航測及遙測學刊 第二十七卷 第 4 期 第 237-246 頁 民國 111 年 12 月
Journal of Photogrammetry and Remote Sensing
Volume 27, No.4, 2022, pp. 237-246
DOI：10.6574/JPRS.202212_27(4).0003

「2022 福衛五號衛星影像應用小論文競賽」優良論文
2022 FORMOSAT-5 Satellite Imagery Application Essay Competition Excellent Paper

Extraction of Water Bodies from Satellite Imagery Using AI Techniques

Daniel-Wilson Lin [1*] Colo Chang [2]

Abstract

Water body segmentation from remote sensing imagery is essential for monitoring and protecting water resources, as well as for assessing the risks of disasters such as flooding. However, traditional index-based approaches to water body identification have significant limitations. In this study, we applied and trained a convolutional neural network (CNN) called U-Net on FORMOSAT-5 imagery of the greater Tainan City area to identify water bodies. The experimental results of the U-Net model were compared with the Normalized Difference Water Index (NDWI) and convincingly showed that the U-Net model had achieved significantly better water body detection performance.

Keywords: FORMOSAT-5, NDWI, U-Net, Water Segmentation, Convolutional Neural Network (CNN)

1. Introduction

Water is a vital resource for sustainable societies and ecosystems alike. With growing global trends of extreme weather events such as droughts and typhoons due to climate change, water security has become an increasingly relevant point of discussion for societies. Achieving an accurate and insightful analysis of water systems is significant in providing the necessary data for water resource investigation, flood monitoring, wetland protection, disaster prevention/mitigation, as well as urban planning.

Many studies have been made on extracting water from satellite imagery, with many well-known methods like the Normalized Difference Water Index (NDWI) (McFeeters, 1996), modified Normalized Difference Water Index (MNDWI) (Xu, 2005), and Automated Water Extraction Index (AWEI) (Feyisa *et al*., 2014). Because of their simplicity and convenience, these different water indices are the most commonly used method for water identification. While they perform relatively well on well-controlled datasets, they are less useful for water body detection in real-world conditions.

This study is aimed at exploring the usefulness of applying CNNs in the identification of water bodies in satellite imagery when compared to traditional index-based identification. In this study, satellite imagery obtained from the FORMOSAT-5 is used to validate the use of CNNs in the prediction of water bodies.

In this paper, the CNN U-Net is applied to extract water bodies from FORMOSAT-5 Imagery. The FORMOSAT-5 RGB and near-infrared (NIR) bands were used to train and test the model. The experimental results show significant improvements compared to traditional NDWI based approach, with a greater area under the Receiver Operating Characteristic (ROC) curve. For certain studies on phenomena such as the Urban Heat Island Effect, differentiation between water and other surfaces becomes increasingly significant. Because roads, buildings and water bodies have drastically different levels of heat retention, the high error rate of the NDWI makes it a sub-optimal option for conducting studies in the area. The main contributions of this research are as follows:

(1) To provide a comparative assessment between deep learning water identification and traditional methods of water identification.

(2) To offer researchers and policy makers a new method to broaden the tools available for water body identification.

2. Related Work

2.1 Traditional Index Based Water Detection

The most common method of water identification using satellite imagery is through the use of the

[1] Senior Student, International Bilingual School at Tainan Science Park, National Nanke International Experimental High School
[2] Mathematics Teacher, International Bilingual School at Tainan Science Park, National Nanke International Experimental High School
[*] Corresponding Author, E-mail: danielwilsonlin@gmail.com

Received Date: Oct. 03, 2022
Revised Date: Oct. 19, 2022
Accepted Date: Nov. 01, 2022

Normalized Difference Water Index, or the NDWI, proposed by McFeeters (1996). The NDWI is computed based on the reflectivity of green and NIR bands. A threshold is then used to determine whether a point on the image is covered with water. The calculation for the NDWI is as follows:

$$\text{NDWI} = \frac{X_{green} - X_{NIR}}{X_{green} + X_{NIR}} \quad \text{.....(1)}$$

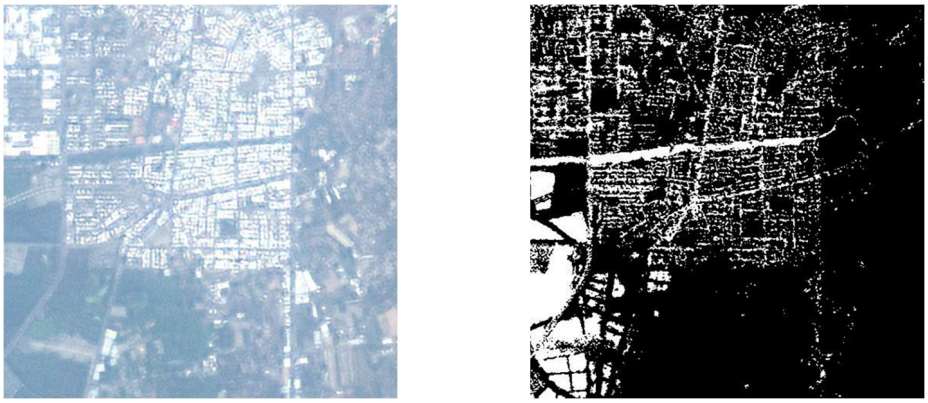
where X_{green} represents the reflectance of the green band and X_{NIR} represents the reflectance of the NIR band. Given a threshold at zero, a positive NDWI value should indicate that the ground is covered with water, a negative NDWI should indicate vegetation coverage, and a value close to zero should indicate rocks and bare soil. The threshold itself is not always zero and is subject to change due to various influences such as the turbidity of the water.

