

NDR U-Net: Segmentation of Water Bodies in Remote Sensing Satellite Images Using Nested Dense Residual U-Net



Naga Surekha Jonnala^{ID}, Neha Gupta^{*ID}

School of Electronics Engineering, VIT-AP University, Amaravati 522237, Andhra Pradesh, India

Corresponding Author Email: neha27brs@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.380324>

ABSTRACT

Received: 16 July 2023

Revised: 10 January 2024

Accepted: 20 February 2024

Available online: 21 June 2024

Keywords:

remote sensing, segmentation, deep learning, water-bodies, nested dense residual network, U-Net

The utilization of remote sensing satellite imagery is crucial for advancing scientific knowledge in identifying water bodies. This technology has diverse applications, including natural disaster forecasting, as well as the detection of droughts and floods. Currently, there are few programs available specifically designed to locate naturally disappearing water bodies. Accurately delineating aquatic features from Very high-resolution (VHR) remotely sensed and moderately high-resolution images poses a significant challenge in remote sensing. The intricate spectrum combinations resulting from aquatic vegetation, various lake/river colours, mud along the sand, and shadows from surrounding plants make it difficult to determine water body boundaries. To enhance water body extraction from VHR and moderately high-resolution remote sensing images, it is imperative to increase feature variety and semantic data. In this study, we employed the nested dense residual network, known as D3net, to detect water bodies in satellite images. The Adam optimizer was utilized for training the satellite images, minimizing losses in the process. The optimizer also identified the optimal values for the activation function and the number of nodes in each layer. For segmentation, 5682 Sentinel-2 satellite images (comprising 2841 satellite images and their respective masks) were used, focusing on the Europe geo location. These masks were generated using the Normalized Water Difference Index (NWDI). The proposed model demonstrated a performance of 95.36% Intersection over Union (IOU) and 96.99% accuracy, making it suitable for edge detection, blurry image recognition, and low-resolution image detection. While the model is reliable and accurate in its predictions, it requires more memory due to the utilization of VHR and moderately high-resolution images for segmentation.

1. INTRODUCTION

The identification of water bodies in remote sensing images is a crucial task for effective management of land and water resources, as well as for preventing disasters like droughts and floods. Retrieving accurate water information from remotely sensed data has long been a significant focus in remote sensing data processing. Many existing water body segmentation methods struggle to precisely locate the boundaries of water bodies. The irregular shape of water bodies poses a challenge for traditional threshold-based and machine learning-based techniques to accurately segment them. Therefore, traditional techniques are not better at segmenting objects in satellite images, especially when dealing with complex structures, cluttered backgrounds, or objects with varying sizes and shapes. Traditional machine learning methods rely on extensive frequency analysis and feature extraction based on prior knowledge to identify water bodies. Deep learning-based approaches have reduced the need for explicit feature extraction, but they require a large amount of training data and processing capacity to perform well. While deep learning models have shown the ability to recognize certain structures present in the current context, they still face challenges such

as the requirement for a substantial amount of training data and the blurring of boundary pixels in the segmentation process. In this research, the aim is to develop a unique deep learning approach that addresses these challenges and improves the accuracy of water body segmentation in remote sensing applications.

This research primarily concentrates on accurately delineating water bodies in intricate and challenging environments, utilizing very high resolution (VHR) and moderate high resolution remote sensing imagery. These images, covering substantial portions of the Earth's water surface, are captured by equipment aboard satellites and aircraft. In VHR and moderate high resolution remote sensing images, discerning the boundaries of water bodies can be difficult. Challenges often arise due to factors such as aquatic plants, mud, boats along the coast, and shadows cast by adjacent tall plants. In recent years, there has been a significant impact on the extraction and detection of natural resources, including water bodies and forests, through the analysis of satellite images. Regular monitoring of water bodies is essential for ensuring their sustained growth. Monitoring serves various purposes, such as the early identification of drought and flood conditions and effective disaster

management. Satellite images have evolved into a crucial tool for locating people, animals, buildings, water bodies, and ships. The primary objective is the identification of water features through the analysis of a substantial dataset of satellite images obtained from Kaggle.

