Automatic Detection of Change of Water Bodies on RGB Satellite Images

NICO FUHRBERG, NIKLAS STOCKFISCH, and HAUKE TIETJEN

Concerning environmental development, landscape monitoring, e.g. of water bodies, may play a feasible role in detecting, predicting and preventing problematic natural scenarios. In this work, we explore a method to segment water on RGB satellite images utilizing convolutional neural networks and use this information for assessing the development of water bodies. We find that our approach yields promising results for moving the complexity of water segmentation from data representation to the underlying computational model, though more sophisticated and extensive measurements will be necessary in order to obtain results that are competitive to works on raw satellite images.

1 INTRODUCTION

Having a tool to forecast desiccation of lakes [11] or coastal erosion [14, 18] is useful to seek alternative water sources in time or to deploy countermeasures to protect the coast from erosion.

Our work took form in pursuit of the following research questions: Can a machine learning model be used to predict water segmentation masks on RGB satellite images? Can the visually apparent change of the water bodies be observed in the image segmentation? If so, can further statements, e.g. concerning the future of the respective water bodies be predicted? How do the results on RGB images compare to the ones on raw satellite images?

2 RELATED WORK

For water segmentation on raw satellite images, various approaches exist ranging from unsupervised, clustering-based techniques [5] to ones involving deep learning techniques [9, 10]. Our approach differs from these, as it uses RGB images as its input.

On landscape segmentation on satellite images, more approaches exist focusing on trees in the western West African Sahara and Sahel using convolutional neural networks [2]; other works on landscape image segmentation with convolutional neural networks focus on segmentation of woody vegetation on repeat photographs [1].

3 DATASETS

Our approach was to train a segmentation model on predicting the water segmentation on RGB satellite images, expecting to yield a model that is suitable to predict said segmentation on unseen images. Therefore, our efforts on accumulating data had been focused on acquiring separate training and application datasets. This methodology implies that the generaliseability of the training dataset will play a significant role in our segmentation system.

3.1 Training Dataset

For our training, we used a dataset found on Kaggle [6]. It contains a collection of 2841 RGB images captured by the Sentinel-2 satellite with respective water segmentation masks. These masks were generated by calculating the Normalized Water Difference Index (NDWI) [12], which is derived from the raw images Near and Short Wave Infrared channels – a common practice in this field of application [9] [13] – and applying a threshold.

Supposedly due to technical interna in the generation of the base images, they often did not show a perfectly cut out satellite view but also had black edges of varying sizes. For one portion of these images, their entire content was majorly black, rendering their segmentation masks unintelligible for our training, as can be seen on Figure 1. Therefore

Authors' address: Nico Fuhrberg, nico6@uni-bremen.de; Niklas Stockfisch, niksto@uni-bremen.de; Hauke Tietjen, hauke2@uni-bremen.de.

we automatically detected and removed all images containing more than 30% of purely black pixels from our dataset, yielding a modified dataset containing 2753 images.

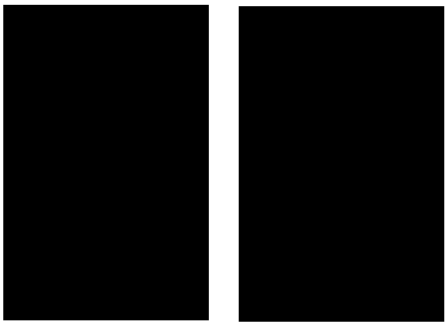

Fig. 1. Example of a base image and corresponding segmentation mask with an inordinate amount of black pixels.

For the remaining images, we still had to take the possible black edges into account which are for example seen on Figure 2, as training on these edges may lead to inaccurate results for our model. As we expected most of the actual satellite images to not contain many purely black pixels, we applied a brightness segmentation with a low, empirically determined threshold on the base images. We then used these images to color our training masks in a different color where the segmented base images were not black, resulting in segmentation masks with colored portions of non-border pixels, rendering white a color to be ignored by our model during training.

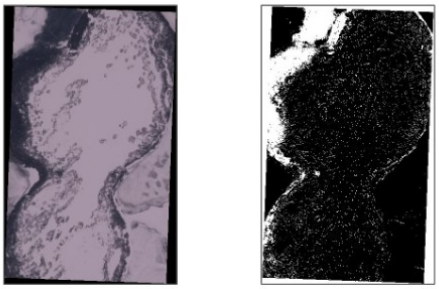

Fig. 2. Example of a base image and corresponding segmentation mask with notable black borders.

