especially in urban areas, which drainage system can often not handle this large amount in a short time. Confirming a flooded scene in a timely manner can help the authorities to take further actions to counter the crisis event or to get prepared for future relevant incidents. This paper studies the detection of flood events comparing two successive in time Sentinel-2 images, a method that can be extended for detecting floods in a time-series. For the flood detection, fine-tuned pre-trained Deep Convolutional Neural Networks are used, testing as input different sets of three water sensitive satellite bands. The proposed approach is evaluated against different change detection baseline methods, based on remote sensing. Experiments showed that the proposed method with the augmentation technique applied, improved significantly the performance of the neural network, resulting to an F-Score of 62% compared to 22% of the traditional remote sensing techniques. The proposed method supports the crisis management authority to better estimate and evaluate the flood impact.

Keywords

Floods, Change detection, Bi-temporal analysis, Sentinel-2, Deep Neural Networks.

INTRODUCTION

In every disaster and crisis, getting accurate information about the scope, extent, and impact of the disaster is critical to creating and orchestrating an effective disaster response and recovery effort. Climate conditions are expected to change worldwide. This includes an increase in intensity and frequency of (among others) extreme weather events. As a result, flooding will become more common in the future.

Additionally, the global warm and arctic ice melt has led to sea level rising, which also has increased the flood risk in coastal urban areas. Urban flooding, on the other hand, is explained by the lack of drainage in an urban area. Specifically, when little open soil exists to be used as water storage, all the precipitation needs to be transported to surface water or the sewage system.

The free, full and open Copernicus images mainly contribute to the recovery phase of a flood crisis event, where a map needs to be created with the areas that suffer most, so as to set priorities in the operational plan. The analysis of a series of flood events over time, also contributes in policy-making, so as to identify the most vulnerable areas for the preparedness phase in future events. For a better overview, the usefulness of such applications can be demonstrated by crisis management systems like Copernicus Emergency Management Service (Copernicus EMS¹), which provides information for emergency response in relation to different types of disasters, including flood events.

Flood event detection involves a sequence of satellite images over a specific area of interest, where the problem is to detect significant changes on water levels. So far the problem has been approached by remote sensing techniques that use thresholds on pre-defined indices from multi-spectral bands (Hussain et al. 2013). On the other hand, computer vision techniques that focus on the use of Deep Convolutional Neural Networks (DCNN) are sensitive to noise and demand very large collections of annotated images for training their models.

In this work, we detect the existence of a flood event between two satellite images by using a DCNN model, that considers as input a combination of Sentinel-2 bands. There are two novelties introduced by this work. The first involves the input of the DCNN, which is the image differences of images, consisting of selected bands for each of the 3 channels that form the false-color PNG image. The second is the data augmentation techniques used for obtaining more data for the training of the DCNN. Specifically, two types of augmentation are realized; the first considers the characteristic features of the satellite images while the second involves taking the difference of all images against all other images in the same collection of images in order to increase significantly the number of positive instances and thus obtain a more balanced dataset.

The paper is structured as follows. Initially, we examine the existing works that are related to change detection techniques using traditional remote sensing techniques and machine learning approaches. Right after, we describe the methodology, followed by the experiments and the corresponding results. Finally, we conclude the paper and discuss future enhancements.

RELATED WORK

There exist several works that consider classic remote sensing techniques for water content estimation based on spectral indices. The work in (Li et al. 2015) compares 10 years of water level fluctuations with seven spectral indices at a 16-day interval. Spectral indices demonstrated good potential for characterizing and monitoring temporal variability in the hydrology of small seasonally-flooded wetlands. Particularly Tasseled Cap Brightness Index (TCBI) proved useful and gave consistent good results for wet and dry years, and for areas characterized by different inundation frequencies. This is relevant as it could provide opportunity to improve hydrological monitoring particularly for data-poor and ungauged wetlands. Nevertheless, it was not tested in urban areas, in contrast with our benchmark dataset that we used to train our DCNN model. Then, we tested our work with several similar approaches based on spectral indices.