Two main drawbacks are experienced when using the NDWI approach. The NDWI was poor at distinguishing between water and roads due to their overlapping reflectance. Any threshold selected would result in a significant amount of water or roads being misinterpreted as the other. Figure 1 shows a satellite image of urban area and the binarization result of the NDWI. The binarized result of applying the NDWI in Figure 1(b) shows the misclassification of significant amounts of road and buildings. Another problem with the NDWI is that the selection of the threshold itself is highly subjective and varies with time and region.

2.2 CNN Based Water Detection

In recent years, convolutional neural networks have become popular for water body detection. Miao *et al*. (2018) proposes the restricted receptive field deconvolution network (RRF DeconvNet) for extracting water bodies from remote sensing images. Wang *et al*. (2020) develop a multidimensional, densely connected, convolutional neural network for identifying water bodies from GF-1 images. Yuan *et al*. (2021) proposes a novel water segmentation method called multichannel water body detection network (MC-WBDN). These models all show the advantage in their ability to extract more distinctive features compared to traditional water index feature. However, the training and designing of CNN models mainly used imagery obtained from Landsat or Sentinel satellite imagery. In comparison, there are significantly fewer deep learning studies done using FORMOSAT5 imagery and even fewer using FORMOSAT-5 Imagery for water body identification using deep learning algorithms.

Many studies have proposed the use of different variations of improved U-Net models in particular for water-body extraction. One such study was conducted by Qin *et al*. (2021) which uses an improved U-Net framework for small water body extraction using satellite hyperspectral imagery of Taihu lake in Suzhou City, Jiangsu province, China. The hyperspectral imagery used in the study comprise of 32 spectral channels with a 10m spatial resolution. The results show a recall and precision of 0.8903 and 0.8950 respectively for their proposed improved U-Net model. In comparison, the U-Net model reached a recall and precision of 0.82 and 0.88 respectively which was a significant improvement over the 0.60 and 0.62 achieved by the NDWI. Another study done by An and Rui (2022) proposed an improved lightweight U-Net model called BU-Net for water body extraction. The study used multispectral imagery of 60 cities across china with 4 spectral channels and a 3.24m spatial resolution. The results of the study indicated a significant improvement in water body detection using both U-Net and BU-Net in comparison with the NDWI. The BU-Net model also achieved better results than other commonly used networks such as ResNet and SegNet.



(a) (b)

Figure 1 (a) The satellite image of an urban area; (b) Binarization result of the NDWI. Note that water bodies can be correctly identified. However, the NDWI also tends to misclassify road and buildings as water

3. Data and Data Preparation

3.1 Study Area and Data Source

Our research area covers the downtown Tainan and its suburban area (including Xinhua, Xinshi, Anding, Shanhua, Danei, and Shanshang districts). We obtained the FORMOSAT-5 imagery of this area from the Data Market of the National Center for High-performance Computing, Taiwan (National Center for High-Performance Computing, 2018). This multispectral (MS) imagery of FORMOSAT-5 has the size of 7824×6935 and comprises of Red, Green, Blue, and NIR bands. Table 1 provides detailed information regarding the multispectral bands and the spatial resolutions of FORMOSAT-5. The satellite image used in this study belongs to grid index G016 and was taken on March 10, 2021.

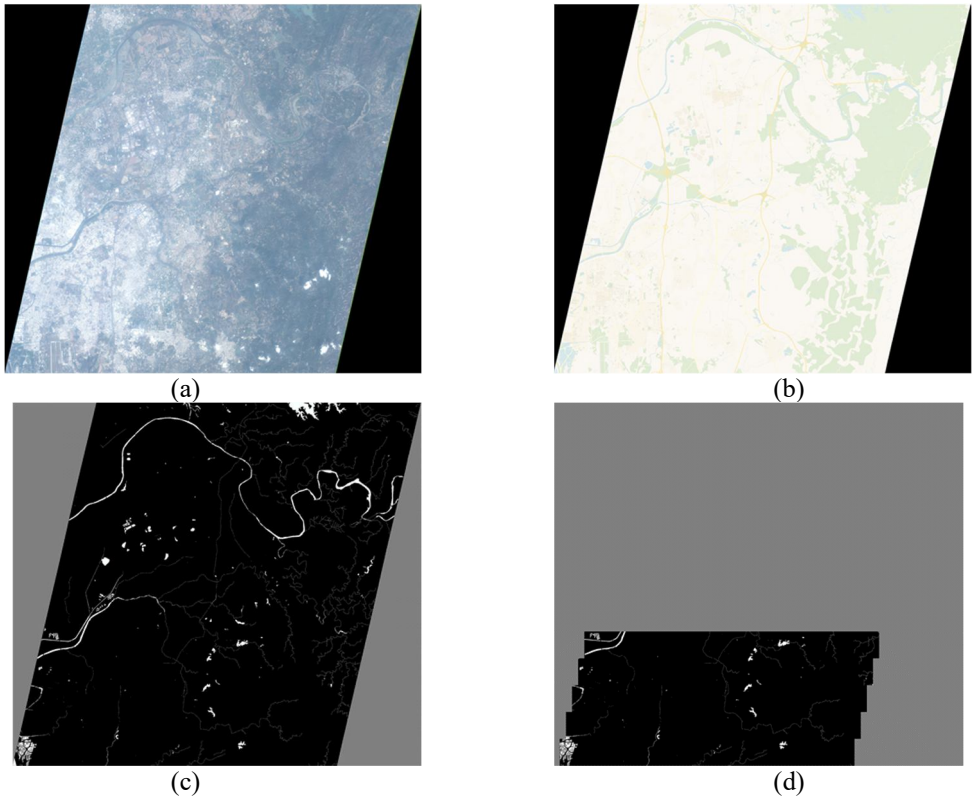
Table 1 Spectral bands of FORMOSAT-5 used in this study