3.2 Application Dataset

For the application of our trained model, we used the time lapse featured of Google Earth Engine [8] on 10 selected water bodies on which we could observe significant development regarding their water surface. The dataset consists of fixed-area screenshots of these water bodies for each year from 1984 to 2020, resulting in 370 images. For each scene we estimated the size of the total area shown using the built-in area measurement tool as a scaling factor for our are calculation. Our observed water bodies as well as their respective visible phenomena are shown in table 1. For our evaluation, we manually created two segmentation masks per water body, therefore twenty in total, using GIMP [17].

Area	Phenomena
Aral sea	Desiccation
Atolls of the Maldives	Land Reclamation
Aydar lake	Lake level rise
Chugach National Forest	Melting Glaciers
Helgoland	Coastal Erosion
Hwanggang dam	Dam building
Iquitos	Meandering rivers
Mississippi River Delta	Land Subsidence and Coastal Erosion
Palm Islands	Land Reclamation
Sylt	Coastal Erosion

Table 1. Overview of observed water bodies and their respective phenomena.

4 SEGMENTATION MODEL

For our segmentation model we chose DeepLabV3+, which is a state-of-the-art semantic segmentation model, using an encoder-decoder architecture (see Figure 3), extending original segmentation models with fully convolutional neural networks, suitable for the images of our datasets of greatly varying sizes. As our encoder (backbone) we used MobileNetV2 [16] for being rather lightweight and thus potentially strengthening the availability of our application.

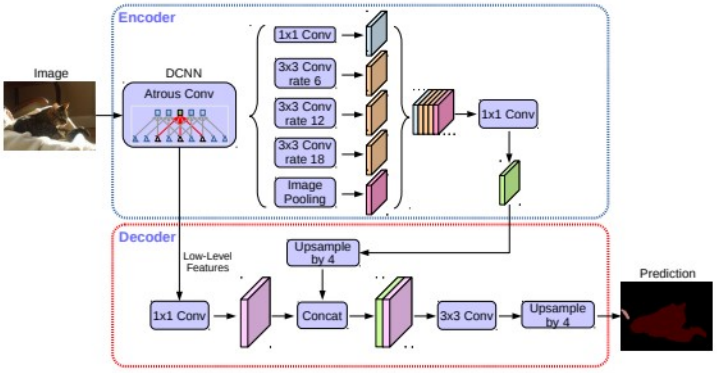

Fig. 3. The DeeplabV3+ architecture (see [4]).

For executing our training, we ended up using Google Colab in order to capitalise on its appropriable GPU support. Regarding our parameter choices, we mainly followed the recommendations given in [4]; as a training batch size we chose 20 in order to prevent running out of memory. We set the count of training steps to 8000, given our dataset size of 2572 and our batch size of 20 resulting in a training of $8000/(2753/20) \approx 58$ epochs.

5 FURTHER PROCESSING

Having our model set up, we needed to focus on using it to turn our raw application dataset into timelines of their respective water surface development. Initially, this was represented in a filesystem-based way in our dataset, as for each water body we had a directory containing the respective screenshots which were annotated the respective year

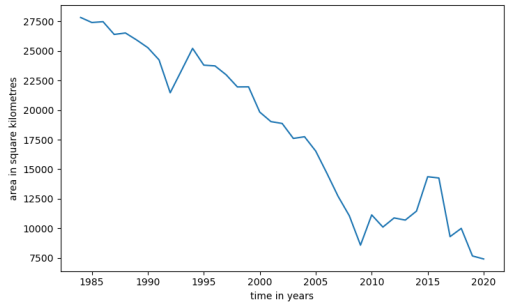

Fig. 4. Our assessed history of the Aral Sea's surface area

via their filename. Thus, we could iterate over our dataset and calculate our model's segmentation mask for each image, converting our dataset into a timeline of segmentation masks.

In order to calculate the size of the water surface area on each image, we combined our estimated total area of an image of a respective water body with the ratio of pixels classified as water on the specific segmentation mask, breaking down the calculation of said area to $A = A_{total} \cdot water\_ratio$. This approach was used in a bulk-assessment for each water body, allowing us to compute a collection of all area sizes, order them by year and write them to a respective CSV file. In the end, our processing resulted in 10 CSV files – one for each water body – with 37 entries each – one per year – containing our calculated water surface size.

6 RESULTS

In the following we will provide an overview of the results of our pipeline. In particular, we will showcase three notable examples of our observed water bodies.

