SARSNet—A Novel CNN Approach for SARWater Body Segmentation

Alhassan A. Kamara[1,*], Md. Rahat K. Khan[2], and Wei Yang[1]
[1] School of Electronic and Information Engineering, Beihang University, Beijing, China
[2] School of Engineering, Ahsanullah University of Science and Technology Dhaka, Bangladesh
Email: al_kamara@buaa.edu.cn (A.A.K.), rahatkader.aust@gmail.com (M.R.K.K.), yangweigigi@sina.com (W.Y.)

***Abstract*—This paper presents the SARSNet architecture, developed to address the growing challenges in Synthetic Aperture Radar (SAR) deep learning-based automatic water body extraction. Such a task is riddled with significant challenges, encompassing issues like cloud interference, scarcity of annotated dataset, and the intricacies associated with varied topography. Recent strides in Convolutional Neural Networks (CNNs) and multispectral segmentation techniques offer a promising avenue to address these predicaments. In our research, we propose a series of solutions to elevate the process of water body segmentation. Our proposed solutions span several domains, including image resolution enhancement, refined extraction techniques tailored for narrow water bodies, self-balancing of the class pixel level, and minority class-influenced loss function, all aimed at amplifying prediction precision and streamlining computational complexity inherent in deep neural networks. The framework of our approach includes the introduction of a multichannel Data-Fusion Register, the incorporation of a CNN-based Patch Adaptive Network augmentation method, and the integration of class pixel level balancing and the Tversky loss function. We evaluated the performance of the model using the Sentinel-1 SAR electromagnetic signal dataset from the Earth flood water body extraction competition organized by the artificial intelligence department of Microsoft. In our analysis, our suggested SARSNet was compared to well-known semantic segmentation models, and a comprehensive assessment demonstrates that SARSNet consistently outperforms these models in all data subsets, including training, validation, and testing sets.**

***Index Terms*—Satellite monitoring, segmentation, convolutional neural networks, multispectral images, Synthetic Aperture Radar (SAR) microwave signals, class balancing**

I. Introduction

Water body segmentation is essential in monitoring the behavior of the water levels in the respective ecosystems [1] and to manage their sources [2]. The anomalous increase and decrease in water levels in various geographical places around the world have recorded a gruesome number of natural disasters. However, it is imperative to detect and monitor water in our environment so that warnings before disasters are in time for prompt actions. Generally, automatic extraction of water bodies in certain locations with complex surfaces rough edges, and type of water, such as rough water body, can be a challenging task [1]. Remote sensing and deep learning have emerged as lieutenants in the automatic extraction of water bodies. For continuous monitoring, cloud cover mostly affects optical data due to seasonality [3]. SAR images whose satellites can transmit and receive signals day or night, with a short revisit time around the globe, regardless of weather conditions [4] are widely applied in water body segmentation tasks.

Over the years, numerous scholars with an interest in water body segmentation have conducted several research using various methods including Normalized Difference Water Indices (NDWI) [5] and NDWI unsupervised deep learning approaches [6]. Most recently, researchers conducted a review of deep learning architectures for water body extraction and provide valuable information on the effectiveness of various deep learning architectures [7]. It follows that, with deep learning principles, researchers have made tremendous progress in producing lightweight [8] and high-precision models [9].

It is evident that prior deep learning models suffer from either of the following pressing issues, narrow waterbodies, class imbalance, and insufficient labeled data. To achieve excellent performance, researchers must treat such inherent problems with maximum attention. For efficient extraction of narrow water bodies, a new deep learning-based method that utilizes image reconstruction techniques to enhance narrow water body segmentation was proposed [10]. Other researchers employed an enhanced Unet to detect narrow water bodies [11]. A recent deep learning approach that uses Sentinel-1 SAR imagery by combining input maps from various scales to increase the accuracy of narrow water body detection was proposed. However, the model did well in narrow water body segmentation, but it demands a high number of parameters, creating high computational requirements, with costly GPU usage during training [12]. Class imbalance is also among the challenges that usually reduce the robustness of deep-learning models to accurately detect waterbodies. It has the potential to make a deep learning model underperform [13]. Numerous techniques, including the use of class re-scaling [14], similarity loss functions [15], and distribution alignment [13], have been proposed, but most have proven to be unreliable or computationally expensive [12]. Another nerve-wracking challenge is insufficient data. Deep learning models often require adequate data, which will enable the model to easily understand the related task

Manuscript received October 22, 2023; revised December 18, 2023; accepted January 15, 2024.
[*]Corresponding author

[16]. To solve this, transfer learning [17] and Albumentations augmentation approaches have been used. Geometric transformations, random erasing, mixing images, color space augmentations, and others, are all popular augmentation methods [18], though some of them possibly will introduce unwanted noise while some might alter data labels [19].

Finally, other issues that often disrupt the efficiency of our deep learning models when used with SAR Dataset include speckle noise, image resolution, and shadow waters. With SAR data, water body detection, and monitoring is reliable and continuous [4] if not for the inherent speckle noise and shadow water effects caused by SAR signal backscatter and tall buildings or trees respectively [8]. This speckle noise aka salt and pepper noise can significantly degrade the quality of the image [20]. Shadow water on the other hand also contributes to misleading the model during training [21].