band	Spatial resolution [m]	wavelength [nm]
1 - Red	4	640 - 690
2 - Green	4	525 - 605
3 - Blue	4	450 - 515
4 - NIR	4	750 - 900

3.2 Data Preprocessing

The satellite imagery is preprocessed in order to train the U-Net model. The details of our data preprocessing procedure are described below.

Except for the satellite imagery, we also need to create a ground truth of the water region to train the AI model. Given the satellite imagery as shown in Figure 2(a), we cropped a corresponding raster map of the same area as shown in Figure 2(b). Here, we chose the CARTO's Voyager basemap over Google Maps or Open Street Map because it provides a free raster map without a legend (CARTO, 2017). Notice that the legend (or text labels) might cause confusion when we extract water regions from a map. The basemap was set to the same spatial resolution as the FORMOSAT-5 imagery when cropping. While we were unable to find details on the specific map version, CARTO's Voyager updates frequency, so there should be negligible discrepancy in the mapping time when compared with the FORMOSAT-5 imagery obtained in 2021. Since the water regions on a map have a fixed color, one can easily extract water regions from the map. Specifically, we let (R,G,B) denote the RGB values at a pixel location and (R_w,G_w,B_w) denote the RGB values of water. In the CARTO's Voyager basemap, the RGB values of water region is (213, 232, 235). We computed the difference d between these two colors as follows.



(a) (b) (c) (d)

Figure 2 (a) The satellite image used in this study; (b) The corresponding Basemap; (c) The mask of the water region in the satellite image; (d) The mask of the water region in the lower part of the satellite image, where we perform evaluation of the NDWI and U-Net. Thus, we can use this mask to generate the distributions of these approaches as shown in Figure 6

$$d = \sqrt{(R - R_w)^2 + (G - G_w)^2 + (B - B_w)^2} \quad \text{..... (2)}$$

If d is greater than a predefined threshold value, then that pixel location is considered a non-water region and converted to black. Otherwise, that pixel location is considered as water region and converted to white. The resulting mask of water region is shown in Figure 2(c).

4. U-Net Model

A convolutional neural network is a class of artificial neural networks mainly used for the purpose of image recognition. The architecture CNNs generally use varied combinations of convolutional and pooling layers alongside other processes like padding and concatenation to achieve the desired outcome. The network utilizes automated learning to optimize the filters in the convolutional layers, giving the major advantage of independence from human intervention.

U-Net is a convolutional neural network that was created for the purpose of biomedical image segmentation (Ronneberger *et al*., 2015). The network itself is composed of a contractionary path and expansionary path, giving it a U-shaped architecture. The network is used for semantic image segmentation, which means the network assigns every pixel in the image a predefined class nomenclature and outputs a high-resolution image of a similar size to the input. In the context of this paper, the purpose of identifying water bodies is not whether they are independent bodies of water, but simply whether there is water in the given image.

Figure 3 shows the architecture of U-Net. The contractionary path continuously samples the feature maps of different resolutions through convolution and pooling while the part of the expansionary path up-samples the features obtained by the final encoder and then splices them with the features before the original down-sampling. The process of decoding occurs until the encoder and decoder have the same depth.

5. Experiment Results and Analysis

5.1 Experimental Setup

In order to generate training, validation, and testing sets for experimental evaluation, the original satellite is divided into three parts. In particular, we generate training, validation, and testing samples from the upper, middle, and lower parts of the original satellite imagery (as shown in Figure 4). In order to have a sufficient amount of training samples, our training samples are overlapped with each other. As shown in Figure 5, we crop a block of 512 × 512 pixels and then shift 128 pixels to crop another one. Figure 5 shows that we can generate 25 training samples of size 512×512 from an image block of size 1024×1024. By doing so, we can generate significant number of 512×512 blocks from the upper part of the original satellite imagery. Blocks that did not cover any water region were excluded from our training set. In total, we generated 824 training samples, each 512×512 pixels in size.