Zhou et al. reviewed and compared existing open surface water body mapping approaches based on six widely-used water indices, including the tasseled cap wetness index (TCW), normalized difference water index (NDWI), modified normalized difference water index (mNDWI), sum of near infrared and two shortwave infrared bands (Sum457), automated water extraction index (AWEI), land surface water index (LSWI), plus three variations, and used as input data imagery of three medium resolution sensors (Landsat 7 ETM+, Landsat 8 OLI, and Sentinel-2 MSI). *NDWI* and *NDWI_{plusVI}* gave best results (Zhou et al. 2017). We compare our approach with respect to the above indices.

Moradi et al. explore the full spectral potential of Landsat8 to calculate the Modified Optimization Water Index (MOWI) taking into consideration the linear combination of bands, where for each band a coefficient is calculated by particle swarm algorithm (Moradi et al. 2017). Particle swarm optimization had been used for the calculation of

¹<https://emergency.copernicus.eu/>

each band coefficient. The result showed a proper performance on different conditions like cloud, cloud shadow and mountain shadow. We estimate the cloud coverage so as to remove the parts of the satellite images that introduce noise to the model creation.

Acharya et al. demonstrated that for the complex Nepal area there is no single spectral index that is able to extract surface water in the entire scene with best accuracy for all its sub-areas (Acharya et al. 2018). Upon selecting optimum thresholds, the overall accuracy (OA) and kappa coefficient (kappa) was improved, but not satisfactory. NDVI and NDWI showed better results for only pure water pixels, whereas MNDWI and AWEI were unable to reject snow cover and shadows. Combining NDVI with NDWI and AWEI with shadow mask improved the accuracy but inherited the NDWI and AWEI characteristics. Segmenting the test scene with elevations above and below 665m, and using NDVI and NDWI for detecting water, resulted in an OA of 0.9638 and kappa of 0.8979. The accuracy can be further improved with a smaller interval of categorical characteristics in one or multiple scenes. Best results were obtained by applying different indices on the various parts of the scenery depending on the elevation of each area. They showed that elevation plays an important role in selecting the water index method and their optimum values. However, data fusion with Digital Elevation Models was beyond the scope of this paper.

For the task of change detection, interest has also grown at the examination of a time series of satellite images, facilitating the creation of reference images that depict the normal state of an area, whereas it allows us to track the time that a change has happened. Clement et al. used a sequence of multiple Synthetic Aperture Radar images to generate a reference image based on the median value of each sequence of pixels (Clement et al. 2018). Then, bi-temporal analysis was applied between the under investigation image and the reference image, to estimate the flooded areas. Contrary to these approaches, we have used changed detection techniques on several spectral indices that are sensitive to water bodies estimation.

A recent approach on exploiting both temporal and spectral features of satellite images using Temporal Convolutional Neural Networks (TempCNNs) is proposed in (Pelletier et al. 2019). They showed that the use of global pooling layers, which drastically reduces the number of trainable parameters, harms the classification of the time series images. Thus, studying the influence of pooling layers before any integration into a TempCNN network is highly recommended. On contrary, the use of local average pooling tends to favor the model. The also showed that manually-calculated spectral features, such as the NDVI, does not seem to improve TempCNN models.

We also created a DCNN model that is able to predict water related changes between any two images within a sequence of images. This model is trained on false-colour composite images that consist of combinations of three spectral bands, that are sensitive in water detection.

More recently, machine learning techniques were used for detecting changes in satellite images. Huang et al. and Bovolo et al. treat change and no - change as a binary classification problem using Support Vector Machine (SVM) (Huang et al. 2008), (Bovolo et al. 2008), which is a well-known supervised non-parametric statistical learning technique. Some other machine learning algorithms used for classification and change monitoring include; random forest (Pal 2005), (Smith 2010). However, these works require segmentation of the image and annotation per pixel. Thus, eventually the results and analysis are made on a pixel-level. Contrary to these approaches, our DCNN-based method does not require per pixel annotation, but simply characterizes the whole image as flooded or not compared to another one.

