APPLICATION OF SEMANTIC SEGMENTATION WITH FEW LABELS IN THE DETECTION OF WATER BODIES FROM PERUSAT-1 SATELLITE'S IMAGES

J. Gonzalez[1], K. Sankaran [2],V. Ayma[1], C. Beltran[1]

[1] Pontifical Catholic University of Peru, Lima, Peru - (jessenia.gonzalezv, vayma, cbeltran)@pucp.edu.pe
[2] Mila, Université de Montréal, Montreal, Canada -(kris.sankaran@umontreal.ca)

Commission VI, WG VI/4

KEY WORDS: Semantic segmentation, remote sensing, water bodies detection, satellite images, PeruSAT-1.

ABSTRACT:

Remote sensing is widely used to monitor earth surfaces with the main objective of extracting information from it. Such is the case of water surface, which is one of the most affected extensions when flood events occur, and its monitoring helps in the analysis of detecting such affected areas, considering that adequately defining water surfaces is one of the biggest problems that Peruvian authorities are concerned with. In this regard, semiautomatic mapping methods improve this monitoring, but this process remains a time-consuming task and into the subjectivity of the experts.
In this work, we present a new approach for segmenting water surfaces from satellite images based on the application of convolutional neural networks. First, we explore the application of a U-Net model and then a transfer knowledge-based model. Our results show that both approaches are comparable when trained using an 680-labelled satellite image dataset; however, as the number of training samples is reduced, the performance of the transfer knowledge-based model, which combines high and very high image resolution characteristics, is improved.

1. INTRODUCTION

Remote sensing is frequently used to gather data from areas we live in, to observe the effects and impact of nature, and evaluate human activities (Van Westen, 2000). Nowadays, satellite images have proven to be a reliable source of information due to the improvements in spatial resolution. Such is the case of PeruSAT-1, which has a resolution of 2.8 m in the Red, Green, Blue and NIR bands; and, 0.7 m in the panchromatic band (eoPortal Directory, 2018).

The recollection of satellite images from several regions of Peru might provide insightful information about population growth, contamination, forest felling, among others. In this research, we focus our attention on the analysis of water-bodies (e.g. rivers, lakes, ponds) from the coast of Peru to assess the affected regions by the Niño costero phenomenon, which is a recurrent natural phenomenon in Peru and has a large impact on agricultural production, social services, and infrastructure (Ramírez, Briones, 2017). Once the catastrophe has occurred, the next step is to quantify the damage so that infrastructure and sanitation projects can be defined to aid recovery of affected areas, restore the normal functioning of services, and improve the life quality of the victims.
When dealing with flooding or contaminated water, the complexity of detecting water bodies is that these frequently appear in non-homogeneous ways due to the combination with other different materials (e.g. rocks, soil, trees, construction materials, among others). This lack of homogeneity entails variability in the parameters or features that can be extracted (e.g. intensity of color, texture) from the image. Besides, due to the huge spread of water bodies and, in some cases, the inaccessibility of the regions affected, the retrieval of images and data represents a challenging task.

Different methods have been proposed to detect water bodies. The use of single-band thresholds and multi-band thresholds were reported in (Nath, Deb, 2010). Spectral water index from satellite images was computed in (Du et al., 2016), (Jiang et al., 2014), and in (Pekel et al., 2016), the authors also proposed a classification scheme based on expert systems, visual analytics, and evidential reasoning to evaluate 32 years of water body data from satellites. Although there exist some approaches using big data analysis and image processing algorithms to cope with these tasks (Ayma et al., 2016), most of the solutions still rely on the expertise of researchers to select the thresholds and determine which features are the most representative. For these reasons, in the last years, there are several investigations using convolution neural networks such as those reported in (Miao et al., 2018), (Nowaczynski, 2017), (Feng et al., 2019), (Talal et al., 2018), and (Hu et al., 2019).

The main objective of this study is to detect water bodies in satellite images captured by PeruSAT-1 and create water body maps to aid in the evaluation of the flood-affected areas. We created a manually labelled dataset from these satellite images, based on which we propose to use semantic segmentation to classify each pixel as part of a water body. We also explore the use of convolutional neural networks to segment water bodies using a training dataset with few labeled images. We first evaluate the performance of a conventional U-Net model using only the PeruSAT-1 dataset. Afterward, we used the idea of knowledge distillation using another dataset with a lower resolution than the PeruSAT-1 dataset. Images for the second dataset are available from the Sentinel-2 satellite (QueryPlanet, 2019).

