A Deep Learning Algorithm for Water Body Mapping from Sentinel 2 Data

Gordana Jakovljević[1,], Miro Govedarica[2]*

[1] Faculty of Architecture, Civil Engineering and Geodesy, University of Banja Luka, Bosnia and Herzegovina, gordana.jakovljevic@aggf.unibl.org

[2] Faculty of Technical Science, University of Novi Sad, Serbia, miro@uns.ac.rs

* corresponding author

doi: 10.5281/zenodo.11657519

Abstract: Water is vital for the life on the Earth. Human health, food security, and economic growth are all water dependent. Monitoring of water bodies and their spatial changes is crucial for understanding the impact of human activities and climate changes on aquatic ecosystems. Remote sensing data provides large amounts of data that have been extensively used for monitoring water bodies, geometry, topology, associated attributes, and their changes. However, water body delineation is challenging due to sensor limitations, cloud presence, and atmospheric conditions. This paper presents a novel approach leveraging Convolutional Neural Networks (CNNs) for extracting water bodies from Sentinel-2 imagery. The efficacy of the proposed algorithm is rigorously evaluated across heterogeneous terrains encompassing diverse riverine and lacustrine features. Key metrics such as overall accuracy, F1 score, precision, and recall are employed to quantitatively assess algorithmic performance. Our findings underscore the promising potential of deep learning techniques in accurately delineating water bodies and monitoring their dynamic behaviour within intricate environmental contexts at local, regional, and global scales.

Keywords: inland water body; Sentinel-2; CNN; geometry.

1 Introduction

Water is essential for life, supporting humans, animals, plants, and entire ecosystems. Recently, the effects of global climate change and human activities on the spatial and temporal variations in water quality and quantity have gained increasing attention. These changes in surface water bodies impact agricultural and industrial production, ecological balance, environmental conditions, and food and health safety. Accurate and rapid information about the spatial distribution, persistence, and quality of surface water is crucial for sustainable water usage and protection from feature degradation. Remote sensing covers large geographic areas at various spatial, spectral, and temporal resolutions, and provides extensive data used for analysing surface water bodies and their dynamics. Satellite images are particularly valuable for obtaining information about water bodies in remote, inaccessible, extremely large, or hazardous areas, such as during floods. Over the past three decades, multi-source satellite imagery of different resolutions has been utilized for extracting data on surface water bodies.

Until now, a variety of algorithms have been used for water body delineation from satellite images, ranging from simple spectral indices to deep learning models. Jiang et al. (2021) have been using the Sentinel-2 Water Index and OTSU thresholding for the detection of water bodies. Similarly, Lekhak et al. (2023) combined the spectral indices and slop information with the threshold method to provide more accurate water body mapping. Although threshold-based methods can be used to delineate water pixels, they yield inconsistent results due to high spatial and temporal variation among regions. A wide range of Machine Learning (ML) algorithms, such as Supported Vector Machine (SVM) and Decision Trees (DT) has been used to address that limitation. Acharya et al. (2019) tested the performance of 6 different ML algorithms for surface water extraction in Landsat 8 images. The Random Forest (RF) provided the highest accuracy followed by Gradient Boosted Machines (GBM). Jakovljevic et al. (2018) was adopted the SVM for water body extraction resulting in a kappa coefficient of 0.89 for Sentinel 2 images. Although ML algorithms can provide high accuracy, they are mainly based on spectral information of the training samples, which can lead to the misclassification of surfaces with similar spectral signatures such as dark areas (topographic and cloud shadows), buildings, snow, etc (Kumar et al. 2014). Moreover, the performance of ML algorithms mainly depends on expert-designed features, which limits generalization ability.

In recent years, deep learning algorithms, particularly Convolution Neural Networks (CNN), have been widely used for image classification (Kim et al. 2022), object detection (Galvez et al. 2018), and semantic segmentation (Alam et al. 2021). The deep learning models have been used for the classification and change detection of remote sensing data. Li et al. (2019) adopted the Fully Connected Network (FCN) model to extract water bodies from Very High Resolution (VHR) images and significantly outperform the indices and ML-based methods. Similarly, Erdem et al. (2021) used U-Net architecture for automatic shoreline extraction from Landsat 8 images with high accuracy (F1: 99.79%).

This paper presents an automatic method for the identification of inland water bodies of different sizes and shapes in complex environment conditions based on Sentinel 2 images.

2 Study area

The Republic of Serbia is in Southeast Europe, covering part of the Pannonian Plain and the Central and Western Balkan Peninsula (Figure 1). Serbia covers 88,361 km^2 of which 56.8 % is cropland, and 36.6 % is covered by forest (OECD 2024). The almost entire territory of Serbia belongs to the Danube (Black Sea) basin. The Danube is the lagers river in Serbia and the second largest river basin in Europe, covering 801,463 km^2 over 19 countries and more than 81 million people (ICPDR 2024). The tributes of the Danube in Serbia are Sava, Tisa, Drina, and Great Morava (Morava) (Figure 1).