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 differences of satellite images either as 'flooded' or 'non-flooded' we build models by using pre-trained Convolutional Neural Networks (CNN). There should be noted that we move from normal state to flooded conditions. Thus, we experimented with the following models: VGG-16 (Simonyan and Zisserman 2014), VGG-19 (Simonyan and Zisserman 2014), Inception-v3 (Szegedy et al. 2016), and ResNet-50 (He et al. 2016). VGG-16 and VGG-19 were originally developed for the ImageNet dataset by the Visual Geometry Group at the University of Oxford. The models involve 16 and 19 layers respectively and the size of input image is 224 x 224 pixels. Regarding, Inception-v3, it is another ImageNet-optimized model. It is developed by Google and has a strong emphasis on making scaling to deep networks computationally efficient, having as input 299 x 299 images. Finally, ResNet-50 is developed by Microsoft Research and uses residual functions to add stability to deep networks, using as input 224 x 224 images. In all aforementioned networks, the input layer consists of 3 nodes,

Table 1. Sentinel-2 Bands description

Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR1	1.610	20
Band 12 - SWIR2	2.190	20

which correspond to the three color channels R,G,B. In order to exploit knowledge captured in existing neural networks, we perform fine-tuning which involves using pre-trained network on a large and diverse dataset like the ImageNet and modify it slightly to capture similar knowledge. Specifically, we removed the last pooling layers and replacing them with a new pooling layer with a softmax activation function with size 2 given that the model recognizes whether there is evidence of change detection due to floods or not.

Therefore, as far as the proposed fine-tuned model, we consider as input the difference between the bands of the images. However, it should be noted that the input images are satellite images and thus the number of bands are more than 3 compared to the standard 3-channel input of the pre-trained models. Table 1 includes the complete list of the Sentinel-2 bands and a small description. Therefore, in order to use the existing pre-trained networks, it is necessary to select a subset of available bands.

Based on the descriptions of the bands and given that we are interested in recognizing water, we evaluated different triples of bands that are water sensitive, in order to form the false-RGB composition. Namely, we considered the following visible bands and bands near the infrared spectrum, although they are inter-correlated to each other:

- NIR - Water absorbs almost all light at the wavelength window. This makes the water bodies to appear very dark. The contrasts with the soil and vegetation bright reflectance makes it good band for the definition of water/land interface.
- SWIR1 - This band is very sensitive to moisture and therefore it is used to indicate and monitor the vegetation and soil moisture in the image.
- Red - Vegetation absorbs almost all the red light in the band. It is used to distinguish between soil and vegetation and also to estimate the vegetation health.
- Green - It senses in a strong chlorophyll absorption region and in a strong reflectance region for most soils. Thus, it can discriminate vegetation and soil.

Thus, we used as input for the DCNN the following combinations of the aforementioned bands for image representation: Red - Green - Blue (RGB), Red - SWIR1 - NIR (RSN), and Green - SWIR1 - NIR (GSN). The output of the DCNN as already mentioned, has 2 nodes since it detects change between two images due to floods.

Another issue that is most common in supervised methods, is the size of training dataset. Given that the proposed approach is based on DCNN, a significant amount of training data is required in order to train efficiently the model. Towards this direction, we applied augmentation techniques to the initial satellite images. In general, data augmentation is a common technique used for increasing the performance of a DCNN model. Since we are analysing satellite images that depict urban cities, transformations that preserve the shape of objects are used to avoid distortion of rigid-shaped constructs such as buildings (Buslaev et al. 2018). Alteration of the brightness and contrast of the images was applied. Rotation and flipping the images can help to effectively increase the size of the dataset (Yu et al. 2017).