The article is divided as follows: Section II describes the state-of-the-art in the analysis of satellite images. Section III defines the proposed methodology. Section IV presents the dataset used in this research. Section V introduces the experiments and results, and Section VI concludes and discusses the research results.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-183-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165643

2. STATE OF THE ART

As mentioned in the introduction, there are different methods used to detect water bodies from satellite images. In the following, we will provide a few key details on these ideas.

1. Threshold segmentation: this method identifies water bodies by applying thresholds to one or more spectral bands. The drawback of this proposal is that there is a possibility to lose some parts of the water body if the threshold value is not selected adequately. In (Jiang et al., 2014), and (Zhiyuan Zhang, 2018), the correct detection of water bodies depends on the selection of the correct bands and appropriate thresholds.

2. Spectral water index: one of the most used methods is the NDWI (Normalized Difference Water Index) proposed by McFeeters (McFeeters, 1996). This index maximizes the reflectance of bodies in the green band and minimizes the reflectance in the NIR band. Some disadvantages of this method are the following: inefficient detection of mixed water pixels, confusion of water bodies with background noise, and thresholding that is dependent on the satellite and the different regions and times analyzed. Also, this index is not robust in the presence of clouds and non-pure water. Shadows from clouds, mountains or buildings are also frequently confused with bodies of water. Xu, *et al.* developed the MNDWI (Modified Normalized Difference Water Index) to address the shortcomings of the NDWI (Xu, 2006). This index uses the Shortwave Infrared (SWIR) band instead of the NIR band used in NDWI. The research suggested that MNDWI is more suitable to enhance water information and can extract water bodies with better accuracy than NDWI (Jiang et al., 2014), (Du et al., 2016); however, MNDWI is not widely used because most high-resolution satellites only capture data in four bands (blue, green, red and near-infrared), such as in the case of PeruSAT-1 satellite.

3. Active contour model: the advantage over other models is that it integrates image data, initial estimates, target contour features, and constraints in the feature extraction process. However, it is necessary to choose an initial point of the contour of the water, and automatic acquisition of this initial position is not frequently easily achieved. Furthermore, these manual decisions impact on the accuracy and efficiency of the detection of water bodies (Feng et al., 2019), (Hemalatha et al., 2018).

4. Object detection based on classification: this method combines the spectral and textural features of remote sensing images with classical machine learning methodologies (e.g. SVM, decision tress, among others). However, in (Huang et al., 2015), the authors found out that due to diminished spectral separability and complications concerning the identification of the same spectrum across different water bodies and/or various spectra within a single water body, object-based technology falls short in several aspects. These features and the method used to select the most relevant ones impacts also in the precision and efficiency of the entire proposal hindering its performance and robustness when applying to different datasets.

5. Deep Learning: it has been verified that CNN (convolutional neural networks) have better performance than classical algorithms due to their robustness, invariance to translation, among other factors (Krizhevsky et al., 2012). However, common CNN architectures do not allow a pixel-wise classification of an image; that is, they do not generate any context information that helps the segmentation of objects in those images. Therefore, the literature describes, initially, a technique that uses CNN to classify small regions (sliding windows) in images and then join said sliding windows to obtain a pixel-wise classification of all the objects recognized in the scene (Cireșan et al., 2012). The computational cost of this proposal was too high, so new ideas emerged such as Fully Convolutional Neural Networks (FCN) (Shelhamer et al., 2017) or other more advanced networks such as U-Net, DeepLab, among others. All these architectures have an end-to-end configuration; that is, they receive images as inputs and then provide images as outputs; however, the U-Net (Ronneberger et al., 2015) has the advantage of requiring a smaller number of images for training, since it is a neural network that has only 23 convolutional layers without additional stages. In addition, following the results of the survey reported in (Hu et al., 2019), its use is suitable for working with remote sensing images.

3. METHODOLOGY

We developed two different approaches. First, we trained a model using only the very high-resolution (VHR) dataset from PeruSAT-1. Then, we used a high-resolution (HR) dataset for the knowledge transfer process. With the second approach, we aim to improve the performance where the first methodology fails, and when we have a limited number of images labeled from the VHR dataset. Additional details about the configuration of the mentioned architectures and datasets are described bellow.

3.1 Data

Two evaluate the performance of both proposed methodologies, we used satellite images from Sentinel-2 and PeruSAT-1. The details are described as follows.