To address these problems, we employed deep learning techniques to reconstruct input images into smaller patches. Reconstruction of images into patches is one of the most efficient and reliable data augmentation mechanisms [14]. Hence one of the key features added to our model for data increment and the detection of narrow waterbodies is the Patch Adaptive Network (PAN) inspired by the Kernel filter of the CNN architecture [22]. For the inherent class imbalance in deep learning, we employed the Class Pixel-level Balancing (CPB) to our PAN and the Tversky loss function [23]. To overcome SAR challenges, we enhance our model right from the preprocessing stage by introducing the SAR polarities (Vertical-Vertical (VV), Vertical-Horizontal (VH)), Fusion Register (FR) inspired by the "Dense-Coordinate-Feature Concatenate Network" [24]. The Digital Elevation Model (DEM) data recommended by [25] is also added to the FR in effect to suppress the disturbances of shadow waters. According to studies, fusing identical polarities (VV, Horizontal-Horizontal (HH)) is efficient for detecting rough surface waterbodies while the fusion of cross polarities (Horizontal-Vertical (HV), Vertical- Horizontal (VH)) is suitable on steady waters [26]. Therefore, the ground interference that affects the single polarization SAR can be suppressed by fusing different SAR image polarities, and by extension, this method also improves the structural characteristics of the feature map [27]. Also, based on the current challenges in SAR for automatic detection and monitoring of water, it seems that the long standing Unet [28] structure is no longer effective in solving current segmentation problems [9]. Therefore, the Unet [28] architecture with four down-scaling and four up-scaling layers has been modified to five layers on both sides with additional preprocessing features and adaptive data augmentations. The main contributions of this study are as follows.

1) We proposed a novel framework for a water-body segmentation technique that integrates the FR to improve the structural characteristics of the input feature map. We also added a CPB to readjust the water and non-water pixel proportions to the PAN.
2) We proposed an advanced Unet architecture that consists of five downblocks and five upblocks, making it robust enough to handle complex segmentation tasks.
3) We set up different experiments purposely to rate the performance between the BCE Jaccard loss and the Tversky loss [23] on the class imbalance issue.
4) Our experimental setup achieved the highest score compared to other state-of-the-art models.

The remaining part of this paper is organized as follows. Section II narrates the training process, the PAN augmentation, the loss function, and the model architecture. The entire experiment is explained in Section III. Finally, the related Discussion and Future Work is explained in Section IV and the Conclusion is summarized in Section V.

II. Methodology

In our research, we analyzed well-known state-of-the-art models and their own flaws. To overcome those flaws, we proposed the SARSNet model, which includes the FR, the PAN with its CPB, and the Tversky loss function [23]. In the FR, we merge the three individual channel inputs, denoted as VV, VH, and Digital Elevation Model (DEM). This advanced fusion technique enhances the performance of the model under challenging environmental conditions, for various applications related to the segmentation of the water bodies [27]. After improving the structural characteristics of the input map, we apply normalization to the Feature Fusion process. To achieve a balanced distribution of classes within the data, we transform the random values of pixels in the data labels into background values. Then the PAN augmentation technique uses the pixel-wise convolution of the kernel filter to augment our training data and improve the water pixel representation. The qualified data enters the training loop, and all parameters are fully updated via back-propagation while the Tversky loss [23] is employed to calculate the loss. After several epochs, the segmented maps are predicted for qualitative analysis while the Metrics are computed for further analysis. The training workflow is depicted in Fig.1.

In deep learning, a low percentage of water pixels in images for water body segmentation is predominant and critical [29]. Recently, Gautam and Singhai [7] have utilized false color processing and a generative adversarial network (GAN) [18] to reconstruct images and enhance the features of narrow water bodies. With this inspiration, our PAN employs CNN pixel-wise movement of the Kernel filter to cut our 512×512 input images into smaller patches of size 256×256. Guided by the convolutional arithmetic for deep learning [28], the convolutional layer's output can be influenced by a variety of parameters, including the size of the kernel (k), the padding value (p), the input size (x) and the step size of the kernel filter called strides (s).

In a square image, the convolutional output when the kernel slides vertically (V) over the image is also the same as when it slides horizontally (H) using the same parameters. For any given values of k, x, s, and p, both vertical or horizontal convolution over the image is defined as