Concerning the Aral Sea, our estimated history of its water surface can be found on Figure 4. For comparison we use a study published by Statista showcasing the water surface area of the Aral Sea in steps of 10 years from 1911 to 2018 [7]. Accordingly, our comparison will not be entirely accurate due to the lack of information in between those years. We find that our system assessed the Aral Sea's water surface area significantly lower than the referenced study shows until 2010. This may partially be related to false negatives in our segmentation as will be further discussed in section 7, as well as inaccuracies with our estimated total area of our scenario. On from 2010 however, our system seems to have assessed the Sea's water surface rather accurately – furthermore the overall trend is the Aral Sea's desiccation is clearly observable as well – except for the local peak around 2015 which again is caused by inaccurate segmentation.

Regarding the Aydar Lake, our estimation can be found on Figure 5 – our comparison is an estimation from an article from 2009's Annual Journal of Hydraulic Engineering ranging from year 1993 to 2004 [15, Fig. 5(b)]. While this again does not cover the entirety of our observed time span, we find that our results are surprisingly close to the estimation in said article.

Lastly, our estimation of our area around southern Sylt seen on Figure 6 should be considered. The results are obviously too noisy for making concise statements about its development, yet the overall trend of the coastal erosion, hence the increasing water surface area, is still fairly observable. It remains to show that our segmentation model definitely is improvable.

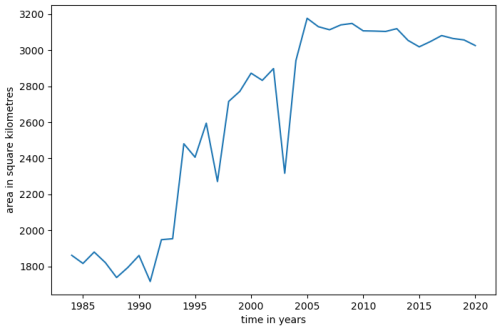

Fig. 5. Our assessed history of the Aydar Lake's surface area

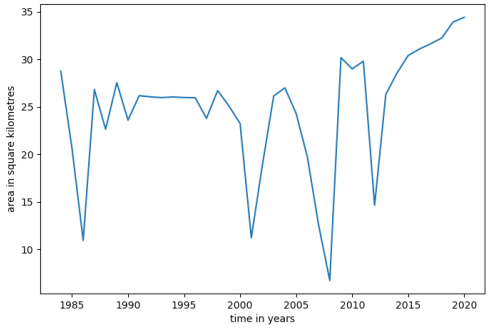

Fig. 6. Our assessed history of the sea surface area around southern Sylt

7 EVALUATION

For the evaluation of our water segmentation, we used an enhanced version of a binary image segmentation evaluation tool initially written during a bachelor's project on waste segmentation on beach images [3]. It can be used to calculate pixel-wise performance metrics such as Accuracy, Precision, Recall, F1-Score and Intersection over Union for a binary image segmentation. Moreover, it can be used to generate visualizations, depicting the quality of a segmentation. Pixels that are classified as water are considered positive, implying the semantics of true positives, etc. in our scenario. As our evaluation tool requires ground truth annotations, we manually annotated two images of each of our selected water bodies, making 20 images in total. Note that this rather low count of images taken into account by our evaluation affects the generalisability of its results; furthermore our manual ground truth annotations are inherently of subjective nature – different people under different circumstances would annotate these images in different ways, allowing for variances in possible evaluation outcomes.

For our evaluation we chose pixelwise Accuracy, Precision, Recall and F1-Score, Intersection over Union as well as the Mean Squared Error of the relative water surface area sizes as our considered metrics. An overview of our evaluation results can be found in table 2 and Figure 7. The evaluation table shows that we reached overall good results with high accuracy and good recall. The scatter plot shows a correlation between low Area-MSE with high F1-Score and high

	Accuracy	Precision	Recall	F1-Score	IoU	Area-MSE
Average	0.876	0.768	0.708	0.722	0.650	0.034
Standard Deviation	0.164	0.362	0.331	0.333	0.332	0.095
Median	0.946	0.988	0.838	0.858	0.752	0.002
Minimum	0.273	4.e-7	2.e-6	7.e-7	0.000	1.e-6
Maximum	0.994	0.999	0.998	0.997	0.994	0.440

Table 2. Tabular summary of our evaluation results.