- HR dataset: taken from a publicly available dataset, containing 7671 Sentinel-2 image patches of size 64×64 pixels each (QueryPlanet, 2019). Sentinel-2 is a European wide-swath, high-resolution, multi-spectral imaging mission. The images from this satellite have 13 spectral bands varying from 443nm to 2190nm. Also, images have the red, green, blue, near-infrared (10m), red edge and short-wave infrared bands (20m), as well as, three atmospheric correction bands (60m) (SUHET, n.d.). In our study, we use 4 bands from this satellite (i.e. red, green, blue, and NIR), each of 10m per pixel resolution, to match the same characteristics of the PeruSAT-1 satellite.

- VHR dataset: built on images captured with PeruSat-1 satellite, which is a very high spatial resolution satellite with 2.8m per pixel in the red, green and blue band, and 0.7m in the panchromatic band. The images were collected from the coast of Peru from different periods, pre and post-disaster. The original images have an approximate size of 6000 × 6000 pixels.

The QGIS software was used to label images and to generate masks for the semantic segmentation training process. To do this, we first added a vector layer to create the polygons. Then, using image processing algorithms, we eliminated missing information (e.g. borders of the satellite image) and further define the edges of water bodies. Finally, the original images were split into small patches. We obtained 1113 patches of 512×512

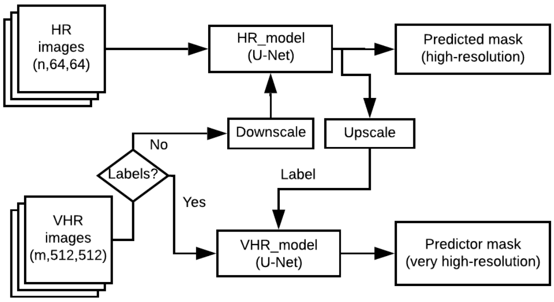

Figure 1. Knowledge transfer based on two models, which uses mapping information from lesser resolution images to improve the segmentation capability of the very high-resolution images.

pixels, from which 945 have binary labels, considering 1 for pixels representing water areas, and 0 otherwise; and the remaining 168 patches correspond to unlabeled images.

3.2 Architectures

Model 1: to train this model, we used images from the VHR dataset. The neural network architecture was based on a variation of the U-Net called TernausNet (Iglovikov, Shvets, 2018), which uses the VGG11 as an encoder. It is a fully convolutional network (FCN) with 23 convolutional layers and its architecture follows an encoder-decoder model. This neural network concatenates low-level feature maps with higher-level ones, enabling a precise pixel-level localization, which is very useful for a pixel-wise classification scheme. The contracting path (i.e. the encoder) consists of 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. The expansive path (i.e. the decoder) increases the resolution of the detected features. In our study, we tuned the network to perform pixel-wise binary classification, for water or no water classes.

Model 2: in this approach, we used the idea of knowledge distillation. Distillation was introduced by (Hinton et al., 2015) and it is a method proposed to do knowledge transfer from larger architectures to smaller ones. Additionally, the authors in (Papernot et al., 2015) extended that idea and proposed a defense distillation architecture in which knowledge transfer is done between the same architectures to improve its own resilience to adversarial samples. Using these ideas, we design the second model which is shown in Figure 1. This architecture has two U-Nets, one working on high-resolution (HR) images, and the other working on very high-resolution (VHR) images, and both networks designed for performing the semantic segmentation, trained at once. The images from the HR dataset are used to train the HR-model, which is in charge of computing the missing labels for the VHR dataset. When a missing label is detected in the VHR dataset, a downscale is applied to the corresponding image and then a label is predicted using the HR model. This label obtained is then upscaled in order to get the very high-resolution label for the original input image.

4. EXPERIMENTS AND RESULTS

For the experiments, we used the following hardware: a Lambda workstation with Ubuntu 18.04 and two NVIDIA RTX 2080 GPUs with 12 GB of memory each. For the implementation of the algorithms, we used Python v3.6 programming language and PyTorch 1.0 framework for deep learning architectures.