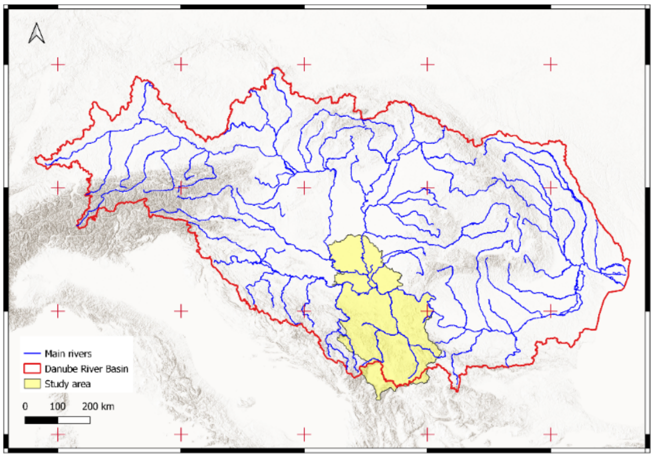

Figure 1. Study area.

3 Materials and methods

Semantic segmentation aims to assign the set of predefined class labels to each pixel in the image (Janai et al. 2020). According to the structure, CNN models for semantic segmentation can be divided into encoder-decoder and spatial pyramid pooling. The encoder-decoder consists of an encoder function that converts the input data into feature maps by using convolution, activation, and pooling layer and a decoder function that up-samples the encoder features maps and converts them to segmentation results. The U-Net architecture (Ronneberger et al. 2015) (Figure 2a) consists of an encoder that captures contextual information and a symmetrical decoder that restores spatial resolution. The encoder followed the typical architecture of CNN (convolution, activation, max pooling), progressively decreasing feature maps resolution, and increasing the number of feature channels per encoder at the same time. The skip connection is used to connect resolution feature maps from the encoder with a corresponding up-sampled output of the decoder, which allows the network to learn back relevant features that are lost after pooling operations and to predict more precise outputs based on that information. In this paper, the ResNet 50 (He et al. 2016) was used as an encoder part of the network. The architecture of ResNet 50 has four stages. The network performs the initial convolution and max pooling using 7x7 and 3x3 kernel sizes, respectively. Afterward, stages 1, 2, 3, and 4 consist of 3, 4, 6, 3 ResNet building blocks (Figure 2b). As the network progress from one stage to another, the feature map resolution is reduced by 2 in terms of height and width while the number of feature channels is doubled. The decoder is fully symmetrical to the encoder, and it is used to restore feature map resolution enabling precise localization. Each step in the decoder consists of 2x2 up sampling that halves the number of feature channels concatenation with the corresponding feature map from the encoder path, followed by two 3x3 convolutions, BN, and ReLU activation functions. In the final layer, a 1 × 1 convolution with the Sigmoid activation function is used to predict the probability of a pixel being assigned to a water or non-water class.

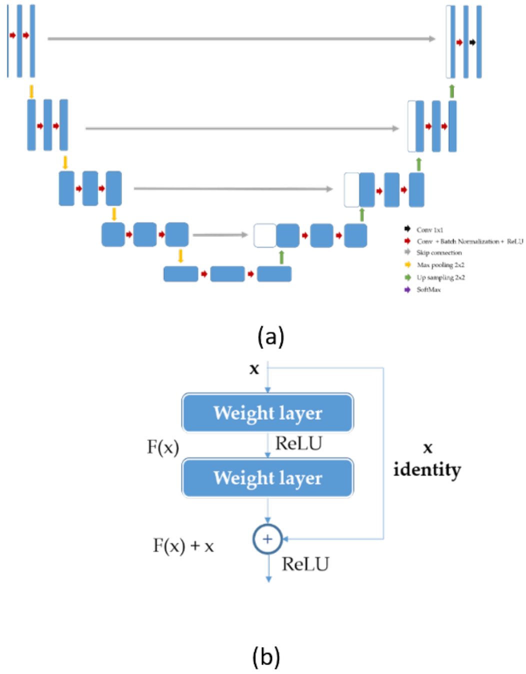

Figure 2. (a) UNet architecture for semantic segmentation, (b) ResNet 50 building block.

The performance of a deep learning network is strongly dependent on a large amount of training data, which is needed to understand hidden patterns of data. Transfer learning has been widely used for solving an insufficient data problem (Castelluccio et al. 2015). Fine-tuning of existing networks that are trained on large datasets such as ImageNet is most used in practice (Penatti et al. 2015). ImageNet is a large and diverse dataset with more than 14 million images labelled into 1000 classes.