For semantic segmentation, a fully convolutional Nested dense residual UNet (NDRU-Net) Deep Learning (DL)-technique is utilized and it is termed as D3net. There have been numerous attempts to use current neural networks to segment images; however, D3net performs the best overall with the least amount of information loss [1, 2]. In order to produce more precise semantic segmentation results, particularly better edges and boundaries, several researchers enhanced the basic prediction findings [3-6]. To accomplish high-resolution (HRL) prediction, Lin et al. [3] employed long-distance residual connections for all multi-scale features throughout the down sampling procedure. The residual refines module (RRM), an independent encoder-decoder created by Qin et al. [4], is used to postprocess the semantic segmentation findings. To improve the feature maps, Yu et al. [5] suggested using the refinement residual block (RRB). In order to maximize the approximate prediction outcomes, Cheng et al. [6] created a special-purpose refinery network using global and local refining. However, the repetitive structural design in the majority of these approaches might result in duplication. The main contribution of our work:

- We introduce a straightforward multi-layered residual framework referred to as the "nested dense residual UNet" designed for segmentation purposes. The NDRU-Net incorporate strategies to enhance accuracy, such as the use of nested dense residual networks, multi-dilated convolutions, or skip connections. These architectural choices can enable the model to capture intricate features and improve the precision of segmentation.

- The Adam optimizer was employed for training the proposed model, aiming for improved performance. The NDR U-Net implement techniques to improve efficiency, such as optimizing network architecture incorporating efficient training algorithms. Additionally, the model transfer learning to expedite convergence.

- Deep learning visualization, including the use of a gradient activation map, was applied to both images and their corresponding masks. This visualization method enhances our understanding of the network's classification process. The NDR U-Net could employ strategies to enhance robustness, such as data augmentation during training to expose the model to diverse scenarios. Robustness also be improved by careful selection of hyper parameters, regularization techniques, and thorough validation on a representative dataset.

The remaining manuscript is structured as follows: section 2 presents previous state-of-the-art methods; section 3 describes the datasets used in the work; Section 4 gives the proposed method of water body segmentation using NDRU-Net; section 5 presents the results of proposed method; and section 6 gives final conclusion of the work.

2. RELATED WORK

Recently, work on DL-based water-body segmentation from remote sensing data has gained attention and progress [7-11]. By taking into account both spectral and spatial information, Yu et al. [7] are pioneers in developing a CNN-based technique for water-body extraction from Landsat

images. This CNN-based method, however, divided an image into small tiles in order to make pixel-level predictions, which led to a high level of redundancy and low efficiency. A constrained receptive field deconvolution network was suggested by Miao et al. [8] to extract water bodies from HR remote sensing images. To extract water bodies from VHR images, Li et al. [9] used a conventional FCN model, It performed significantly better than techniques using the normalized difference water index (NDWI), support vector machine (SVM), and sparsity model (SM). In order to segment water bodies more precisely, Duan and Hu [10] suggested a unique multi-scale refinement network (MSR-Net) that fully used the multiscale properties. Although the MSR-Net has a multi-scale module, this model does not take into account channel links between feature maps and does not reuse high-level semantic information. Guo et al. [11] introduced a multiscale feature extractor that included four dilated convolutions with varying rates and was deployed on top of the encoders. They used a straight forward FCN-based technique for water-body extraction. This FCN-based technique did not extract entire features at different scales; instead, it utilized the multi-scale information of high-level semantic features. It is clear that feature extraction and prediction optimization were prioritized in recent FCN based water extraction experiments, but there is still much potential for improvement. In order to identify minor water bodies and conduct quality analyses in Wuhan, Wang et al. [12] developed a model. The approach is based on band analysis and image categorization. For quality analysis, a semi-empirical deep learning algorithm is employed. Since it is entirely based on an inversion model, accuracy evaluation is performed at regular intervals then finished by the inversion of water quality. The methodology has certain flaws, including the difficulty of detecting water features in mixed images. A technique for segmenting water bodies from remote sensing images using mask Region Based Convolutional Neural Network (R-CNN) is suggested by Yang et al. [13]. The proposed method addresses the drawbacks of inadequate resolution and pixel clarity in remote sensing images. Both the ResNet-50 and the ResNet101 techniques were used to train the model. To solve a number of issues in the field of remote sensing, Lira [14] offers a technique. This model makes use of optical reflectance's and segmentation methods. The above approach has a problem in that false overlays is shown at various points during detection. Lalchhanhima et al. [15] proposed a method. That use a hybrid CNN that takes into account U-Net and Inception, makes use of sparsely distributed SAR data, and speckle noise image segmentation. The output of this model depends on a number of aspects, including the training, dataset size, and accuracy, among others. Due to a variety of variables, Zhang et al. [16] discussed the difficulties in precisely determining the limits of water bodies. A new combination and multi-feature extraction are provided as solutions to these problems. To get rich feature representation, three feature extraction sub modules are employed for different spaces and connections. A change for the current ECDSA and their uses was made by Shang [17] The ECDSA's efficiency and security are enhanced by this work. The ECDSA method is chosen because it has fewer drawbacks and more advantages than the RSA one. It increases effectiveness by analyzing various assaults, such as side-channel attacks. A technique, verifiable ECDSA, was introduced by Yang et al. [18]. The main objective is to be able to use a public key provided by the reliable service to encrypt