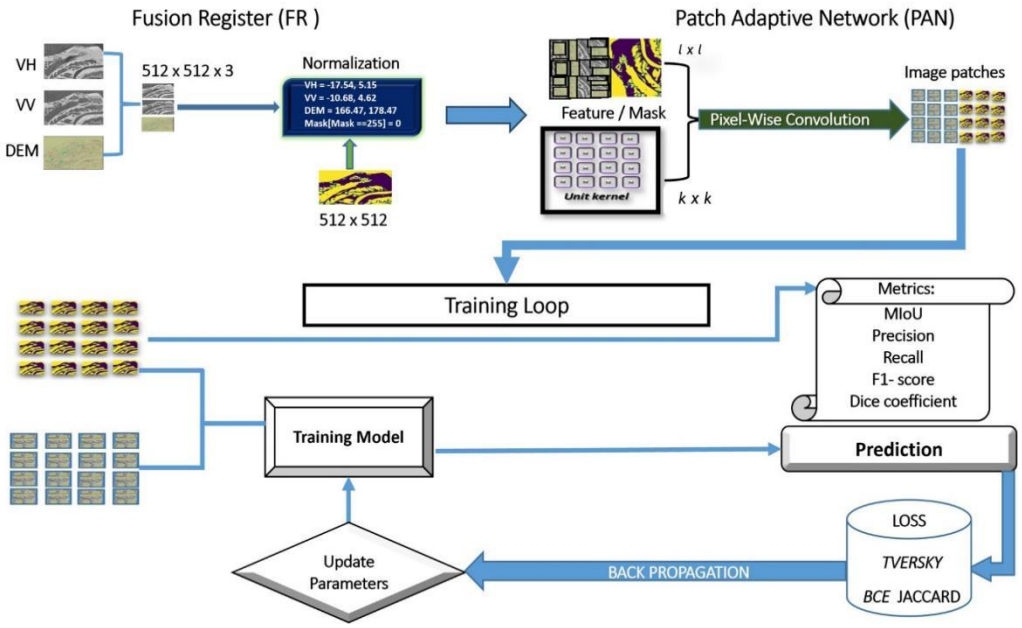

Fig. 1. Transfer Fusion data and Mask into the Training Loop according to the experimental configuration, employing a blend of FR and PAN augmentation modules. The loss function is computed utilizing the Tversky loss [23].

$$H \text{ or } V = \left(\frac{x+2p-k}{s}\right) + 1 \quad (1)$$

Hence the total number of small images extracted from a single image is represented as N in (2):

$$N = VH \quad (2)$$

From Eqs. (1) and (2), our PAN augmentation can extract multiple smaller image patches from every single input image without distorting their fine features. This implies that our 434 training samples can be automatically augmented to a thousand patches. However, with these anticipated figures, there is a high probability of class imbalance considering the nature of our dataset. As mentioned earlier, the percentage of water pixels is not evenly distributed in all the images and there are plenty of images with just a few or negligible pixels of water. Therefore, we introduce the CPB technique to the PAN augmentation approach by modifying Eq. (2) as shown in Eq. (3):

$$N = VH - \alpha \quad (3)$$

The modified equation sets α to 0.20 as a threshold value to perform automatic augmentation and then applies water CPB, which truncates patches with water pixels below the threshold. Overall, this approach increases our training inputs and adjusts the water and background proportions for a better pixel distribution.

A. Proposed Architecture

Our model presents an efficient SAR image segmentation model, referred to as SARSNet, designed to each downblock consisting of ReLU activation, dropout, and double 3×3 convolutional layers. A 2×2 max pooling layer downsampled the data, akin to downblocking1. Conversely, upblock 1mirrors downblock1, with SoftMax activation and a 2×2 convolutional layer with zero padding. Skip connections link each downblock to its corresponding upblock. Kernel quantities double from downblock 1 to downblock 3, while downblocks 4 and 5 reduce the second convolutional layer kernels by 25% and 50% compared to downblock 3. The SARSNet middle block involves a double 3×3 convolutional network, a ReLU activation, a dropout, and a 2×2 convolutional transposition with consistent padding. A skip connection is included for information flow. Table I gives a comprehensive summary of the proposed architecture.

TABLE I: SUMMARY OF THE PROPOSED SARSNET ARCHITECTURE

Layers	Conv2D$_1$	Conv2D$_2$	Pooling	Non-Linearities
FR & APN	256, 256, 3	-	-	-
Downblock1	256, 256, 16	256, 256, 16	Same	ReLU
Downblock2	128, 128, 32	128, 128, 32	Same	ReLU
Downblock3	64, 64, 64	64, 64, 64	Same	ReLU
Downblock4	32, 32, 128	32, 32, 128	Same	ReLU
Downblock5	16, 16, 256	16, 16, 256	Same	ReLU
middleblock	(8, 8, 512)	(8, 8, 512)	-	ReLU
Upblock5	16, 16, 256	16, 16, 256	Same	ReLU
Upblock4	32, 32, 128	32, 32, 128	Same	ReLU
Upblock3	64, 64, 64	64, 64, 64	Same	ReLU
Upblock2	128, 128, 32	128, 128, 32	Same	ReLU
Upblock1	256, 256, 16	256, 256, 16	Same	ReLU
Conv2D	256, 256, 2	-	-	Softmax

B. Loss Function

Since water is drastically underrepresented in our employed dataset, we added the Tversky loss function [23], which is known to influence models over the delicate imbalance between false negative predictions and false positive predictions [23]. It encourages the model to pay more attention to the minority water class than the non-water class. Consider true positive (TP), false negative (FN), and false positive in image segmentation, while setting $\alpha > 0.5$ emphasizes false negatives more while downplaying the role of false positive in (4):

$$\text{Tversky}=1-\frac{\text{TP}}{\text{TP}+\alpha\times\text{FN}+(1-\alpha)\times\text{FP}} \quad (4)$$

C. Evaluation Metrics

Irrespective of a pixel's assigned category, our primary concern typically revolves around determining the accuracy of pixel classification. Since the water pixels are in the minority, MIoU becomes the perfect metric for our task. Therefore, MIoU and various other metrics are expressed below, where x_i is the label and y_i is the expected outcome.

$$\text{Recall}=\frac{\text{TP}}{\text{FP}+\text{TN}} \quad (5)$$

$$\text{Precision}=\frac{\text{TP}}{\text{FP}+\text{TP}} \quad (6)$$

$$\text{F-1}=2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (7)$$

$$\text{MIoU}=\frac{\text{TP}}{\text{FN}+\text{TP}+\text{FP}} \quad (8)$$

$$\text{Dice Coefficient}(x_i,y_i)=\frac{2*|x_i\cap y_i|+1}{x_i+y_i+1} \quad (9)$$

III. Experiment

A. Data Domain Description

To evaluate the efficiency of our proposed model, we chose the Sentinel-1 dataset provided by the Artificial Intelligence department of Microsoft [25]. They selected several SAR (VV, VH) suites of water body scenes from 13 different countries stored as high-resolution GeoTIFF image files with 512×512 pixels value. The 542 (VV, VH) SAR image suites with their corresponding labels were collected under different types of climates and geographical conditions. Radar satellites offer a diverse array of invaluable features for Earth observation. These include their exceptional all-weather capability, and the ability to capture high-quality imagery both day and night.