For the validation samples, we cropped non-overlapping blocks from the middle part of the original satellite imagery. In total, we generated 44 samples in the validation set. Similarly, the testing samples do not overlap and we can generate 55 testing samples from lower part of the original satellite imagery.

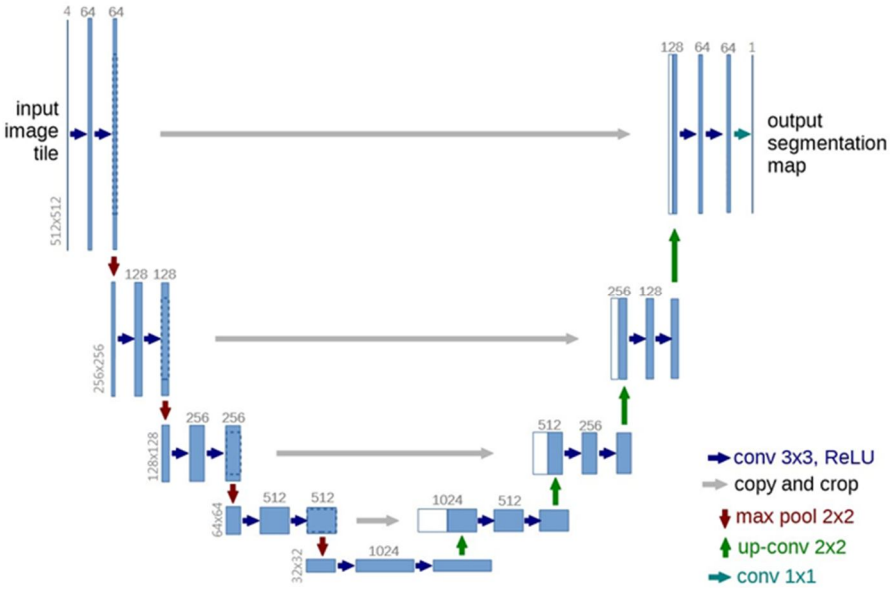


Figure 3 The U-Net architecture (Ronneberger *et al*., 2015)