The aforementioned modifications could simulate the alterations of the images' illumination due to the daylight conditions. Apart from altering the images by applying several types of transformations, and given that the provided

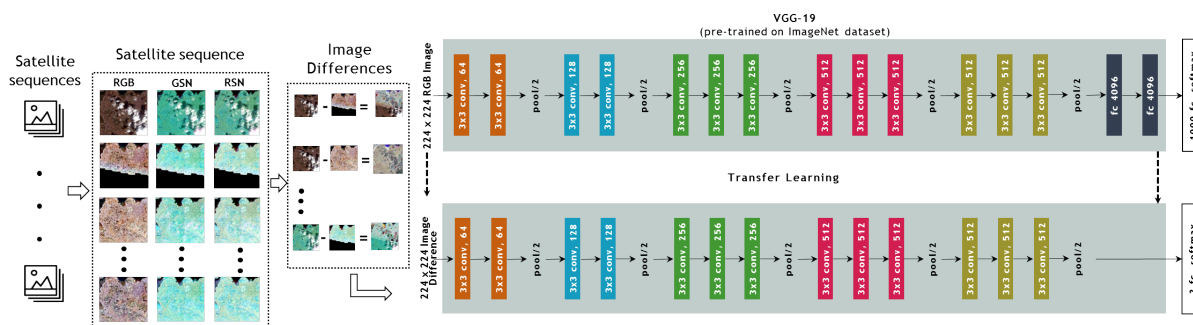


Figure 1. DCNN-based method framework.

dataset consists of time-series with N images, we increased the images by considering not only the $N - 1$ differences of two sequential images but the k -combination:

$$\binom{N}{k} = \frac{N!}{k!(n-k)!} \quad (1)$$

where N is the number of images inside a time-series and k is the number of images considered per selection, which equals to 2. This type of augmentation allows us to increase significantly the number of positive instances, which are usually few in such cases and thus allows us to obtain a more balanced, but also enriched dataset. In our case, the dataset increases approximately about a factor of 50% because the $N - 1$ differences become:

$$\binom{N}{2} = \frac{N(N-1)}{2} \quad (2)$$

For the implementation we used TensorFlow ² and Keras ³ 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.

Figure 1 depicts the outlier detection framework.

EXPERIMENTS

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

Data set description

The dataset consists of a set of sequences of satellite images that depict a certain city over a certain length of time and are provided for the MediaEval 2019 Satellite Task ⁴, and specifically for the "City-centered satellite sequences" subtask (Bischke et al. 2018). In total 335 events, each consisting of varied number of images, are provided, 267 of which are considered as the training set whereas the remaining 68 are the test set. Each event has a number of layers that depict different acquisition dates of the satellite images. Therefore, each sequence can be represented as a 3D array with size equal to image width x image height x number layers. Layers are of size 512x512 pixels for the 10m bands (B03, B04, B08). The 20m band (B11) band was sharpened to 10m thus re-sized to the same dimensions as the previous bands.

For the DCNN-based approach, as we have already mentioned, data augmentation was applied in order to increase the training set. Specifically, for each image sequential rotations of 90 degrees were applied, combined with 3 different modifications of the contrast and brightness ratio, also flipping horizontally, resulting that dataset increased 24 times. In order to further augment the dataset we considered as input not only the difference between two consecutive layers (dates) but between any two images within the same event. This led eventually to a dataset of 180,000 image differences as training set and 58,000 image differences as test set. However, given the limitations imposed by GPU, we have narrowed down the dataset to 30,000 records as train set and 9,000 records as test set.

²<https://www.tensorflow.org/>

³<https://keras.io/>

⁴¹<http://www.multimediaeval.org/mediaeval2019/multimediasatellite/>

The selection of the images that would form the two datasets was random. Moreover, it should be noted that both the training and test dataset were balanced. Furthermore, for the creation of the false-color PNG images the following triplets of bands were considered: Red - Green - Blue, Red - SWIR1 - NIR and Green - SWIR1 - NIR. Finally, an extra parameter that was considered for creating the dataset was whether the satellite image has any missing part due to the partially covered tiles that correspond to those at the edge of the swath path at a satellite pass. To tackle this issue, we produced two different train and test datasets. In the first one, we considered as input images that had 100% of the pixels appearing correctly, while in the second one, we considered as input images that had at least 50% of the pixels appearing correctly. It should be noted that in both cases images that had less of 50% of the pixels appearing were omitted.

