Enhancing Water Body Detection in Satellite Imagery Using U-Net Models

Jiongyi Li

Electrical Engineering and Computer Science, Pennsylvania State University, PA, 16802, USA
Jzl6831@psu.edu

Abstract. Precise and efficient detection of water bodies in satellite pictures is essential for diverse applications, like environmental surveillance, urban development, and disaster response. This study investigates the effectiveness of utilizing the U-shaped network (U-Net) models with input shapes of 128x128 and 256x256 to detect water bodies in satellite photos acquired from the Sentinel-2 Satellite. This research aims to address the dual challenge of recognizing global features in images while also capturing detailed characteristics, such as the boundaries of water bodies. It observes that both models achieve a commendable accuracy of approximately 0.8, accompanied by a modest loss of about 0.3. Notably, the model with a smaller input shape demonstrates a faster convergence during training but exhibits slightly diminished delineation of water body edges compared to its counterpart with a larger input shape. These findings contribute valuable insights into the optimization of water body detection algorithms, offering avenues for both broad-scale previews and fine-scale segmentation in satellite imagery analysis.

Keywords: Water Body Detection, Satellite Imagery, U-Net.

1 Introduction

Water body segmentation is a general name for technologies that know the space distribution of water on the surface. Since humans have the ability to monitor the surface, acknowledging the distribution of surface water is an important topic. To achieve that, scientists developed many methods, like threshold segmentation, before importing machine learning [1]. However, because of the significant difference between the data from the dataset and the real world, the well-developed, successfully matched methods are incompatible with data from the real world [2]. So, there is still ample space for improvement in the accuracy and robustness of the method for water body detection.

Recently, other researchers have tried different methods for the segmentation. For research that used normalized difference water index (NDWI) and normalized difference vegetation index (NDVI), both of which are threshold methods, to search the status of plant life and the stress caused by drought in an Indian district and confidently find the significance of the impact from varying patterns of rainfall on the severity of meteorological drought [3]. Then, in the medical direction, researchers are re-

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_87

searching making the model with U-shaped network (U-Net) structure in segmentation better by changing the typical U-Net structure to two directions. In the first direction, researchers let the U-Net model work with other models in sequence or parallel to detect their specific targets [4, 5]. Their results found the potential for higher accuracy and lower loss while there are outstanding requirements for computation resources. Then, in the second direction, researchers edit the typical U-Net model and slightly increase the model complexity [6, 7]. Their results show that their models have greatly improved accuracy and other points, with some cost of increasing computation [6, 7]. Then, going back to the machine learning research for water body segmentation, researchers use the concept of convolutional neural network (CNN) to build a new model. Then, the new model is used to find the distribution of the water body in satellite photos, and their results prove the meaning of bringing machine learning to water body segmentation [8, 9].

Improving the precision and efficacy of water body detection by applying advanced image segmentation techniques to satellite imagery is the main objective of this study. Initially, the data undergoes preprocessing, wherein both the image and its corresponding mask are resized to a standardized format suitable for subsequent model training. Subsequently, the U-Net model is employed for training, leveraging its effectiveness in medical image segmentation to address the spatial complexities inherent in satellite imagery. Finally, the trained model is rigorously evaluated using cross-validation and confusion matrices. The experimental findings unequivocally demonstrate the approach's remarkable accuracy and high confidence levels. The practical importance of this research lies in its ability to bolster sustainable water management practices and enhance disaster preparedness through the provision of precise, real-time insights into water body dynamics across diverse geographic landscapes.

2 Methodology

2.1 Dataset Description and Preprocessing

The dataset utilized consists of a compilation of images of water bodies captured by the Sentinel-2 Satellite on Kaggle [10]. The dataset includes 2841 pieces of images and the corresponding masks of water in binary. Masks are generated by NDWI, which is a method compatible with satellite images [1]. Here, as Fig. 1 shows below, for using U-Net, images and corresponding masks are stretched to square images. For comparing the model's performance with different input sizes, the shape of images and their masks are stretched to 128x128 or 256x256. Also, a train dataset and a validation dataset are obtained by splitting the dataset with a ratio of 8:2, which means the train dataset sources from 80% of the original dataset, and the validation dataset is generated by using 20% of the original.