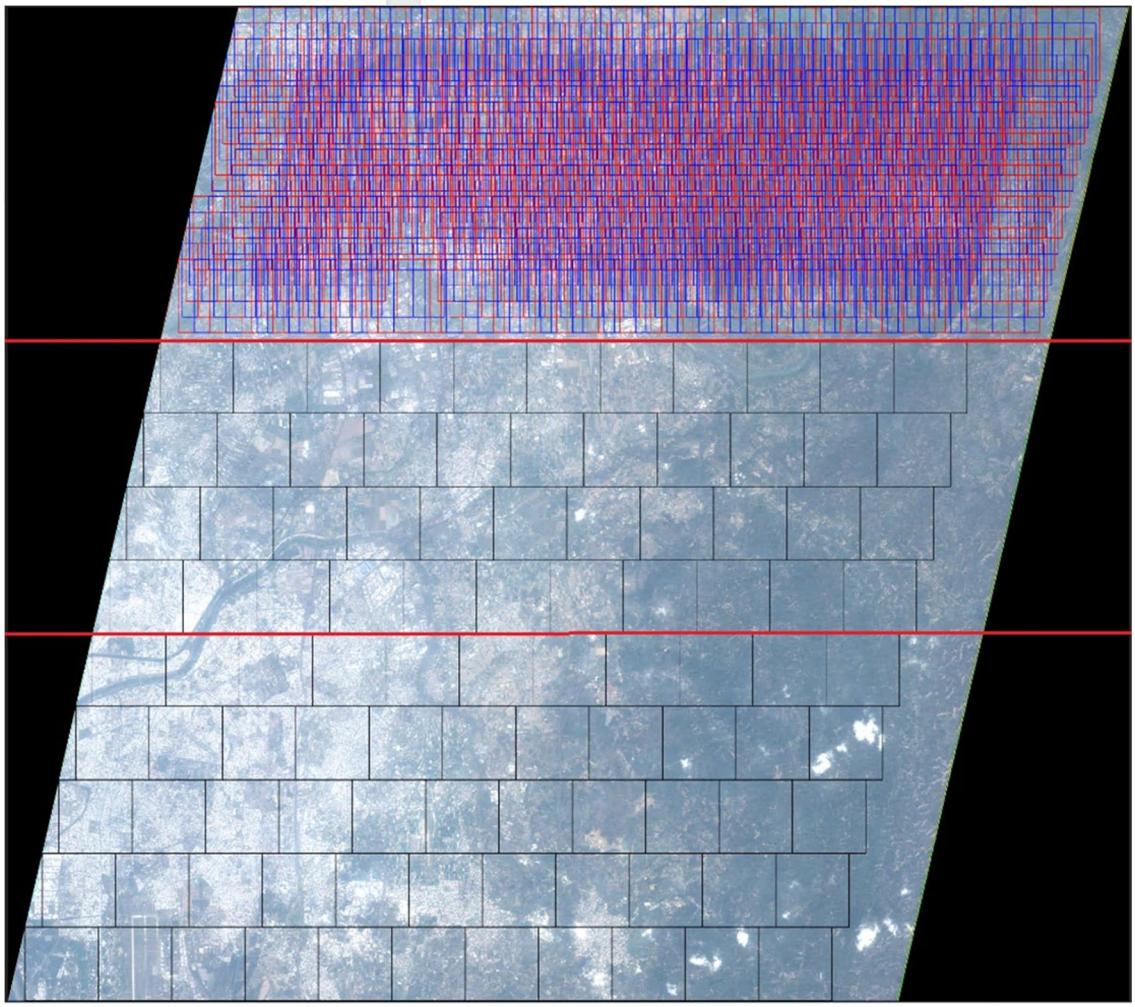


Figure 4 We divide the original satellite image into three parts (indicated by the red lines). We extract training, validation, and testing samples from the upper, middle, and lower parts, respectively. The training samples overlap with each other while the validation and testing samples do not overlap. Each sample is of size 512×512 pixels

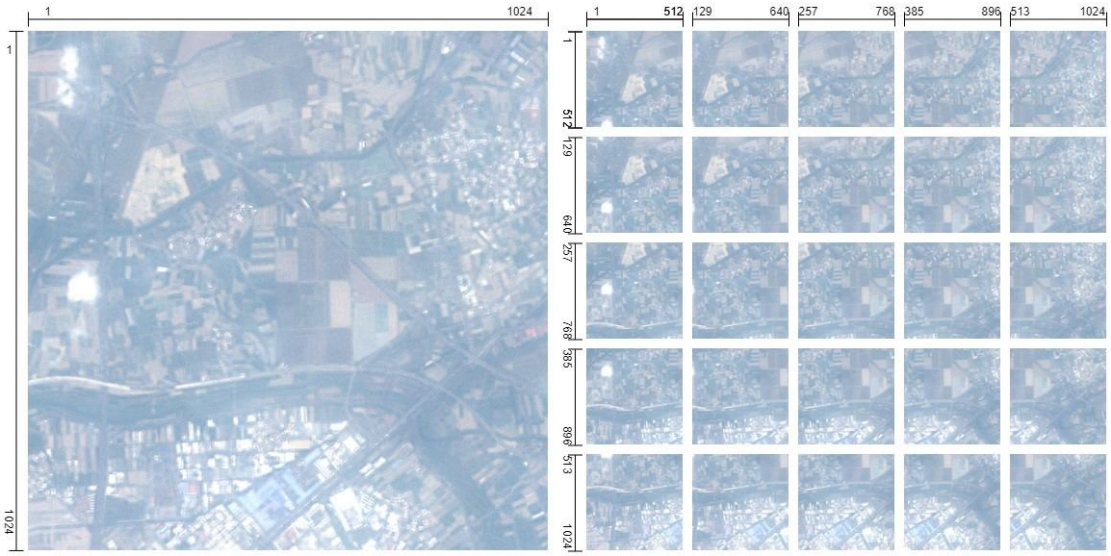


Figure 5 A satellite image block (left) and the cropped patches (right) from this block. The values on each line segment denote the first and last pixel locations of that cropped patch in the image block. In particular, we crop a 512×512 patch starting from the upper-left corner. Then, we shift horizontally or vertically by 128 pixels to crop another patch. Since these 512×512 patches are partially overlapped, we can obtain 25 patches from a 1024×1024 image block

The implementation details of training the U-Net model are listed as follows. We utilize Google Colab cloud computing service to train the U-Net model. We combine both the Binary Cross-Entropy (BCE) loss and the dice loss as our loss function. The batch size is set to 11 and the learning rate is set to 0.0005. The number of epochs is set to 20. Additionally, the Adam optimizer (Kingma and Ba, 2014) is used to speed up the training process.

5.2 Experimental Results

After training the U-Net model using the training and validation sets, we can feed the testing samples into the U-Net model and see if it can correctly identify the water regions. The lower part of Figure 4 shows where we obtain our 55 testing samples and Figure 2(d) shows the corresponding mask of water regions. To establish a baseline for comparison, we also compute the NDWI in the same area. Figure 6 show the distribution of the NDWI values and the U-Net model's output values.

In Figure 6(a), the distributions of NDWI values from non-water and water regions overlap significantly. Therefore, if we use the NDWI for identifying water regions, both false negative rate and the false positive rate would be very high.

In Figure 6(b), the distributions from the two regions are much more separate, with only a small portion of results overlapping with non-water region values. There is still a considerable amount of false negatives but the false positives are significantly reduced. Compared to the results of the NDWI, it indicates a noticeable improvement in the discrimination between non-water and water regions.

To objectively evaluate whether the U-Net model is better at identifying water than the NDWI, an effective and quantitative comparison is needed to further verify the qualitative analysis above. Among many quantitative metrics for evaluating segmentation results, we choose Receiver Operating Characteristic (ROC) curve because it is not affected by threshold setting. An ROC curve is plotted with true-positive rate (TPR) against false-positive rates (FPR) (i.e, TPR is on the *y*-axis and FPR is on the *x*-axis). The calculation formulas of the TPR and FPR are as follows:

$$TPR = \frac{TP}{TP+FN} \quad \text{.....(3)}$$

$$FPR = \frac{FP}{FP+TN} \quad \text{.....(4)}$$

where TP denotes the number of true-positives (i.e., the number of pixel locations which belong to water region and are correctly classified as water region), FN denotes the number of false-negatives (i.e., the number of pixel locations which belong to water region but are incorrectly classified as non-water region), FP denotes the number of false-positives (i.e., the number of pixel locations which belong to non-water region but are incorrectly classified as water region), and TN denotes the number of true-negatives (i.e., the number of pixel locations which belong to non-water region and are correctly classified as non-water region).