Image differencing

The purpose of this methodology is to determine if inside a series of satellite images of a specific area has occurred a flood event. At start, for each timely consecutive pair of the imagery a difference image for this pair is calculated, depicting the extent of the changed value for each pixel. A significant change at a minimum amount of water pixels between two consecutive images denotes a flood incident.

To discover the type of input data that would provided optimum results, the following eight distinct types of images were used to calculate the differences, resulting to eight separate executions:

i) The vegetation indices are produced separately for two images and then subtracted to each other, in order to detect changes in vegetation (Mancino et al. 2014). Similarly, we experiment with spectral indices that specialise in the extraction of water bodies. The four known indices that are generated are these of:

- The Modification of normalised difference water index (*MNDWI*) (Xu 2006) can enhance open water features while efficiently suppressing and even removing built-up land noise as well as vegetation and soil noise.
- *NDWI₁* was introduced for the first time by Gao (Gao 1996), reflects moisture content in plants and soil.
- *NDWI₂*, another variation of NDWI (McFeeters 1996) is used to differentiate water from the dry land. Water bodies have a low radiation and strong absorbability in the visible infrared wavelengths range. It's subtle in land built-up and often ends up in overestimated water bodies.
- *NDVI* is a measure of the state of plant health based on how the plant reflects light at certain frequencies (some waves are absorbed and others are reflected). But can also be used for the detection of water bodies. Values near zero speak to shake and exposed soil, whereas negative qualities speak to water, snow and mists (Ganie and Nusrath 2016).

These indices were selected for their ability to discriminate water from non-water and since they are using different spectral bands, they demonstrated different behaviour at the estimation of the water bodies.

ii) 4 Raw bands: B03 (Green), B04 (Red), B08 (NIR), B11 (SWIR1).

Image ratioing

This method resembles the image differencing technique, with the differentiation that in here the two consecutive days are divided by each other. If the intensity of reflected energy is nearly the same in each image then the ratio pixels with no change will have values close to union, otherwise changed pixels will significantly diverge (A. Singh 1989). The same 8 executions as image differencing are performed.

Change vector analysis

Change Vector Analysis (CVA) is a change detection tool that characterize dynamic changes in multi-spectral space by a change vector over two different time instances in terms of magnitude and direction (S. Singh and Talwar 2013). It is able to process any number of spectral bands that are required. We used the combination of bands SWIR1, NIR and Red due to their sensitivity at water detection. To create the binary classification image we used the change magnitude (CM). An optimum threshold was detected using the formula (Leutner n.d.): $\text{threshold} = \text{multiplier} * \text{median}(\text{CM}[\text{CM} > 0])$.

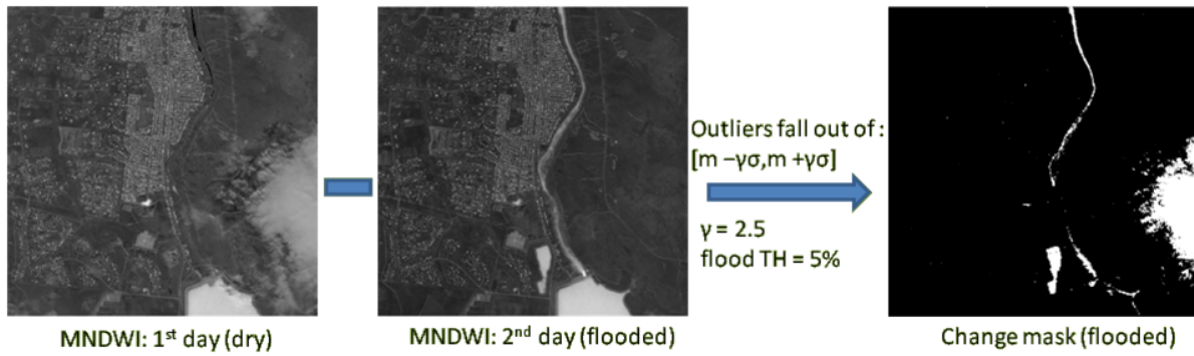


Figure 2. Difference of consecutive MNDWI.