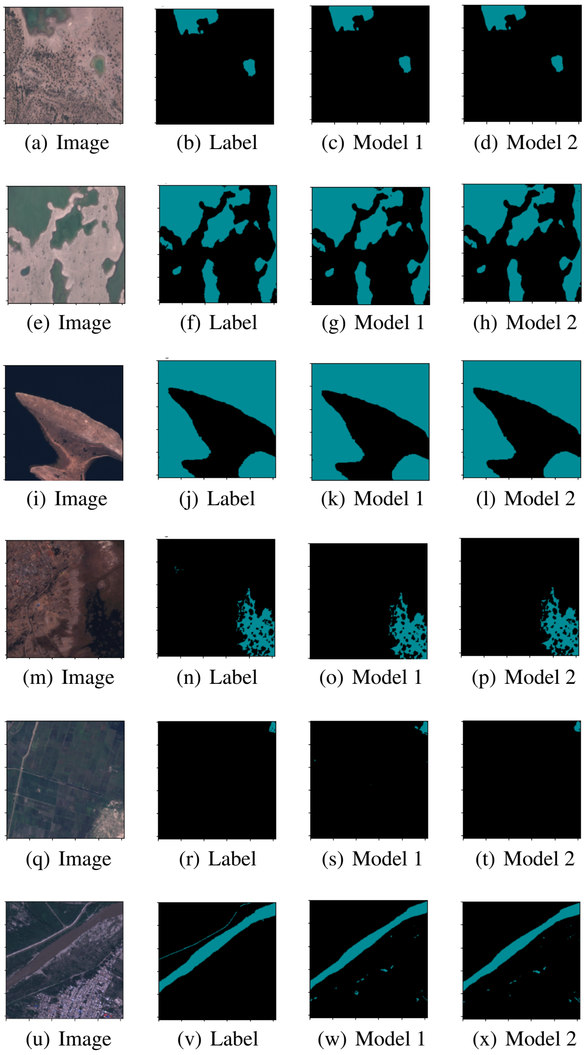

Figure 2. Example image patch (a), (e), (i), (m), (q), (u), ground-truth label (b), (f), (j), (n), (r), (v), predictions of the model 1 (c), (g), (k), (o), (s), (w), and predicitions of the model 2 (d), (h), (l), (p), (t), (x).

The HR-dataset was divided into training, and validation splits, each with 7057, and 613 samples, respectively. For the labelled VHR-dataset, we used 680, 170 and 95 samples for the training, validation, and testing dataset; and regarding the unlabelled data, we used 131, and 37, for the training, and validation.

All the experiments used the same VHR labelled testing dataset. For our first model, we used all the labelled VHR data, and the results are shown in the third column of Figure 2. For the second approach, we used the HR-dataset, and both, labelled and unlabelled VHR-dataset and their outcomes are presented in the fourth column of the same figure.

In a qualitative evaluation of the predicted masks, we observed that for several cases (such as in Figure 2(a) - 2(q)), both models compute the water bodies labels correctly as they match well with the ground truth label. However, in some cases such as the one in Figure 2(u), we realized that the narrower river in the image is not segmented and additional noise appears in both models. Therefore, having the same qualitative results for both

	Model 1		**Model 2**	
Training samples	**F1 score**	**IoU**	**F1 score**	**IoU**
680	0.9317	0.9124	0.9038	0.8823
340	0.9092	0.8844	0.8935	0.8844
170	0.9111	0,8846	0.9099	0,8849
68	0.8946	0,8665	0.9093	0,8819

Table 1. Average IoU and F1 score values for 4 datasets of training and evaluated on a 95-image test dataset using model 1 and model 2.

	Model 1		**Model 2**	
Training samples	F1 score %	IoU %	F1 score %	IoU %
68 (Fig. 3(a))	78.43	64.62	88.40	79.61
680 (Fig. 3(e))	97.63	95.36	95.78	92.14
68 (Fig. 3(i))	21.79	18.22	77.64	66.57
680 (Fig. 3(m))	87.15	77.56	84.53	74.14
68 (Fig. 3(q))	0.08	0.00	0.13	0.00
680 (Fig. 3(u))	21.71	20.00	20.49	20.00

Table 2. Average values of IoU and F1 score for specific samples which are part of 95 test dataset using model 1 and model 2.

models would imply that in our case the information from HR images does not directly help the segmentation of VHR data, following the configuration of the training datasets.