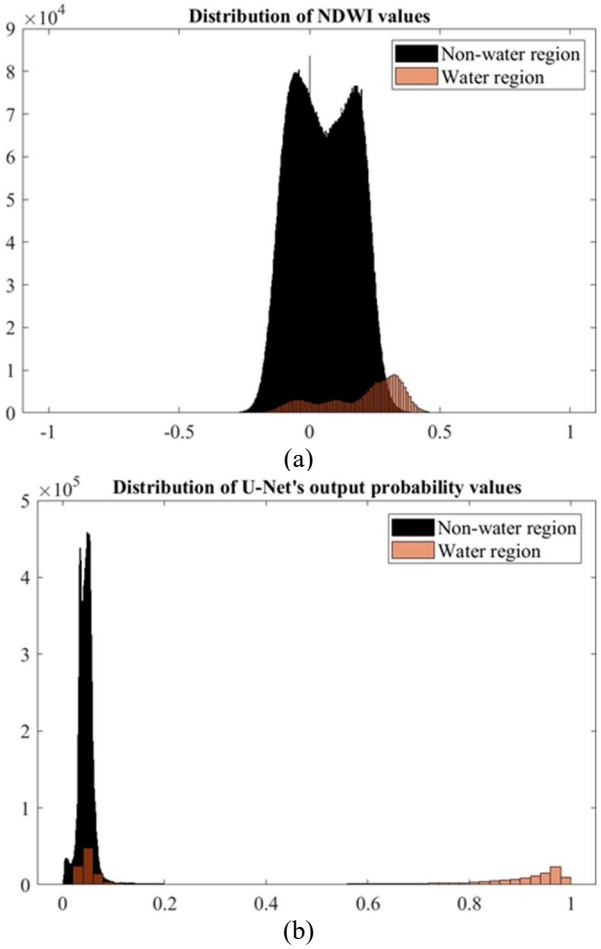


Figure 6 (a) The distribution of NDWI values from the non-water and water regions; (b) The distribution of U-Net's output probability values from the non-water and water regions. Notice that Figure 2(d) shows the non-water and water regions used for this evaluation

Thus, one can plot an ROC curve by obtaining the TPR and FPR at varied discrimination thresholds. Figure 7 shows the ROC curves of the NDWI and the U-Net. The ROC curve of the U-Net model is above the ROC curve of the NDWI for all points. The NDWI and the U-Net achieves AUC of 0.774 and 0.857, respectively. In other words, at any given FPR, the U-Net can achieve higher TPR than the NDWI. For instance, at the FPR of 20%, the U-Net model can achieve TPR of 72.8% while the NDWI only achieves of TPR of 64.3%. Or, we can also compare the FPRs of two approaches at any given TPR. It is clear that the U-Net consistently achieves lower FPR than the NDWI.

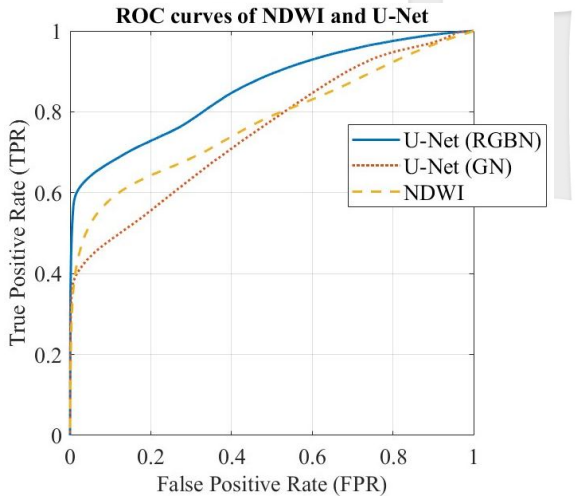


Figure 7 The ROC curves of using NDWI and U-Net for detecting water bodies from FORMOSAT-5 satellite image. Here, U-Net (GN) denotes the model trained using the same bands as NDWI: Green and NIR. U-Net (RBGN) denotes the model trained using all four spectral bands available: Red, Green, Blue and NIR. The NDWI, U-Net (GN) and U-Net (RGBN) achieve AUC of 0.774, 0.753 and 0.857, respectively

For instance, at the TPR of 80%, the FPR can be reduced from 52.4% to 32.9% by replacing the NDWI with the U-Net. Notice that the number of negative samples (i.e., non-water region) is usually much larger than the number of positive samples (i.e., water region) as shown in Figure 6. Thus, reducing about 20% of FPR indicates that a significant amount of false-positives are eliminated without compromising the accuracy of water body detection.