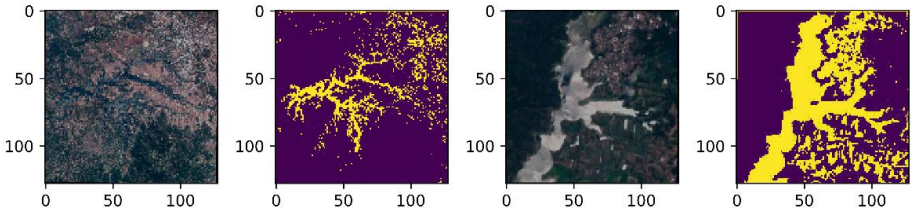

Fig. 1. Examples of stretched images and corresponding masks.

2.2 Proposed Approach

When applying the U-Net model, to search whether the model is suitable for recognizing the location of the water body in the graph, the same train dataset and validation dataset generated in a preprocessing dataset, whose shape is uniformed to 128x128 or 256x256, were used previously, and the train dataset is 80% of the original dataset. Then, as Fig. 2 shows on the next page, the train dataset, which includes a graph matrix in shape 2272x128x128x3 and mask matrix in shape 2272x256x256x1, is inputted to U-Net model, which is built with input shape 128x128x3 and output shape 128x128x1, and get the history of the accuracy and loss of 100 training times. Then, the validation dataset, which includes a graph matrix in shape 569x128x128x3 and a mask matrix in shape 569x128x128x1, is inputted into the model and gets the history of the accuracy and loss of the test after the training. Similarly, the train dataset, which includes a graph matrix in shape 2272x256x256x3 and mask matrix in shape 2272x256x256x1, and the validation dataset, which includes a graph matrix in shape 569x256x256x3 and mask matrix in shape 569x256x256x1, are inputted into the model and get the history of the accuracy and loss of 100 training times. Here, the input shape of the model is 256x256x3 and an output shape of the model is 256x256x1.

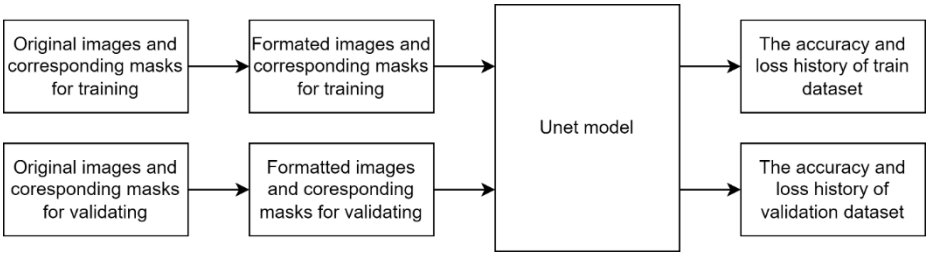

Fig. 2. The pipeline of the model.

U-Net. The U-Net model can be seen as a specific CNN mainly used for completing image segmentation tasks. Then, the model can capture both local and global context and tolerate the shortage of sample images by down-sampling and up-sampling the input. So, that is the reason the model is needed to identify the water body in the satellite images here. As Fig. 3 shows below, there are generally multiple layers in differ-

ent shapes in a U-Net model. The model first keeps two convolutional layers with input shape, then copies the matrix from the last convolutional layer for future use and uses a 2×2 max pool to reduce the original matrix to half the size in width and height, which can make the model capture some more global features. Then, the previous behavior is done similarly again, and the shape of the matrix is one-fourth of the original input shape. By this, the model can capture the most global features. Then, after the matrix passes a convolutional layer and has multiple feature layers, as the bottom of the model in Fig. 3, it passes a convolutional layer and has one feature. Then, the matrix is up-sampled by 2×2 up-sampling to put global features back to a more detailed matrix, and the matrix is combined with the matrix with half the width and height of the input shape, which got from previous convolutional layers. Then, the up-sampling and passing convolutional layers are done similarly again, as shown on the right side of each model image in Fig. 3. Here, the same as the right matrix at the bottom of the model, the matrix with the original input shape and multiple feature layers passes a convolutional layer and gets a matrix with the original input shape and one feature layer, which is the output layer of the model. The model generates a matrix with the width and height of the input shape and one layer to represent the probability that a pixel is a part of the water body.