In addition, Microsoft AI also recommends the inclusion of NASA's supplementary elevation data (NASADEM) which also ensures the resilience and generality of our model. However, we observed from the data source [25] that the dataset employed in this research is entangled with several challenges such as label errors, zero water pixels on some images, and fewer available valid images. In the analysis of water body segmentation, the water Pixel-wise distribution across the Dataset is always paramount. In our case, the background pixels have occupied a gruesome portion leaving the water pixel with only a small portion.

Due to the nature of these images, data augmentation is absolutely necessary. Hence the introduction of the PAN for image augmentation with (CPB) for re-scaling the proportion of water and non-water image patches. Consider (1) and (2), substitute 64 as stride (s), 256 for kernel filter (k), 0 for padding (p) on both sides of the image, and our input (x) is 512, our PAN approach was able to extract 25 image patches from every image and overall, up to 10.8k patches were generated. Equation (3) was then activated to qualify only patches with the required water pixels to enter the training loop. Table II above shows the performance of our augmentation approach.

TABLE II: ANALYSIS OF IMAGE PATCHES BEFORE AND AFTER PAN TOGETHER WITH CPB

Analysis	Total (100%)	Patches (80%)	Train (10%)	Test (10%)
Before PAN	542	434	54	54
After PAN + CPB	3025	2359	333	333

In the analysis of water body segmentation, the water Pixel-wise distribution across the dataset is always paramount. In our case, the background pixels have occupied a gruesome portion leaving the water pixel with only a small portion. Fig. 2 shows updates on the distribution of pixels in the test, train, and validation dataset.

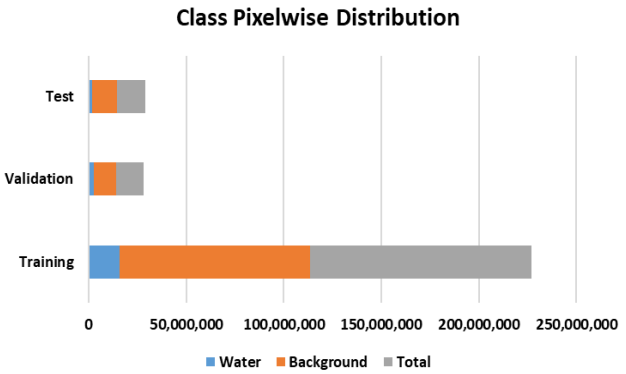

Fig. 2. Pixel distribution of water and background for training, validation, and test dataset.

B. Experimental Setup

The dataset was partitioned into separate subsets, comprising 10% for validation, 80% for training, and another 10% for testing purposes. During model development, a validation sample was created as a subclass of the training data, allowing us to fine-tune hyperparameters and evaluate the model's performance effectively. All experiments were conducted using the Tensor Flow framework and were executed on a NVidia GeForce RTX 2080 GPU. The training duration for each model typically ranged from 5 to 9 h, depending on factors such as architecture and batch size. Every model was trained using the whole training dataset for a maximum of 175 epochs, or until we saw no improvement in validation loss for a continuous period of 10 epochs. To determine proficiency and robustness we compare SARSNet against various other segmentation models, including Unet [28], DNCNN [3], U2NET [30], Unet++ [29], Attunet [31], FPN [32], LINKNET [33], and VNET [34]. To avoid bias, all experimental procedures use the same parameter values, data preparation methods, and configurations. During these experiments, we also alternated the BCE Jaccard loss recommended by Microsoft AI with that of our suggested Tversky loss function [23]. As anticipated, SARSNet with the Tversky loss function [23] consistently outperforms these models in our quantitative and qualitative analyses. Eventually, even the models of the top three winners of the

Microsoft Al competition [25] are dominated by our approach as shown in Fig. 3.

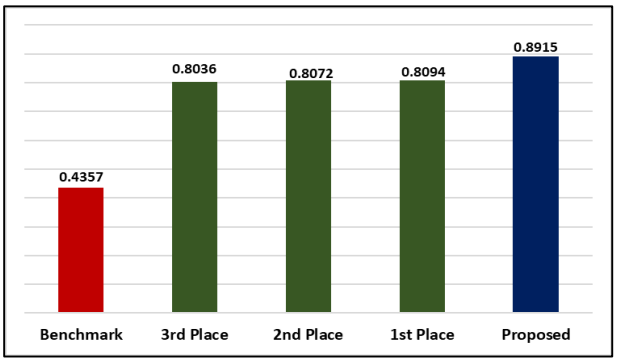

Fig. 3. Evaluation of MIOU scores for the top three contenders in the Microsoft AI competition with our proposed SARSNet.

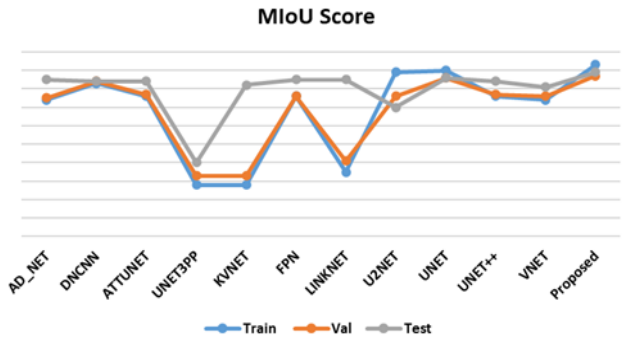

Fig. 4. Evaluating models performances using the MIoU score on the training, validation, and test datasets.