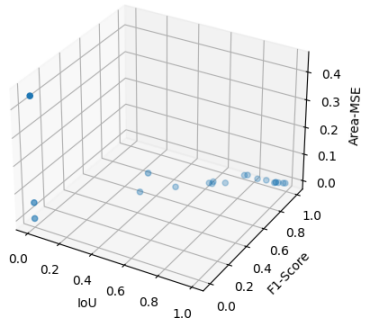

Fig. 7. Scatter plot of our evaluation results.
For visibility's sake, we omitted Accuracy, Precision and Recall for a simple reduction of our results to three dimensions.

IoU. Which makes sense because having accurate prediction results i.e. high IoU should yield low Area-MSE and high F1-Score. Only very few data points have very low IoU scores, most of them are above 0.6. The figure 8 shows the visualizations of the aforementioned evaluation tool. Areas are color coded by red for false positive, green for true positive, blue for false negative and normal pixels for true negative detection of water. Large, color uniform water bodies were almost always classified correctly. One example is the image of Dubai from 1984. The model has difficulties with narrow bodies of water for example on the Aydar lake images. Shallow water that changes the usual water color can be mistaken for land as shown in the picture of Sylt from 2020. The low quality and distortion of the Sylt image from 1984 lead to chunks of false negatives. The evaluation supports that our method for water body segmentation is working.

8 DISCUSSION

Regarding our proposed predictions about the future development of these water bodies, we find that most of our observed scenarios are already subject to human impact and therefore only problematically modelled by natural development. An application of our system on a bigger scale would be advisable in order to observe more less strictly-controlled water bodies in order to obtain more results that are interpretable as natural processes.

A training dataset of greater size and quality would be desirable in order to obtain more accurate results on the segmentation. For our application dataset, more water bodies and especially more annotated ground truth segmentation for our evaluation should be taken into account as well. Concerning our model, there could always be improvement of segmentation found by empirical parameter fine-tuning.

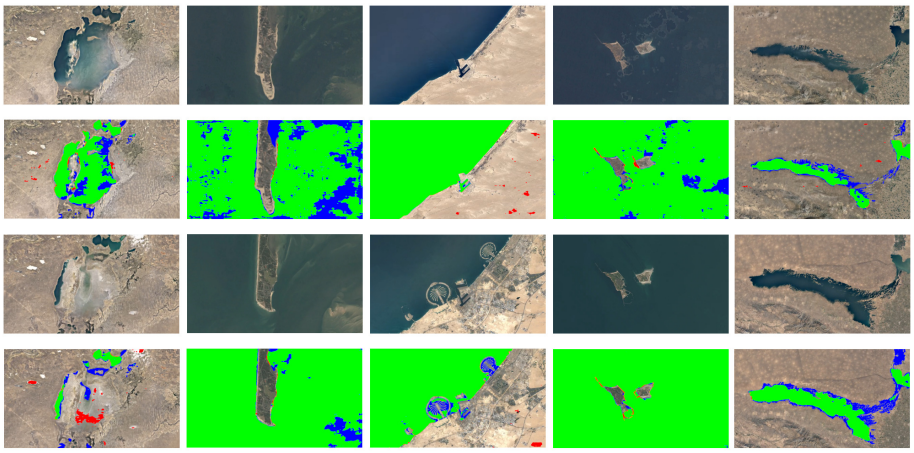

Fig. 8. Evaluation of the masks. The top two images in one column are always from 1984 and the bottom two are from 2020. From left to right Aral sea, Sylt, Dubai, Helgoland, Aydar lake.

For the sake of shortness, we omitted most of our observed water bodies from our results and evaluation. In a larger context, a discussion of more results and an evaluation against other water body segmentation systems would be welcome in order to gain more in-depth knowledge about our segmentation system.

In our evaluation the choice of metrics plays an important role on assessing the quality of our segmentation system. We used a fair amount of pixel comparison-based metrics which give a simple overview over the general trend of the quality of the segmentation, however are purely syntactic and give little information about our system application context. Therefore we also introduced the MSE of relative water surface sizes to have a more semantic perspective on our system. As our system could completely by chance assess the same amount of water as in our ground truth annotation, this metric needs to be taken alongside with the rather syntactic ones in order to gather a rather qualitative view on our segmentation quality. Lastly, the use of other, more complex metrics, which take into account more semantic information about our scenario of water segmentation could be taken into account.

9 CONCLUSION

To come to conclusions on our initial research questions, we have explored a method for water body segmentation based on RGB satellite images which can be used for landscape monitoring using fully convolutional neural networks as provided by our chosen DeepLabV3+ segmentation model. This method shows promising results which are expected to be improved by more extensive and sophisticated training on a larger and higher-quality dataset as well as empirical parameter tuning on our model. For water bodies such as the Aral Sea or the Aydar Lake, we could observe their respective development concerning their respective phenomena. Given a model that would provide better segmentation results on various water bodies, we would expect to obtain more interpretable information.