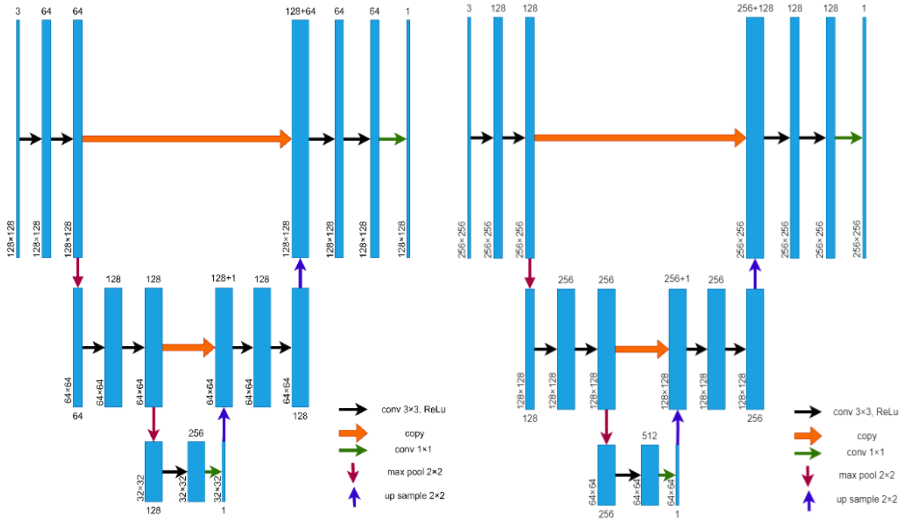

Fig. 3. Shape of U-Net models with different input shapes.

Loss Function. For the model is used for highlighting the water bodies from the satellite images, the model actually is classifying whether a pixel in the image is a part of the water body. So, a loss function that is suitable for binary classification applications is needed, and the cross-entropy loss function, whose function is shown below, is found to be the most suitable loss function here.

$$H(m,n) = -\sum_{x \in X} m(x)\, log(\, n(x)) \tag{1}$$

In the function, “m” and “n” are two types, and the function searches the match level between the two types. So, for the model here, the actual equation, which is

shown below, is searching the match level between the "y," which means whether the water body includes the pixel, and the "q," which is the predicted probability of whether the water body includes the pixel, for a pixel in pixels. So, the loss function can penalize incorrect identification with a logarithmic function, and small deviations from the true answer are not punished as harshly as large deviations. By this feature, the function can be directly used as a gradient descent method for optimizing the model and leads to training the model effectively and aligning the predicted result with the correct result closely.

$$H(y,q) = -\sum_{pixel \in pixels} \begin{matrix}[y(pixel) \cdot log(\, q(pixel)) + \\ (1 - y(pixel)) \cdot log(\, 1 - y(pixel))]\end{matrix} \tag{2}$$

2.3 Implementation Details

For repeating the research, information about the system and some arguments is important. First, the pipelined models, which have different input and output shapes, are running over Python 3.9 and TensorFlow 2.6 in Windows 11 on a machine with NVIDIA RTX 3090. Second, the argument, batch size, of models is set as 16, which means the model is trained with 16 data at one time. Third, the argument, epoch, of models are set as 100, which means models are trained 100 times. Fourth, models use the Adam optimizer as their optimizer for its high computational efficiency, low memory requirements, less sensitivity over most hyperparameters, and faster convergence. Overall, the Adam optimizer is suitable for this research.

3 Results and Discussion

As Fig. 4 shows on the next page, when the U-Net model is built with input shape 128x128x3, the accuracy of recognizing the water body in the train dataset and validation dataset is increased quickly at first few pieces of training, then slowly increased after the accuracy for both datasets is over 0.8. Then, after training 50 times, the accuracy for the train dataset is increased a little more aggressively, while the accuracy for the validation dataset falls slightly and keeps between 0.8 and 0.825. Generally, the overfit can be observed after 40 times of training over the U-Net model when its input shape is 128x128x3.