Settings

To evaluate the performance of the baseline approach, a comparison with three change detection techniques that are based on subtraction of indices or bands was performed. Pixels with no change in radiance are distributed around the mean in the difference image (Lu et al. 2005). Pixel values that fall within $[m - \gamma\sigma, m + \gamma\sigma]$ denote no change, i.e. no flood. The m and σ are the mean and standard deviation of the differences image. For the optimum γ value multiple values between 1.0 and 4.0 were tested. In order to characterize a difference image as "changed" a minimum amount of changed pixels needs to be detected. Various water pixel ratios were used to find the one that gives the optimum results. Tested for values from 0.01 to 0.20.

Regarding our approach, we have experimented with the following parameters:

- type of network, i.e. VGG-16, VGG-19, Inception v3, ResNet-50
- learning rate, i.e. 0.001, 0.01, 0.1
- optimizer, i.e. Adam and SGD
- bands used for image representation: Red - Green - Blue, Red - SWIR1 - NIR, and Green - SWIR1 - NIR
- image missing part: 0.5% (i.e. at least half of the image should be visible) 1.0% (i.e. whole image should be visible - do not have missing parts)

The batch parameter is set to 32 and the epochs parameter is set to 50. To evaluate the performance of the different networks, we considered precision, recall and F-Score as the evaluation metrics.

Results

The results of our analysis and the comparison with the aforementioned baseline methods are shown in Table 2. It should be noted that the metrics provided refer to the test part of the dataset.

Following the baseline approach, a series of executions of all combinations of change detection base techniques and the available input data was conducted. The ability of each method to detect any water related change between two consecutive days is evaluated. The results with the optimum parameters for water ratio and gamma are presented in Table 2. For the baseline methods, the image differencing with $NDWI_2$ data provided the best results, followed by CVA with the Red-SWIR1-NIR (RSN) dataset. Image ratioing finished third among the baseline techniques with best results on the NIR band. It should be noted that different type of input were considered for the baseline method, including four fundamental spectral indices known for their ability to discriminate water bodies and four Sentinel-2 bands, which exist at the formulas of the aforementioned indices, thus associated with water sensitivity. In this way a comparison between indices and bands is conducted on change detection. Figure 2 depicts an example of the remote sensing technique, where two MNDWI images of consecutive days are being subtracted to each other and then the outlier technique highlights the changed pixels.

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 *Settings* sub-section. However given the space limitations, only the top 7 results are provided. The best performing runs were highlighted in bold, they were achieved for VGG-19 and VGG-16 when RGB bands were used and reached an F-Score of 62,35% and 60,06% respectively, which is

Table 2. Evaluation of change detection techniques for two images

Method		Parameters			Precision	Recall	Fscore		
		Index	Water Ratio	Gamma					
Baseline Methods	Image differencing	$NDWI_1$		0,1	1,5	0,116	0,6911	0,1986	
		$NDWI_2$		0,07	1,9	0,1577	0,3821	0,2233	
		$MNDWI$		0,08	1,7	0,1262	0,5366	0,2043	
		$NDVI$		0,1	1,7	0,1778	0,2602	0,2112	
		NIR		0,1	1,7	0,159	0,3089	0,2099	
		SWIR		0,12	1,5	0,129	0,5285	0,2073	
		Green		0,12	1,5	0,1245	0,2683	0,1701	
		Red		0,02	3,2	0,1224	0,3333	0,179	
	Image ratioing	$NDWI_1$		0,01	4	0	0	0	
		$NDWI_2$		0,01	2,1	0,0994	0,1463	0,1184	
		$MNDWI$		0,01	1	0,0226	0,0325	0,0267	
		$NDVI$		0,01	1,1	0,0983	0,1382	0,1149	
		NIR		0,01	3,9	0,1023	0,4865	0,169	
		SWIR1		0,01	4	0,0923	0,3243	0,1437	
		Green		0,01	3,6	0,0865	0,2432	0,1277	
		Red		0,01	3,3	0,0909	0,2703	0,1361	
	CVA	RSN		0,05	2	0,1121	0,5285	0,1849	
		RSN		0,15	2,3	0,1291	0,4472	0,2004	
		Index	Optimizer	Learning rate	Augme- ntation	Visible image	Precision	Recall	Fscore
DCNN-based Methods	VGG-16	GSN	Adam	0,001	2-comb.	100%	0,5142	0,4525	0,4814
		RSN	Adam	0,001	2-comb.	>50%	0,5265	0,5439	0,5351
		RGB	Adam	0,001	2-comb. & transf.	>50%	0,5499	0,6616	0,6006
		RGB	Adam	0,01	2-comb. & transf.	100%	0,4504	1,0000	0,6211
	ResNet-50	RGB	Adam	0,001	2-comb.	100%	0,5265	0,5998	0,5608
	VGG19	RGB	Adam	0,001	2-comb. & transf.	>50%	0,4530	1,0000	0,6235
		RGB	Adam	0,001	2-comb.	100%	0,5799	0,2473	0,3468