To evaluate whether the second model, based on knowledge transfer, increases the performance over the first one when using less data, we divided the number of labeled VHR training samples into smaller datasets of 340, 170, and 68 samples. Results shown in Table 1 suggest that even with these reduced datasets, the overall performance using the second model is similar to the first one. However, further evaluation of each image in the testing dataset shows that when we decreased the training samples, images with narrow rivers or sediments, and some cases with no water bodies, are better segmented using the second model rather than the first one. These specific cases are shown in Figure 3, and its corresponding metrics appear in Table 2. Figures 3(g), 3(h), 3(o), 3(p), 3(w), and 3(x), describe similar performances when training both models using 680 samples from the VHR-labeled dataset, and it can be observed in rows 2, 4 and 6 from Table 2 a small variation in the metrics between the two models. When we decreased the training dataset to 68 images from the VHR labeled dataset, we observed that the performance of the second model increased. Figures 3(d), 3(l), and 3(t)) reflect a better segmentation when using the knowledge-transfer based model, compared to the outcomes achieved by the first one and presented in Figures 3(c), 3(k), 3(s). Regarding the metrics from this scenario, rows 1, 3 and 5 in Table 2 confirm such increase in the F1 score and IoU metrics for every case analyzed.

5. CONCLUSION AND DISCUSSION

In this work, we compared two different deep learning methods to segment water bodies in satellite images from Peru. We used data collected from PeruSAT-1 and Sentinel-2 satellites. The first approach consists of variations of the U-Net architecture; meanwhile, the second approach uses the concept of distillation to enhance the U-Net response when fewer training images are available. For this second methodology, HR images from Sentinel-2 were used to guide the learning process of VHR image segmentation of PeruSAT-1 data.

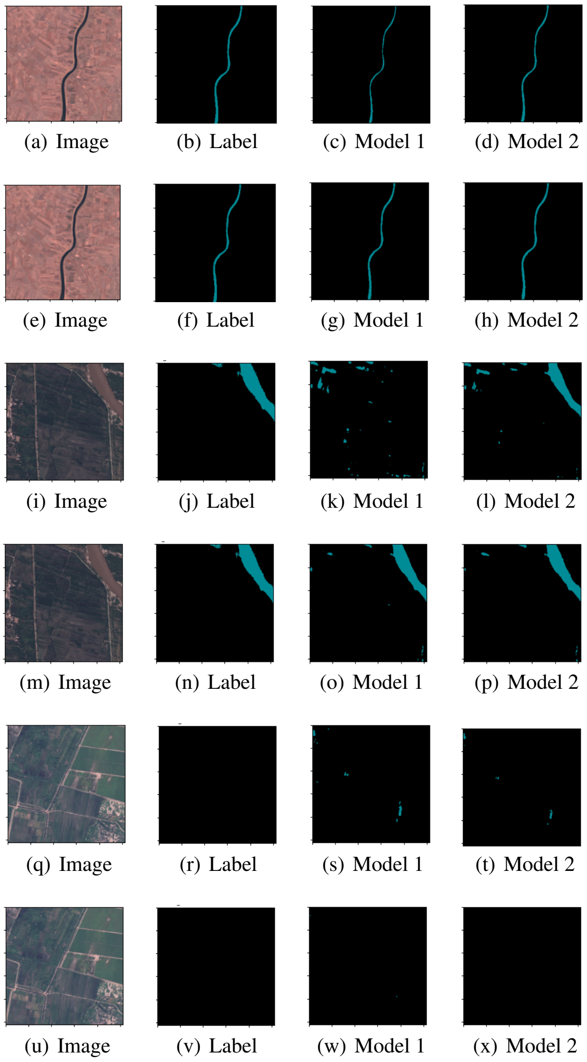
(a) Image (b) Label (c) Model 1 (d) Model 2
(e) Image (f) Label (g) Model 1 (h) Model 2
(i) Image (j) Label (k) Model 1 (l) Model 2
(m) Image (n) Label (o) Model 1 (p) Model 2
(q) Image (r) Label (s) Model 1 (t) Model 2
(u) Image (v) Label (w) Model 1 (x) Model 2

Figure 3. Example image patch (a), (e), (i), (m), (q), (u), ground-truth label (b), (f), (j), (n), (r), (v), the predictions when it has 68 training samples from model 1 (c),(k),(s) as well as of the model 2 (d),(l),(t), and the predictions when it has 680 training samples with the model 1 (g),(o),(w) as well as with the model 2 (h),(p),(x).

One of the reasons we found for the overall performance to be similar in both models is that images in which the segmentation using the model 2 outperforms model 1 represent examples of narrow rivers, water-bodies with sediments, among other complex scenarios, which are cases that do not appear frequently in the training/testing dataset. This behaviour may suggest that model 2 takes advantage of different resolution in the training dataset to make the segmentation process more robust in the presence of complex cases. In addition, following the results presented in Table 1, we could argue that the variance in model 1 is increasing when adding more images to the training dataset. On the other hand, the variance on model 2 remains the same even when more data is added. This hypothesis translates in the performance shown Table 2.