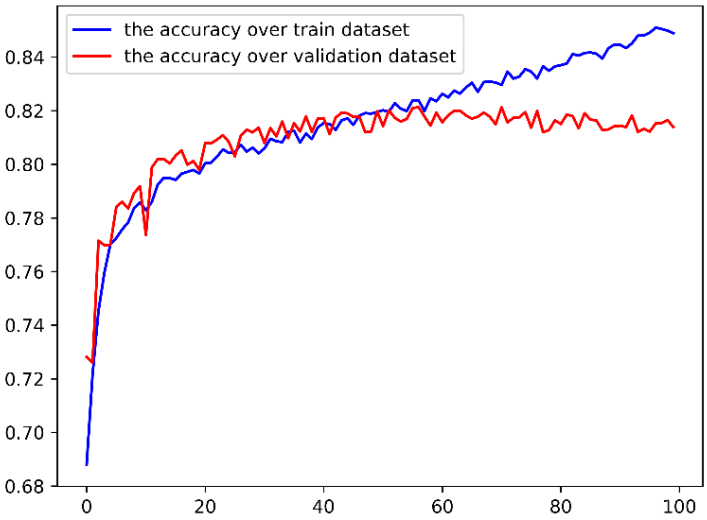

Fig. 4. The accuracy of the model, which sets 128x128x3 as its input shape and 128x128x1 as its output shape.

Then, as Fig. 5 shows on the next page, when the U-Net model is built with input shape of 128x128x3, the loss of the model over the train dataset and validation dataset decreased fast in the first few trainings. Then, the loss over two datasets is decreased simultaneously slowly to between 0.25 and 0.30 until about 30 times of training. After that training times, the loss over the train dataset keeps decreasing toward 0 and even faster after 50 times. At the same time, the loss over the validation dataset increased slightly, approaching 0.35. For the loss value of the model, the overfit can be observed after about 30 times of training.

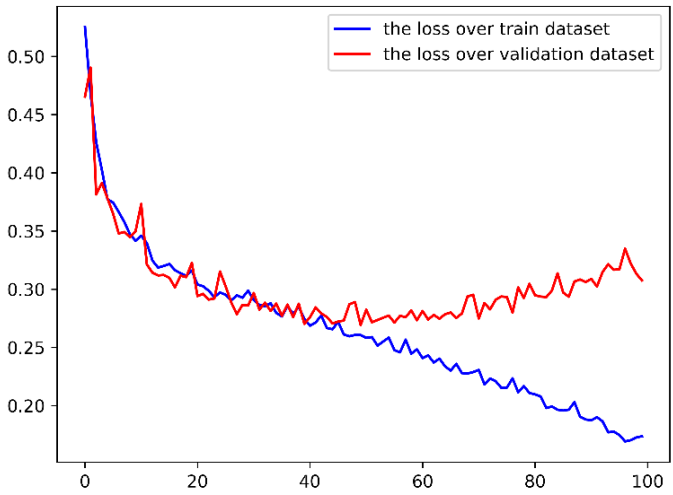

Fig. 5. The loss of the model which sets 128x128x3 as its input shape and 128x128x1 as its output shape.

For the accuracy of the U-Net model, when its input shape is 256x256x3, as Fig. 6 shows below, the accuracy over both the test dataset and validation dataset is increased quickly at the first few training times and slowly increases together then. After training about 70 times, the U-Net model overfits the train dataset, for the accuracy over the train dataset keeps increasing while the accuracy over the validation dataset keeps around 0.8.

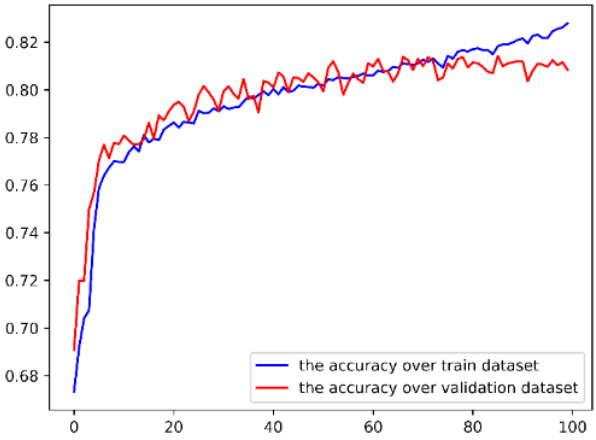

Fig. 6. The accuracy of the model when it sets 256x256x3 as its input shape and 256x256x1 as its output shape.

For the loss of the U-Net model with input shape 256x256x3, as Fig. 7 shows below, the loss over both the train dataset and validation dataset decreased quickly in the first few trainings, then slowed down at the same time until around 60 training times. After around 70 training times, the overfit can be observed, for the loss over the train dataset keeps decreasing steadily while the loss over the validation dataset is increased at a tiny rate. Also, before the overfit is observed, the accuracy over datasets is around 0.8.