In addition to limited size, datasets for the classification of inland water bodies are highly imbalanced since most pixels represent non-water classes. To prevent imbalance learning, enlarge dataset, and reduce over-fitting the data augmentation was used. In this research, clipping, rotating, flipping, and translating were used.

To assess the accuracy of the implemented model the recall, precision, F1-score, and estimate of Kappa coefficient were calculated, as shown in Foody (2008).

Implementation: Dataset was split into 80% for training and 20% for validation. The network is fine-tuned on the dataset created during preprocessing. The cross-entropy and Stochastic Gradient Descent were selected as loss function and optimization algorithm. The GPU limited the batch size, and it was chosen as big as possible for each network. The models were implemented in the Python 3 programming language by using artificial intelligence libraries such as PyTorch, TensorFlow, Keras, and Matplotlib. The training of the networks was done using the publicly available cloud platform Collaboratory (Google Colab).

4 Results

Accuracy assessment of the proposed model for Sentinel 2 is based on 861 and 13600 image patches for validation and test phases. The results of the accuracy assessment are presented in Table 1.

Table 1. Results of accuracy assessment for water body extraction from Sentinel 2.

Phase	Precision	Recall	F1-score	Kappa
Validation	0.90	0.95	0.92	0.92
Test	0.81	0.99	0.89	0.89

The visual inspection shows (Figure 3) that detected wetlands and channels are more completed comparing to masks, which also decrease the precision.

The visual inspection of results (Figure 3) shows that the water bodies extracted from the satellite images followed a similar pattern with true data. As can be seen from the figure, the algorithm can detect lakes, large rivers, and even small ponds or reservoirs with high accuracy (Figure 3a, Figure 3b, Figure 3g). As expected, the lowest accuracy is obtained for small and narrow streams. The small water bodies were overestimated (Figure 3d, Figure 3e, Figure 3f) due to mixed pixels producing lower precision.

5 Discussion

Motivated by recent success in deep learning, this research focused on using those methods to improve the water body mapping from satellite images. As presented results indicated, the proposed approach provides water body detection in the complex environment from optical with consistently high F1-score and kappa coefficient despite varying topology, land-use/land cover, and atmospheric conditions. Similarly, Yan et al. (2022) has reported an F1-score of 0.9 for mapping lakes from Sentinel-2 by using the Unet architecture.

The difference between F1 score during the validation and test phase was 3% indicating the algorithm's high generalization ability. Therefore, it can be used for automatic water body detection from different areas without manual intervention. These results are also confirmed by Billson et al. (2023). It is observed that during the test phase recall value increases while the precision decreases meaning that on the one hand algorithm is more secure that the pixel labelled as water represents the water body in the real world but on the other hand, it includes more non-water pixels in water class. Although the proposed approach produces stable results on large and medium size water bodies, the extraction of narrow streets and small lakes remains challenging.

6 Conclusions

In this paper, the methodology for automatic inland water body mapping based on UNet architecture and Sentinel 2 images have been proposed. The Kappa coefficient, precision, recall, and F1-score were calculated to evaluate the performance. The high value and visual inspection show that ResUNet 50 is not sensitive to low albedo surfaces such as built-up areas, roads, or shadows, which is one of the primary sources of errors during water body extraction from remote sensing data. Comparison of validation and test accuracy (F1: 0.89 vs 0.92) indicates great generalization ability and the possibility to apply the algorithm for automatic water body detection over different areas.