C. Quantitative analysis

In Section II C above, we listed five metrics and since our task has a very high interest in the minority class, we used the MIoU for evaluation. This score serves as a comprehensive reflection of our model's performance following the whole training process. Fig. 4 shows the MIoU score curves on all the models. The curves illustrated that Unet3PP [35], KVNET [34], and U2NET [30] perform poorly on the test data while our proposed model shows the best performance followed by Unet [28] and FPN [32].

Nevertheless, for the sake of reference, we also provide scores of Recalls, Precision, Dice Coefficient, and F-1.

Table III presents a statistical evaluation of various models using identical parameters. Significantly, our SARSNet excels by achieving the highest MeanIoU scores: 0.89 in testing, 0.90 in validation, and 0.93 in training.

In water body detection and monitoring, one of the inherent challenges is dealing with data imbalance, which can hinder efficiency. To reduce its effect, the dataset providers, Microsoft AI for Earth STAC API, recommend using the BCE Jaccard loss function. However, in our investigations, we have found that the Tversky loss function [23] offers a better solution for addressing data imbalance. Table IV presents a summary of our results. We conducted training sessions for 175 epochs for each loss function and documented their performance on the test set, focusing on metrics such as MIoU, precision, and F-1 score. However, it is evident that the Tversky loss function [23] consistently outperforms the BCE Jaccard loss in all experiments. These findings informed our decision to exclusively employ the Tversky loss function in all our experiments.

TABLE III: SARSNET EXPERIMENTAL RESULTS, DERIVED FROM THE DATASET [25], COLLECTED FOLLOWING 175 EPOCHS WITH AN ERROR RATE OF ±2.50%

Model	Tversky Losss			MIoU			Dice Coff			F-1 score			Precision			Recall		
	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test
AD_NET	0.37	0.33	0.07	0.74	0.75	0.85	0.63	0.67	**0.93**	0.61	0.58	0.79	0.61	0.61	0.82	0.61	0.68	0.88
DNCNN	0.21	0.08	0.07	0.83	0.84	0.84	0.79	0.92	0.92	0.77	0.8	0.79	0.78	0.83	0.82	0.77	0.87	0.88
ATTUNET	0.46	0.43	0.08	0.76	0.77	0.84	0.54	0.57	**0.93**	0.53	0.5	0.79	0.53	0.53	0.82	0.54	0.59	0.87
UNET3PP	0.5	0.49	0.35	0.28	0.33	0.4	0.5	0.51	0.64	0.49	0.45	0.47	0.5	0.5	0.53	0.5	0.54	0.6
KVNET	0.5	0.5	0.08	0.28	0.33	0.82	0.5	0.51	0.92	0.49	0.45	0.79	0.5	0.5	0.81	0.5	0.54	0.88
FPN	0.45	0.41	0.07	0.76	0.76	0.85	0.55	0.59	0.92	0.54	0.51	**0.8**	0.55	0.54	0.82	0.55	0.6	0.88
LINKNET	0.5	0.49	0.07	0.35	0.41	0.85	0.5	0.51	0.92	0.49	0.45	0.79	0.5	0.5	0.82	0.5	0.54	0.88
U2NET	0.23	0.2	0.24	0.89	0.76	0.7	0.55	0.61	0.59	0.59	0.58	0.55	0.63	0.61	0.68	0.74	0.71	0.72
UNET	0.18	0.07	**0.06**	0.9	0.86	**0.86**	0.82	0.93	**0.93**	0.8	0.81	**0.81**	0.82	0.84	**0.83**	0.79	0.88	**0.89**
UNET++	0.48	0.47	0.07	0.76	0.77	0.84	0.52	0.53	0.92	0.51	0.47	0.79	0.52	0.52	0.82	0.52	0.56	0.87
VNET	0.48	0.47	0.09	0.74	0.76	0.81	0.52	0.53	0.9	0.51	0.47	0.76	0.51	0.51	0.81	0.52	0.56	0.85
Proposed	0.17	0.06	**0.06**	0.93	0.87	**0.89**	0.83	0.94	**0.94**	0.81	0.82	**0.81**	0.83	0.84	**0.84**	0.8	0.89	**0.9**

TABLE IV: EVALUATION OF SARSNET FOR DIFFERENT LOSS FUNCTIONS

Loss Function	MIoU (%)	Precision (%)	F-1 Score (%)
Tversky Loss [23]	89	84.3	81.1
BCE Jaccard Loss	86.1	78.4	77.8

D. Qualitative Analysis

Here we analyze the output quality and compare the performance of our model to earlier models in segmenting both large and narrow water bodies. We randomly selected six predictions from our testing dataset, as shown in Fig. 5. The initial four rows correspond to VH, VV, DEM, and Ground Truth (GR), while the subsequent rows display the predicted result generated by the models. Our proposed model's predictions are placed in the last row. In the presentation of results for each of the six scenarios, we observe distinct outcomes derived from the models.