5.3 Discussion

This study shows the application of a U-Net CNN for identifying water bodies in the greater Tainan City area utilizing FORMOSAT-5 imagery. In Figure 8, the several sample results of the study are shown. The first column on the left shows the reference satellite imagery obtained from FORMOSAT-5. The 2^{nd} column shows the Basemap used to obtain the ground-truth for training the U-Net model. The 3^{rd} column shows the reference Google map image. When observed carefully, the water regions are not always consistent between the Google map and the Basemap. One noticeable example are the aquaculture ponds visible on the bottom left side of the images in the 3^{rd} row. The 4^{th} column shows the water segmentation results of the NDWI. Notice that in rows 1 to 3, while the NDWI was able to identify most of the areas covered with water, it also misinterpreted a lot of roads and buildings as water. The final column of Figure 8 shows the segmentation results of the U-Net model discussed in this study. While the U-Net software was not able to perfectly identify the water bodies in the satellite imagery, it was able to effectively avoid identifying roads and buildings as false positives, which demonstrates the advantage of using a U-Net CNN model. The difference is further quantified in Figure 7 and explained in Section 5.2.

In order to evaluate the usefulness of using all four spectral bands, the U-Net model was trained using two spectral bands (Green and NIR) as the input. In Figure 7, the ROC curves of the U-Net models using two bands and four bands are denoted as U-Net (GN) and U-Net (RGBN), respectively. When using two bands, the U-Net achieves a lower AUC (0.753) than the NDWI does (0.774). This indicates the usefulness of the NDWI in water body segmentation when there are only Green and NIR bands available. When using four bands as the input to the U-Net model, the AUC is significantly improved from 0.774 to 0.857. This demonstrates the advantage of using a deep learning approach for extracting discriminative information from other spectral bands to improve segmentation results. The advantage would be more considerable if we have more bands like hyperspectral remote sensing images.

6. Conclusions

This study effectively shows the comparison between the results of water body segmentation between the traditional NDWI based approach and a U-Net convolutional neural network model using FORMOSAT-5 imagery. While index based approaches still have the benefit of being simple and easy to apply, considering the improvement in recognition accuracy, the training time is negligible. As such, this method can be used for other hydrological studies such as the exploration of the Urban Heat Island Effect. Because roads, buildings and water bodies have drastically different levels of heat retention, the high error rate of the NDWI makes it a sub-optimal option for conducting studies in the area. The effectiveness of using the U-Net model in this study warrants the exploration of other CNN models in not only the image segmentation of water other aerially visible objects such as roads and buildings using imagery obtained from FORMOSAT-5.

Acknowledgment

The authors would like to acknowledge the National Space Organization (NSPO) for providing the opportunity to access FORMOSAT-5 imagery and participate in the 2022 FORMOSAT-5 satellite imagery essay competition. We would also like to thank Chia-Ming Yeh for instructing us on the usage of U-Net. Finally, we would like to express our sincere gratitude to Prof. Tee-Ann Teo, Prof. Yi-Hsing Tseng, and two anonymous reviewers for their time and constructive suggestions. Their insightful comments are very helpful in improving the quality and readability of our paper.