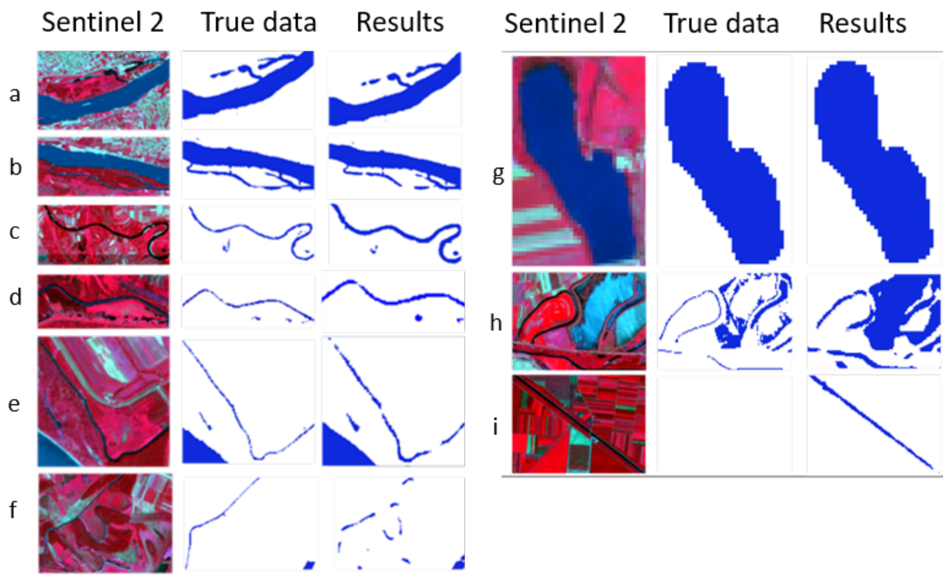

Figure 3. Visual comparison of extracted water bodies for different water body types (a), (b) large river (> 400 m width), (c) medium river (width around 100 m), (d), (e), (f) small rivers (width between 10-35 m), (g) lake, (h) wetland, (i) artificial channels (Jakovljevic 2020).

7 References

Acharya, T. D., Subedi, A., Ha Lee, D., 2019. Evaluation of Machine Learning Algorithms for Surface Water Extraction in a Landsat 8 Scene of Nepal. Sensors 19 (12), 2769.

Alam, M., Wang, J.-F., Guangpei, C., Yunrong, L., Chen, Y., 2021. Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images. Mobile Networks and Applications, 200-215.

Billson, J., Islam, M., Sun, X., Cheng, I., 2023. Water Body Extraction from Sentinel-2 Imagery with Deep Convolutional Networks and Pixelwise Category Transplantation. Remote Sensing 15 (5), 1253.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks.

Erdem, F., Bayram, B., Bakirman, T., Vayrak, O. C., Akpinar, B., 2021. An ensemble deep learning-based shoreline segmentation approach (WaterNet) from Landsat 8 OLI images. Advances in Space Research 67 (3), 964-974.

Foody, G. M., 2008. Harshness in image classification accuracy assessment. International Journal of Remote Sensing 29 (11), 3137-3158.

Galvez, R. L., Bandala, A. A., Dadios, E. P., Vicerra, R., Maningo, J., 2018. Object Detection Using Convolutional Neural Networks. TENCON 2018 - 2018 IEEE Region 10 Conference, 2023-2027, Jeju, Korea.

He, K., Zhang, X., Ren, S., 2016. Deep Residual Learning for Image Recognition. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).

ICPDR. (2024, 05 10). Countries of the Danube River Basin. Retrieved from ICPDR: https://www.icpdr.org/main/danube-basin/countries-danube-river-basin. (Accessed 13 June 2024)

Jakovljevic, G., 2020. Multidimensional model of use remote sensing data and geospatial services in water management according to INSPIRE and WFD specification, PhD thesis. Novi Sad: Faculty of Technical Science, University of Novi Sad.

Jakovljevic, G., Govedarica, M., Alvarez-Taboada, F., 2018. Waterbody mapping: a comparison of remotely sensed and GIS open data sources. International Journal of Remote Sensing 40 (8), 2936-2964.

Janai, J., Guney, F., Behl, A., Geiger, A., 2020. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. Foundations and Trends in Computer Graphics and Vision 12 (1-3), 1-308.

Jiang, W., Ni, Y., Pang, Z., Li, X., Ju, H., He, G., Qin, X., 2021. An Effective Water Body Extraction Method with New Water Index for Sentinel-2 Imagery. Water 13 (12), 1647.

Kim, H., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., Ganslandt, T., 2022. Transfer learning for medical image classification: a literature review. BMC Medical Imaging 22 (1), 69.

Kumar, L., Sinha, P., Taylor, S., 2014. Improving image classification in a complex wetland ecosystem through image fusion techniques. Journal of Applied Remote Sensing 8 (1), 083616.

Lekhak, K., Rai, P., Budha, P. B., 2023. Extraction of water bodies from Sentinel-2 imagery in the foothills of Nepal Himalaya. International Journal of Environment and Geoinformatics 10 (2), 70-81.

Li, L., Yan, Z., Shen, Q., Cheng, G., Gao, L., Zhang, B., 2019. Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks. Remote Sensing 11 (10), 1162.

OECD. (2024, 05 10). Retrieved from Land cover in countries and regions: <https://stats.oecd.org/Index.aspx?DataSetCode=LAND_COVER. (Accessed 13 June 2024)

Penatti, O. A., Nogueira, K., dos Santos, J. A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scene domains? 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Boston: IEEE.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015.

Yan, K., Li, J., Zhao, H., Wang, C., Hong, D., Du, Y., Wang, S., 2022. Deep Learning-Based Automatic Extraction of Cyanobacterial Blooms from Sentinel-2 MSI Satellite Data. Remote Sensing 14 (19), 4763.