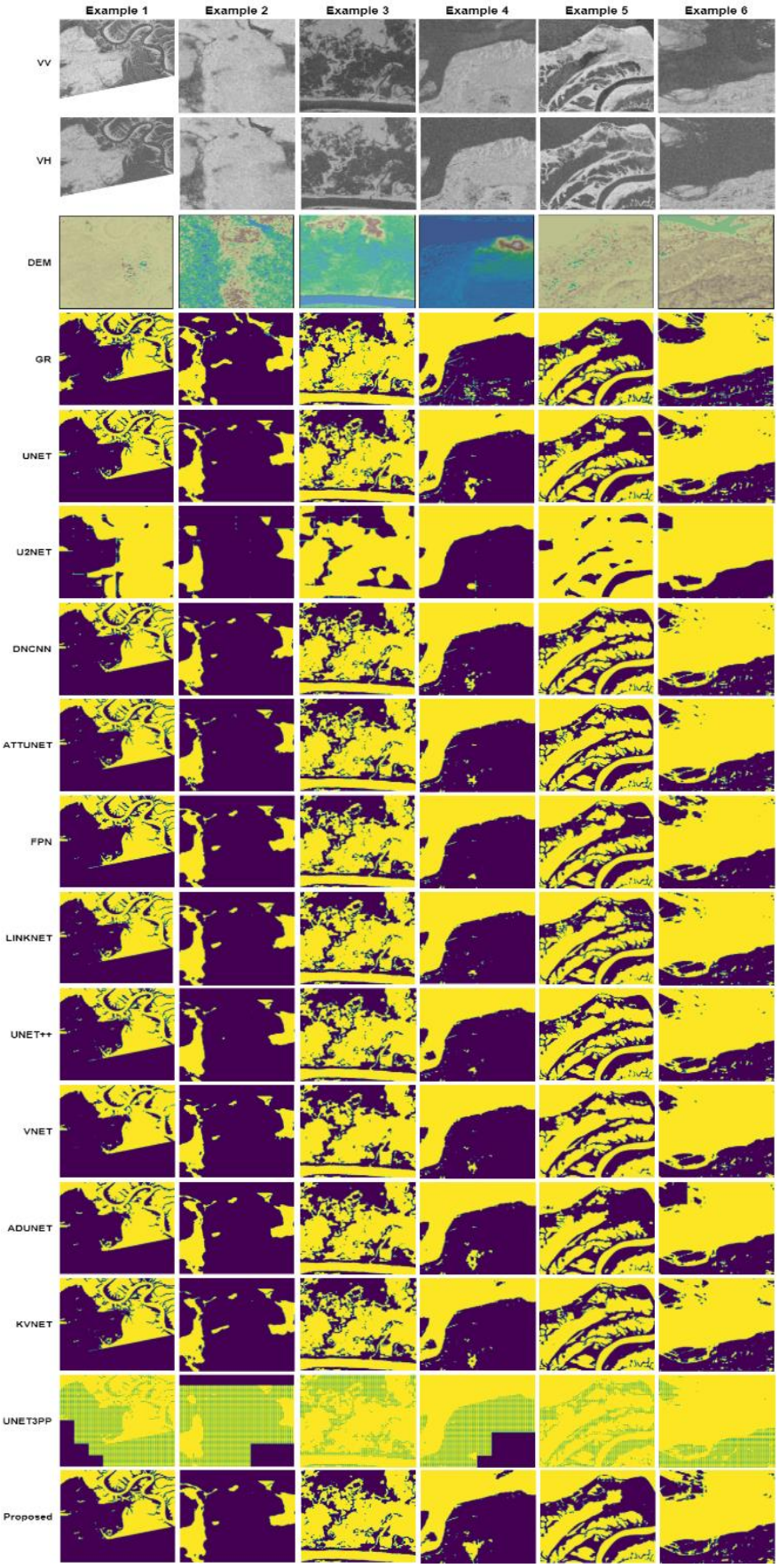

Fig. 5. A total of six samples, selected at random, have been chosen for the purpose of evaluating the performance of our proposed SARSNet in comparison to the existing samples.

A noteworthy observation is that U2NET [30], Attunet [31], DNCNN [3], Unet++ [29], and VNET [34] exhibit limitations in accurately predicting water pixels across all six samples. In contrast, our newly proposed SARSNet, along with Unet [28], LINKNET [33], and FPN [32], demonstrated superior abilities to forecast water pixels accurately when compared to alternative models. However, FPN [32] and LINKNET [33] also exhibit weak performances in the fourth column while the U2NET [30] and Unet3PP [35] had the worst performance. SARSNet excels in detecting both small and large water areas. Clearly, our proposed SARSNet consistently outperforms all other models, demonstrating exceptional performance in accurately predicting nearly all water pixels while maintaining a remarkably low level of noise.

IV. Discussion and Future Work

The SARSNet architecture was developed to address two key challenges. Firstly, obtaining sufficient annotated data is a significant hurdle in deep learning, with a growing demand for diverse Dataset. Secondly, deep learning relies on complex feature representations, reducing the need for extensive feature engineering but increasing the need for annotated data. Our dataset had specific limitations, and we can enhance the model's performance by exploring advanced data augmentation techniques like style transfer, GAN generated data, and different input shapes. Balancing class distribution, customizing the loss function, and fine-tuning the hyperparameters, such as dropout rates and kernel numbers, can improve the model's performance. SARSNet, optimized for water body detection using SAR data, outperforms Unet with similar training times.

In our upcoming research, we will explore advanced methods like style transfer and GANs. We'll modify our model to also handle RGB bands, employing transfer learning for comparison.

We'll also assess the effects of different input SAR polarities on model performance, adjusting model layers for further advancement in this field. Our future work aims to enhance the model's performance and versatility.