Figure 3. Two consecutive RGB images and their difference image.

significantly higher than 22,3%, which was the F-Score of the best run for the baseline methods. The VGG-16 RGB with learning rate equal to 0,01, although it has the second best F-Score, it has very low precision and therefore, it was not preferred as second best run. This can be partly explained by the fact that original networks were trained for RGB images. Another conclusion that can be drawn is that among the optimizers, the Adam provided the best results, with a learning rate almost always equal to 0,001. Finally, we should stress the crucial role that augmentation played, as the best performing DCNNs proved the ones on which full augmentation was applied (i.e. includes both 2-comparison and image transformations) and of course compared to the baseline methods, all the runs provided were significantly better which can be attributed both to the nature of the method and to the improved and balanced dataset that was created. In figure 3 the two input RGB images and their subtraction is presented. The difference is more obvious on the top part of the area highlighting the flooded area.

CONCLUSION AND FUTURE WORK

In this work we presented our approach in change detection for flood mapping using Sentinel-2 images by considering the recent advances in Deep Neural Networks. Our method was compared with three well-known baseline change detection techniques, i.e. image differencing, ratio differencing and change vector analysis that used as input data spectral indices or Sentinel-2 bands. The DCNN proposed method was evaluated for several settings including different type of models, optimizer functions and datasets and the results showed, apart from the fact that it is superior to the standard remote sensing techniques, that applying data augmentation to the training dataset is crucial for obtaining good results.

The dataset used is a set of Sentinel 2 images and we focus our analysis in Sentinel 2 mission. The Copernicus Programme of Europe is one of the largest set of missions and provides S2 images in a free, full and open basis, and we therefore aim at providing a mature solution for S2 images.

The specification of the flooded areas is a crucial issue for the disaster management authorities. With this information, they are able to correlate the flooded areas with their relevant characteristics (resources, infrastructures etc) in order to get prepared for future events. Additionally, the authorities can estimate the impact of the flood for a specific area and the results for the measures that have been taken.

Future work includes the evaluation of the DCNN-based approach to the whole time-series of Sentinel-2 data, also the use as input of images with different resolution such as the ones provided by Worldview, LandSat and other satellite constellations and finally, the full-training of a DCNN model using satellite images which involves not using pre-trained weights but rather training the model from scratch.

ACKNOWLEDGEMENTS

This work was supported by the EC-funded projects H2020-832876-aqua3S and H2020-776019-EOPEN.

REFERENCES

- Acharya, T., Subedi, A., and Lee, D. (2018). "Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal". In: *Sensors* 18.8, p. 2580.
- Bischke, B., Helber, P., Zhao, Z., Bruijn, J. de, and Borth, D. (2018). "The Multimedia Satellite Task at MediaEval 2018." In: *MediaEval*.
- Bovolo, F., Bruzzone, L., and Marconcini, M. (2008). "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure". In: *IEEE Transactions on Geoscience and Remote Sensing* 46.7, pp. 2070–2082.