Finally, we do believe our findings will be of interest to other researches currently working with limited-label datasets, which nowadays is a common case in most remote sensing researches.

ACKNOWLEDGEMENTS

The authors would like to thank the support of FONDECYT (National Fund for Scientific, Technological Development and Technological Innovation) under the financing agreement No. 131 - 2018 (FONDECYT - SENCICO), the Artificial Intelligence Laboratory at Pontificia Universidad Catolica del Peru, and CONIDA (National Aerospace Research and Development Commission).

REFERENCES

Ayma, V., Costa, G., Nigri Happ, P., Feitosa, R., Ferreira, R., Oliveira, D., Plaza, A., 2016. A New Cloud Computing Architecture for the Classification of Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP, 1-8.

Cireșan, D., Giusti, A., Gambardella, L. M., Schmidhuber, J., 2012. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. *NIPS*.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sensing*.

eoPortal Directory, 2018. PeruSat-1 Earth Observation Minisatellite.

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*.

Hemalatha, R., Thamizhvani, T., Dhivya, A., Joseph, J., Babu, B., Chandrasekaran, R., 2018. Active Contour Based Segmentation Techniques for Medical Image Analysis.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the Knowledge in a Neural Network.

Hu, J., Li, L., Lin, Y., Wu, F., Zhao, J., 2019. A Comparison and Strategy of Semantic Segmentation on Remote Sensing Images.

Huang, X., Xie, C., Fang, X., Zhang, L., 2015. Combining Pixel- and Object-Based Machine Learning for Identification of Water-Body Types From Urban High-Resolution Remote-Sensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 1-14.

Iglovikov, V., Shvets, A., 2018. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *ArXiv e-prints*.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J., Xiao, T., 2014. An automated method for extracting rivers and lakes from Landsat imagery. *Remote Sensing*.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *ImageNet Classification with Deep Convolutional Neural Networks*.

McFeeters, S. K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425-1432. https://doi.org/10.1080/01431169608948714.

Miao, Z., Fu, K., Sun, H., Sun, X., Yan, M., 2018. Automatic Water-Body Segmentation from High-Resolution Satellite Images via Deep Networks. *IEEE Geoscience and Remote Sensing Letters*.

Nath, R., Deb, S., 2010. Water-Body Area Extraction From High Resolution Satellite Images-An Introduction, Review, and Comparison. *International Journal of Image Processing (IJIP)*, 3(3), 353–372.

Nowaczynski, A., 2017. Deep learning for satellite imagery via image segmentation.

Papernot, N., McDaniel, P. D., Wu, X., Jha, S., Swami, A., 2015. Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks. *CoRR*, abs/1511.04508. http://arxiv.org/abs/1511.04508.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*.

QueryPlanet, 2019. QueryPlanet: AI meets EO.

Ramírez, I. J., Briones, F., 2017. Understanding the El Niño Costero of 2017: The Definition Problem and Challenges of Climate Forecasting and Disaster Responses. *International Journal of Disaster Risk Science*, 8(4), 489–492. https://doi.org/10.1007/s13753-017-0151-8.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 234–241.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

SUHET, n.d. Sentinel-2 User Handbook. MIT.

Talal, M., Panthakkan, A., Mukhtar, H., Mansoor, W., Almansoori, S., Ahmad, H. A., 2018. Detection of water-bodies using semantic segmentation. *2018 International Conference on Signal Processing and Information Security (ICSPIS)*, 1–4.

Van Westen, C., 2000. Remote sensing for natural disaster management. *ISPRS 2000 congress : geoinformation for all : Amsterdam, the Netherlands, 16-23 July, 2000. pp. 1700-1707*, International Society for Photogrammetry and Remote Sensing (ISPRS), 1700–1707.

Xu, H., 2006. Modification of Normalized Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *International Journal of Remote Sensing*, 27, 3025–3033.

Zhiyuan Zhang, Haixia He, C. Y. W. Z. L. L. L. M., 2018. Using the modified two-mode method to identify surface water in Gaofen-1 images. *Journal of Applied Remote Sensing*, 13(2), 1 - 16 - 16. https://doi.org/10.1117/1.JRS.13.022003.