V. Conclusion

This study presents an innovative methodology for the delineation of water bodies from satellite imagery, drawing upon advanced deep learning techniques in the realm of satellite remote sensing. Notably, our proposed model uses a reduced number of parameters in comparison to its predecessors, while delivering superior performance. A significant contribution of our research lies in the introduction of two innovative preprocessing techniques. Firstly, the FR module is designed to integrate and normalize the VH and VV polarizations alongside the DEM data. This fusion aids in enhancing the overall quality of the input data. Secondly, the PAN augmentation algorithm was employed to generate image patches. Critically, this algorithm ensures that the image resolution is preserved throughout the augmentation process. In our evaluations, the SARSNet approach demonstrates superiority over existing state-of-the-art methodologies. This was particularly evident when assessing the performance of metrics such as the Mean Intersection over Union (MeanIoU) scores. On the qualitative analysis, our model shows proficiency in accurately delineating water bodies. Our approach differs from conventional Unet models, which predominantly focus on the segmentation of large water bodies. Instead, our methodology harnesses the power of advanced augmentation techniques to detect and delineate even the smallest water bodies, thereby expanding the scope and applicability of satellite-based water body detection systems.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Alhassan Alie Kamara: Conceptualized the study, developed the methodology, data analyses, performed validation, general analyses, and drafted the original manuscript. Md. Rahat K Khan: Contributed to writing, reviewed and edited the manuscript, handled software-related tasks, and contributed to the visualization of the results. Yang Wei: Provided supervision, resources and technical advice throughout the project, administered the overall project, and acquired funding for the research. All authors had approved the final version.

Funding

This research received funding from National Science Foundation of China (NFSC) under Grant No. 62271028 and Grant No. 62101014, and National Key Research and Development Program of China 2022YFB3901601.

References

[1] Z. Liu, X. Chen, S. Zhou, H. Yu, J. Guo *et al*., "Dupnet: Water body segmentation with dense block and multi-scale spatial pyramid pooling for remote sensing images," *Remote Sensing*, vol. 14, no. 21, #5567, Nov. 2022.

[2] Z. Zhang, D. Liu, Z. Liu, Y. Qiao *et al*., "Deep learning based methods for water body extraction and flooding evolution analysis based on sentinel-1 images," in *Proc. 7th ICHCE and SWIDR*, Dec. 2021, pp. 191–195.

[3] K. Zhang, W. Zuo, Y. Chen *et al*., "Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. on Image Processing*, vol. 26, pp. 3142–3155, Feb. 2017.

[4] Z. Guo, L. Wu, Y. Huang *et al*., "Water-body segmentation for sar images: Past, current, and future," *Remote Sensing*, vol. 14, #1752, Apr. 2022.

[5] E. ̈Ozelkan, "Water body detection analysis using ndwi indices derived from landsat-8 oli," *Polish Journal of Environmental Studies*, vol. 29, no. 2, pp. 1759–1769, Feb. 2020.

[6] J. Li, Y. Meng, Y. Li, Q. Cui *et al*. "Accurate water extraction using remote sensing imagery based on normalized difference water index and unsupervised deep learning," *Journal of Hydrology*, vol. 612, #128202, Jul. 2022.

[7] S. Gautam and J. Singhai, "Critical review on deep learning methodologies employed for water-body segmentation through remote sensing images," *Journal of Multimedia Tools and Applications*, vol. 23, pp. 1869-1889, 2024.

[8] Y. Wang, S. Li, Y. Lin *et al*. Wang, "Lightweight deep neural network method for water body extraction from high-resolution remote sensing images with multisensory," *Sensors*, vol. 21, #7397, Nov. 2021.