Predictions about the future development of our observed water bodies have proven to be rather problematic choice is subject to high human impact, rendering models for natural development rather unusable as the given, theoretically arbitrary human measurements would need to be taken into account. Overall we consider our approach as worth exploring, yet as of now inferior to the more extensively researched field of water segmentation on raw satellite images which can capitalise on more information such as infrared values.

REFERENCES

[1] Ulrike Bayr and Oskar Puschmann. 2019. Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network. *Ecological Informatics* 50 (02 2019). https://doi.org/10.1016/j.ecoinf.2019.01.012

[2] Martin Brandt, Compton J. Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, Laurent Kergoat, Ole Mertz, Christian Igel, Fabian Gieseke, Johannes Schöning, Sizhuo Li, Katherine Melocik, Jesse Meyer, Scott Sinno, Eric Romero, Erin Glennie, Amandine Montagu, Morgane Dendoncker, and Rasmus Fensholt. 2020. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* 587 (2020), 78–82. https://doi.org/10.1038/s41586-020-2824-5

[3] Finn Brüggemann, Nico Fuhrberg, Andreas Gärtner, Luca Gee, Dieudonne Fiacre Koulahoue, Alexander Krenz, Aidin Moghinan, Malte Neysters, Michelle Oppermann, Marlow Rischmüller, and Niklas Stockfisch. 2022. *Meeresblick*. Project Report. Universität Bremen. https://www.meeresblick.uni-bremen.de/doc/Bericht_Meeresblick.pdf

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision – ECCV 2018*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 833–851.

[5] Maurício C.R. Cordeiro, Jean-Michel Martinez, and Santiago Peña-Luque. 2021. Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors. *Remote Sensing of Environment* 253 (2021), 112209. https://doi.org/10.1016/j.rse.2020.112209

[6] Francisco Escobar. 2020. *Satellite Images of Water Bodies*. https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies

[7] Interstate Commission for Water Coordination of Central Asia of Central Asia. 2020. *Wasseroberfläche des Aralsees in den Jahren 1911 bis 2018*. https://de.statista.com/statistik/daten/studie/1179776/umfrage/wasseroberflaeche-des-aralsees/

[8] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* (2017). https://doi.org/10.1016/j.rse.2017.06.031

[9] Furkan Isikdogan, Alan C. Bovik, and Paola Passalacqua. 2017. Surface Water Mapping by Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10, 11 (2017), 4909–4918. https://doi.org/10.1109/JSTARS.2017.2735443

[10] Leo F. Isikdogan, Alan Bovik, and Paola Passalacqua. 2020. Seeing Through the Clouds With DeepWaterMap. *IEEE Geoscience and Remote Sensing Letters* 17, 10 (2020), 1662–1666. https://doi.org/10.1109/LGRS.2019.2953261

[11] René Létolle and Monique Mainguet. 1996. *Der Aralsee* (1 ed.). Springer Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-60954-1

[12] European Drought Observatory. 2011. *NDWI: Normalized Difference Water Index*. https://edo.jrc.ec.europa.eu/documents/factsheets/factsheet_ndwi.pdf

[13] Emre Ozelkan. 2019. Water Body Detection Analysis Using NDWI Indices Derived from Landsat-8 OLI. *Polish Journal of Environmental Studies* 29 (08 2019). https://doi.org/10.15244/pjoes/110447

[14] A. Richter, D. Faust, and H.-G. Maas. 2013. Dune cliff erosion and beach width change at the northern and southern spits of Sylt detected with multi-temporal Lidar. *CATENA* 103 (2013), 103–111. https://doi.org/10.1016/j.catena.2011.02.007 Long-term degradation of fragile landscape systems.

[15] Wahyuni S., Oishi S., Sunada K., Toderich K.N., and Gorelkin N.E. 2009. Analysis of water-level fluctuations in Aydarkul-Arnasay-Tuzkan Lake System and its impacts on the surrounding groundwater level. *Annual Journal of Hydraulic Engineering* 53 (February 2009), 37–42. http://www.cawater-info.net/pdf/wahyuni-et-al.pdf

[16] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[17] The GIMP Development Team. 2019-06-12. *GIMP*. https://www.gimp.org

[18] J. Thiede and K. Ahrendt. 2000. *Klimaänderung und Küste – Fallstudie Sylt*. Abschlussbericht. http://www.iczm.de/geomarab.pdf%3F