- Buslaev, A., Parinov, A., Khvedchenya, E., Iglovikov, V. I., and Kalinin, A. A. (2018). “Albumentations: fast and flexible image augmentations”. In: *arXiv preprint arXiv:1809.06839*.
- Clement, M., Kilsby, C., and Moore, P. (2018). “Multi-temporal synthetic aperture radar flood mapping using change detection”. In: *Journal of Flood Risk Management* 11.2, pp. 152–168.
- Ganie, M. A. and Nusrath, A. (2016). “Determining the vegetation indices (NDVI) from Landsat 8 satellite data”. In: *International Journal of Advanced Research* 4.8, pp. 1459–1463.
- Gao, B.-C. (1996). “NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space”. In: *Remote sensing of environment* 58.3, pp. 257–266.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- Huang, C., Song, K., Kim, S., Townshend, J. R., Davis, P., Masek, J. G., and Goward, S. N. (2008). “Use of a dark object concept and support vector machines to automate forest cover change analysis”. In: *Remote Sensing of Environment* 112.3, pp. 970–985.
- Hussain, M., Chen, D., Cheng, A., Wei, H., and Stanley, D. (2013). “Change detection from remotely sensed images: From pixel-based to object-based approaches”. In: *ISPRS Journal of photogrammetry and remote sensing* 80, pp. 91–106.
- Leutner, B. (n.d.). *Change Vector Analysis*. <https://bleutner.github.io/RStoolbox/rstbx-docu/rasterCVA.html>.
- Li, L., Vrieling, A., Skidmore, A., Wang, T., Muñoz, A.-R., and Turak, E. (2015). “Evaluation of MODIS spectral indices for monitoring hydrological dynamics of a small, seasonally-flooded wetland in southern Spain”. In: *Wetlands* 35.5, pp. 851–864.
- Lu, D., Mausel, P., Batistella, M., and Moran, E. (2005). “Land-cover binary change detection methods for use in the moist tropical region of the Amazon: a comparative study”. In: *International Journal of Remote Sensing* 26.1, pp. 101–114.
- Mancino, G., Nolè, A., Ripullone, F., and Ferrara, A. (2014). “Landsat TM imagery and NDVI differencing to detect vegetation change: assessing natural forest expansion in Basilicata, southern Italy”. In: *iForest-Biogeosciences and Forestry* 7.2, p. 75.
- McFeeters, S. K. (1996). “The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features”. In: *International journal of remote sensing* 17.7, pp. 1425–1432.
- Moradi, M., Sahebi, M., and Shokri, M. (2017). “MODIFIED OPTIMIZATION WATER INDEX (MOWI) FOR LANDSAT-8 OLI/TIRS.” In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42.4/W4.
- Pal, M. (2005). “Random forest classifier for remote sensing classification”. In: *International Journal of Remote Sensing* 26.1, pp. 217–222.
- Pelletier, C., Webb, G. I., and Petitjean, F. (2019). “Temporal convolutional neural network for the classification of satellite image time series”. In: *Remote Sensing* 11.5, p. 523.
- Simonyan, K. and Zisserman, A. (2014). “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *CoRR* abs/1409.1556. arXiv: [1409.1556](https://arxiv.org/abs/1409.1556).
- Singh, A. (1989). “Review article digital change detection techniques using remotely-sensed data”. In: *International journal of remote sensing* 10.6, pp. 989–1003.
- Singh, S. and Talwar, R. (2013). “Review on different change vector analysis algorithms based change detection techniques”. In: *2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*. IEEE, pp. 136–141.
- Smith, A. (2010). “Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm”. In: *Journal of Spatial Science* 55.1, pp. 69–79.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.
- Xu, H. (2006). “Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery”. In: *International journal of remote sensing* 27.14, pp. 3025–3033.

- Yu, X., Wu, X., Luo, C., and Ren, P. (2017). “Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework”. In: *GIScience & Remote Sensing* 54.5, pp. 741–758.
- Zhou, Y., Dong, J., Xiao, X., Xiao, T., Yang, Z., Zhao, G., Zou, Z., and Qin, Y. (2017). “Open surface water mapping algorithms: A comparison of water-related spectral indices and sensors”. In: *Water* 9.4, p. 256.