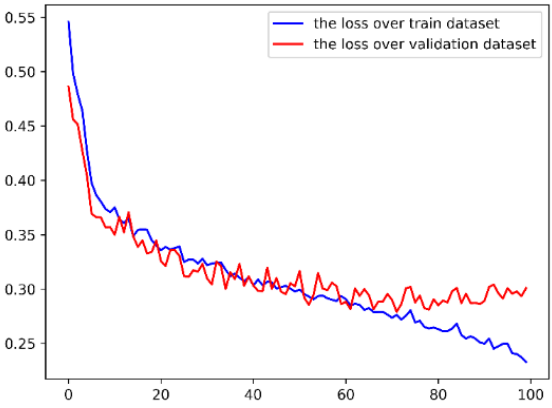

Fig. 7. The loss of the model when it sets 256x256x3 as its input shape and 256x256x1 as its output shape.

Overall, the overfit can be observed in shorter training times in both accuracy and loss when the U-Net model is built with an input shape of 128x128x3, compared to the model with an input shape of 256x256x3. Then, before the overfit is observed, the accuracy and loss of the model with different input shapes are similar.

4 Conclusion

In this study, the author investigates the efficacy of the U-Net model in locating water bodies in satellite imagery. The graphs and associated masks were resized to dimensions of 128x128 or 256x256, facilitating preprocessing for subsequent training. The preprocessed data were then utilized for training the U-Net model with input shapes 128x128x3 or 256x256x3. The findings indicate that the accuracy of the trained models across different input and output shapes consistently approached 0.8 prior to overfitting, while the loss remained around 0.3. Interestingly, the U-Net model with an input shape of 128x128x3 required less training to reach the overfitting threshold compared to its counterpart with an input shape of 256x256x3. This suggests that future improvements in model architecture or preprocessing methodologies could enhance water body detection in satellite imagery. Specifically, enhancing the model structure by incorporating additional max-pooling and up-sampling layers may afford a more comprehensive view of the data, thus further refining the model's performance.

References

1. Cao, Z., Feng, X., Hu, M., Li, J., Ma, R., Xiong, J., Xue, K.: Satellite Detection of Surface Water Extent: A Review of Methodology. Water, 14(7), 1148 (2022).
2. Fang, H., Feng, J., Guan, L., Schaefer, G., Xu, Z., Yuan, K.: Deep-Learning-Based multispectral satellite image segmentation for water body detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 7422–7434 (2021).
3. Jagtap, M.P., Khatri, N., Madan, H., Patil, P., Patodia, T., Vadduri, A.A.: Exploration and advancement of NDDI leveraging NDVI and NDWI in Indian semi-arid regions: A remote sensing-based study. Case Studies in Chemical and Environmental Engineering, 9, 100573 (2024).
4. Agrawal, P., Hooda, N., Katal, N.: Segmentation and classification of brain tumor using 3D-UNet deep neural networks. International Journal of Cognitive Computing in Engineering, 3, 199–210 (2022).
5. Geng, L., Liu, B., Liu, Y., Xiao, Z., Zhang, F.: Segmentation of lung nodules using improved 3D-UNET neural network. Symmetry, 12(11), 1787 (2020).
6. Cheng, C., Le, M., Liu, D., Nguyen, T., Tran, S.: TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation. Healthcare, 9(1), 54 (2021).
7. Cai, S., Chen, G., Lui, H., Tian, Y., Wu, Y., Zeng, H.: Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quantitative Imaging in Medicine and Surgery, 10(6), 1275–1285 (2020).

8. Fang, H., Feng, J., Guan, L., Schaefer, G., Xu, Z., Yuan, K.: Deep-Learning-Based multi-spectral satellite image segmentation for water body detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 7422–7434 (2021).
9. Bai, Y., Liu, X., Mas, E., Wu, W., Yang, H., Yang, Z., Yu, J., Zhao, B.: Koshimura, S.: Enhancement of detecting permanent water and temporary water in flood disasters by fusing Sentinel-1 and Sentinel-2 imagery using deep learning algorithms: Demonstration of SEN1FloodS11 benchmark datasets. Remote Sensing, 13(11), 2220 (2021).
10. Satellite images of water bodies. https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies/data, last accessed 2024/2/3.