[9] S. An and X. Rui, "A high-precision water body extraction method based on improved lightweight u-net," *Remote Sensing*, vol. 14, no. 14, #4127, Aug. 2022.
[10] Y. Luo, A. Feng, H. Li *et al*. "New deep learning method for efficient extraction of small water from remote sensing images," *PLOS ONE*, vol. 17, no. 8, p. e0272317, Aug. 2022.
[11] P. Qin, Y. Cai, and X. Wang, "Small waterbody extraction with improved u-net using zhuhai-1 hyperspectral remote sensing images," *IEEE Geo-Sci and Remote Sensing Letters*, vol. 19, pp. 1–5, Jan. 2022.
[12] C. Jiang, H. Zhang, C. Wang, J. Ge, and F. Wu, "Water surface mapping from sentinel-1 imagery based on attention-unet3+: A case study of poyang lake region," *Remote Sensing*, vol. 14, #4708, Sep. 2022.
[13] J. Billson, M. S. Islam, X. Sun *et al*., "Water body extraction from sentinel-2 imagery with deep convolutional networks and pixelwise category transplantation," *Remote Sensing*, vol. 15, no. 5, #1253, Feb. 2023.
[14] D. Misra, C. Crispimunior, and L. Tougne, "Patch-based CNN evaluation for bark classification," in *Computer Vision – ECCV Workshops: Glasgow*, UK, Aug. 23–28, Ch. 5, pp. 197–212.
[15] J. Zhang, M. Xing, G.-C. Sun, J. Chen *et al*. "Water body detection in high-resolution sar images with cascaded fully- convolutional network and variable focal loss," *IEEE Trans. on Geo. Sci. and Remote Sensing*, vol. 59, pp. 316–332, Jan. 2021.
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84–90, May 2017.
[17] J. Chae and J. Kim, "An investigation of transfer learning approaches to overcome limited labeled data in medical image analysis," *Applied Sciences*, vol. 13, #8671, Jul. 2023.
[18] M. R. K. Khan, M. S. Islam, M. Al-Mukhtar *et al*., "Sagan: Maximizing fairness using semantic attention based generative adversarial network," in *Proc. of the IEEE 3rd Int'l. Conf. on Big Data and AI (ICIBA)*, Jul. 2023, vol. 3, pp. 721–726.
[19] A. Buslaev, V. Iglovikov, E. Khvedchenya *et al*., "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, #125, Feb. 2020.
[20] V. Kumar, A. Dubey, M. Gupta *et al*. "Speckle noise reduction strategies in laser-based projection imaging, fluorescence microscopy, improved image sharpness, and resolution," *Optics & Laser Technology*, vol. 141, no. 141, #107079, Apr. 2021.
[21] N. Inoue and T. Yamasaki, "Learning from synthetic shadows for shadow detection and removal," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 31, pp. 4187–4197, Nov. 2021.
[22] F. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *ArXiv*, pp. 15-64. Jan. 2018,
[23] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention u-net for lesion segmentation," in *Proc. of the 16*th *IEEE Int'l. Symp. on Biomed. Imaging (ISBI 2019)*, Italy, Apr. 2019, pp. 683– 687.
[24] W. Xue, H. Yang, Y. Wu, P. Kong, H. Xu, P. Wu, and X. Ma, "Water body automated extraction in polarization sar images with dense-coordinate-feature-concatenate network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, #1207312087, 2021.
[25] Stac overflow: Map floodwater from radar imagery. [Online]. Available: https://www.drivendata.org/competitions/81/detect-flood-water/page/386/#metadata
[26] S. Grimaldi, J. Xu, Y. Li, V. Pauwels, and J. Walker, "Flood mapping under vegetation using single sar acquisitions," *Remote Sensing of Environment*, vol. 237, #111582, 2 2020.
[27] M. S. Islam, X. Sun, Z. Wang, P. Ghuman, and I. Cheng, "Mnet: Semantic segmentation for satellite images based on multi-channel decomposition," *Engineering Proceedings*, vol. 21, no. 1, 2022.
[28] W. Weng and X. Zhu, "Unet: Convolutional networks for biomedical image segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021.
[29] Z. Zhou, M. Rahman., N. Tajbakhsh *et al*., "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, *et al.* Ed. 2018, pp. 3–11.
[30] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jager sand, "U2-net: Going deeper with nested u-structure for salient object detection," *Pattern Recognition*, vol. 106, #107404, 2020.
[31] O. Ozan, S. Jo, F. Loic, L. Matthew *et al*., "Attention u-net: Learning where to look for the pancreas," *arXiv:1804.03999*, Apr. 2018.
[32] A. Xiao, X. Yang, S. Lu, D. Guan, and J. Huang, "Fps-net: A convolutional fusion network for large-scale LIDAR point cloud segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 176, pp. 237–249, Jun. 2021.
[33] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. of 2017 IEEE Visual Communications and Image Processing (VCIP)*, Dec. 2017, pp. 1–4.
[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. of 2016 4*th *Int'l. Conf. on 3D Vision (3DV)*, Dec. 2016, pp. 565–571.
[35] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.W. Chen, and J. Wu, "Unet 3+: A full-scale connected Unet for medical image segmentation," in *Proc. of 2020 IEEE Int'l. Conf. on Acoustics, Speech and Signal Processing*, Apr. 2020, pp. 1055-1059.

Alhassan A. Kamara obtained his BSc. in Electronics and Telecommunications and the MSc. in Telecommunications and Networking, from Njala University, Sierra Leone, in 2016 and 2020, respectively. Between 2016 and 2018, he served as Information and Technology officer at the Ministry of Youth Affairs, Government of Sierra Leone. From 2018 to 2020, he held the position of Network Operation Center Engineer at the Dedicated National Security Information System, State House Freetown, Sierra Leone. He assumed the role of Researcher and Teaching Assistant at the Department of Physics and Computer Science, Njala University, from 2020 to 2022. Currently pursuing another master's in information and communication engineering at the School of Electronics and Information Engineering, Beihang University, Beijing, China. His research interests include the monitoring and detection of surface water dynamics, multitemporal synthetic aperture radar (SAR) imagery, Deep Learning, Machine Learning and Artificial Intelligence.

Md. Rahat Kader Khan is currently a master's student in Computer Engineering at the University of Mississippi, USA. He completed his B.Sc. in Computer Science and Engineering at the Ahsanullah University of Science and Technology, Bangladesh, in 2022. He also worked as a machine learning engineer at Canada Syntax Machine Learning. His current research areas of interest are computer vision, machine learning, and artificial intelligence.

Wei Yang (Member, IEEE) was born in 1983. He received the M.S. and Ph.D. degrees in signal and information processing from Beihang University (BUAA), Beijing, China, in 2008 and 2011, respectively. From 2011 to 2013, he held a post-doctoral position at the School of Electronics and Information Engineering, BUAA. Since July 2013, he has been a Lecturer with the School of Electronics and Information Engineering, BUAA. From 2016 to 2017, he researched as a Visiting Researcher at the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield, U.K. Since 2018, he

has been an Associate Professor with the School of Electronics and Information Engineering, BUAA. He has authored or coauthored more than 60 journal and conference publications. His research interests include moving target detection, high-resolution space borne synthetic aperture radar (SAR) image formation, SAR image quality improvement, and 3-D imaging.