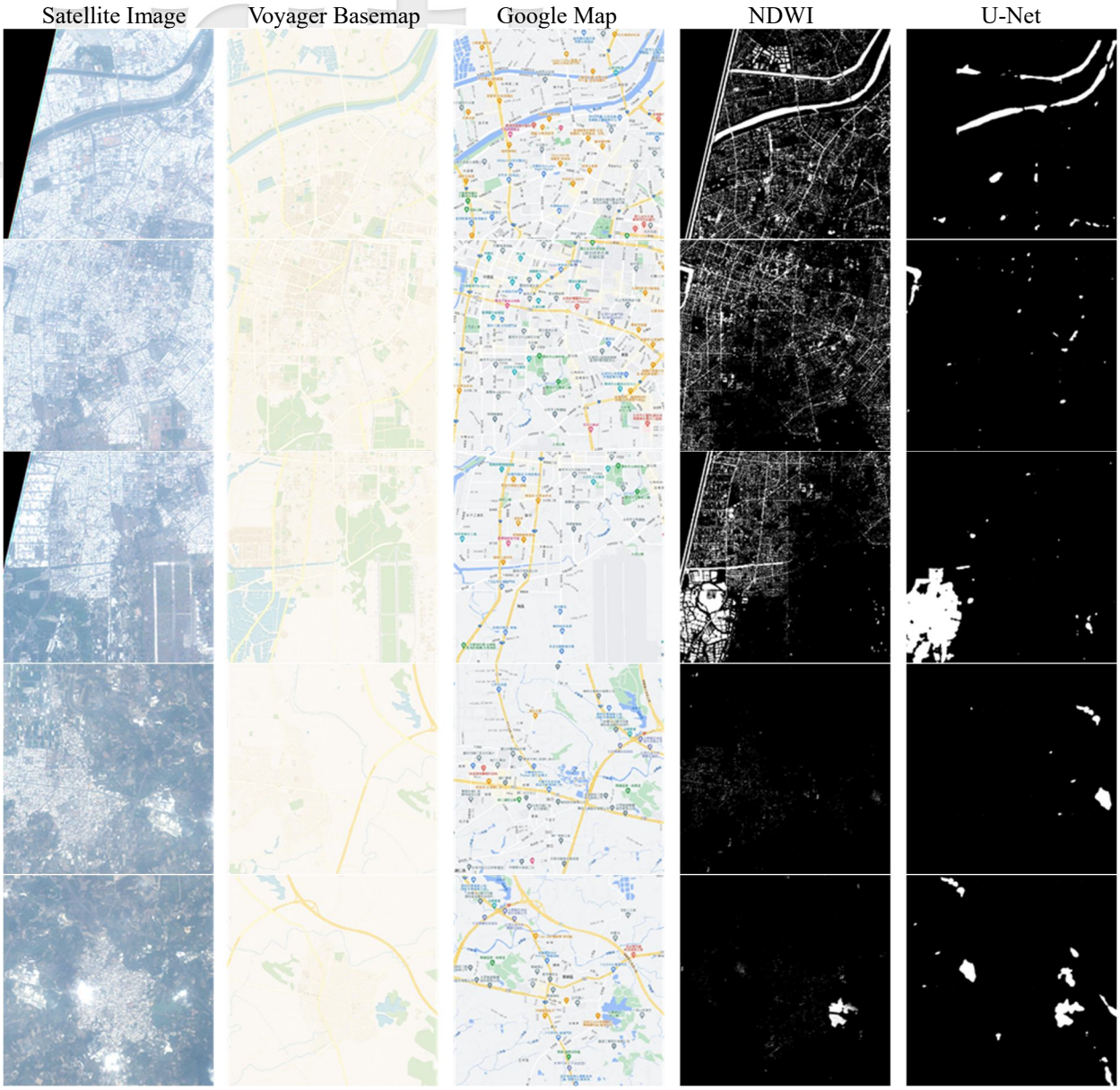


Figure 8 The 1st column shows the satellite images covering the North, West Central, South, Guiren, and Guanmiao Districts of Tainan City (from top to bottom). The 2nd column shows the corresponding Basemaps which serve as the ground truth in this study. The 3rd column shows the Google map of the same areas. Notice that the water regions on Basemap and Google map are not always consistent, especially for rural areas. The 4th column shows the water segmentation results of the NDWI. The 5th column shows the water segmentation results of the U-Net

References

An, S.H., and Rui, X.P., 2022. A high-precision water body extraction method based on improved lightweight u-net, Remote Sensing, 14(17): 4127.

CARTO, 2017. Introducing Voyager: The New CARTO Basemap, Available at: https://carto.com/blog/new-voyager-basemap/, Accessed October 08, 2022.

Feyisa, G.L., Meilby, H., Fensholt, R., and Proud, S.R., 2014. Automated water extraction index: A new technique for surface water mapping using landsat imagery, Remote Sensing of Environment, 140: 23–35.

Kingma, D.P., and Ba, J., 2014. Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

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.

Miao, Z., Fu, K., Sun, H., Sun, X., and Yan, M., 2018. Automatic water-body segmentation from high-

resolution satellite images via deep networks, IEEE Geoscience and Remote Sensing Letters, 15(4): 602-606.

National Center for High-Performance Computing, 2018. Data Market, Available at: https://scidm.nchc.org.tw, Accessed May 25, 2022.

Qin, P., Cai, Y., and Wang, X., 2021. Small waterbody extraction with improved U-Net using Zhuhai-1 hyperspectral remote sensing images, IEEE Geoscience and Remote Sensing Letters, 19: 1-5.

Ronneberger, O., Fischer, P., and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351, pp. 234–241.

Wang, G., Wu, M., Wei, X., and Song, H., 2020. Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks, Remote Sensing, 12(5): 795.

Xu, H.Q., 2005. A study on information extraction of water body with the modified normalized difference water index (MNDWI), Journal of Remote Sensing, 9: 589–595. (in Chinese)

Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., and Fang, H., 2021. 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.

運用 AI 技術進行衛星影像的水域辨識

林威昕 [1*] 張揚暉 [2]

摘要

許多衛星影像的分析應用 (例如：分析熱島效應、揚塵區域、河道變遷、水資源的監測與保護、評估洪水災害) 都仰賴於正確地找出水域範圍，進而才能夠得到有意義的分析結果。目前，最普遍被用於衛星影像的水域偵測方法為常態差異化水體指標 (Normalized Difference Water Index, NDWI)，但我們觀察到該方法用於福衛五號衛星影像時，要找出水域會有一些問題 (例如容易將道路、建築物誤判為水域)。基於這樣的觀察，我們率先嘗試運用深度學習技術來進行福衛五號衛星影像的水域分割。我們採用的卷積神經網路架構被稱為 U-Net，實驗結果顯示相較於 NDWI，U-Net 的水域分割準確率有著顯著提升。

關鍵詞：福爾摩沙衛星五號、NDWI、U-Net、水域分割、卷積神經網路

[1] 國立南科國際實驗高級中學雙語部 學生
[2] 國立南科國際實驗高級中學雙語部 數學科老師
[*] 通訊作者, E-mail: danielwilsonlin@gmail.com

收到日期：民國 111 年 10 月 03 日
修改日期：民國 111 年 10 月 19 日
接受日期：民國 111 年 11 月 01 日