the digital signature. The biggest disadvantage of this proposed work is the storage need of a 270 MB assessment look-up table. A model based on Mobile-UNet was created by Jing et al. [19] To find fabrics, it uses the CNN approach. Segmenting fabric flaws from end to end can be done effectively. Simple imbalance is reduced by applying the loss function of the median frequency. Using a softmax layer, segmentation mask detection is performed. The model's cutting-edge precision is one of its main advantages. Tensorflow was used by Wang et al. [20] to give the scientific computation of fluid flows. In this project, the graph-based module of Tensorflow is utilized. Additional performance and accuracy analyses revealed great scalability up to TPU v3 pod. The TPU platform is used to create the model. A grid search for solar power forecasting with LSTM using Nadam Optimizer was found by Sharma et al. [21]. The LSTM provides better results for time-series data. In this study, eight alternative neural network models and an optimizer are compared against the LSTM along with two time series. A research of semantic segmentation for aquatic bodies was introduced by Erfani et al. [22] The ATLANTIS dataset is the foundation of this effort. This might stop emergencies from occurring during floods. For the purpose of distinguishing between aquatic and non-aquatic environments, the AQUANET model was created. The strongest part of the work is its dataset, which can cover a wide range of classes with different types of dataset. Using high quality satellite images, Rajyalakshmi et al. [23] established a method to identify water bodies. Thresholding methods are used in this paper. The dataset consists of images with a wide range of spectral and temporal characteristics. In this study, a brand-new technique called single-band threshold employing bilateral filtering is used. A technique for IOV with a fault-tolerant ECDSA signature was proposed by Lin et al. [24] Moreover, it is vulnerable to various attacks. Pedestrians and vehicles can communicate with each other. There is a possibility that a third party can view this sensitive information during transmission. A U-net has been proposed by Chen et al. [25] as a building identification technique. Using remote sensing images, this model helps identify buildings. It is created by replacing the n-channel feature maps with a series of smaller meshes.

Convolutional Neural Networks (CNNs) and U-Net architectures have proven to be powerful tools in various computer vision tasks, especially in image segmentation. Here are some common limitations associated with CNNs and U-Net: Lack of Global Context: CNNs and U-Net architectures, particularly in their basic forms, may struggle with capturing extensive global context information. This limitation can affect their ability to understand and contextualize relationships between distant pixels or regions in an image. High Computational Requirements: Training deep CNNs, including U-Net, can be computationally intensive. As the depth and complexity of the network increase, so does the demand for computational resources. This can be a limitation in resource-constrained environments or for real-time applications. On a multi-spectral image with 13 bands, Kaplan and Avdan [26] describe a process that combines the usage of a pixel-based index and an object-based strategy. Sentinel-2 images with resolutions ranging from 10 meters to 60 meters are used in the model. The model considers a mountainous and urban setting for a better understanding of the performance. In above all mentioned methods used limited datasets and achieved less accuracy. The mentioned models did not have the security for predicted masks and computational cost is also

more. So, we have developed NDRUNet model for the segmentation to achieve the more accuracy and ECDSA is used to protect the predicted mask details.

3. MATERIALS

A collection of Sentinel-2 aerial images has been gathered, specifically focusing on the geographic location of Europe. These images encompass numerous water bodies within the region [27]. The Sentinel-2 satellite carries a Multi Spectral Instrument (MSI) with specific characteristics. Here are some specifics about the Sentinel-2 dataset. Spatial Resolution: Sentinel-2 provides different spatial resolutions for its spectral bands. Each image includes a respective black-and-white mask in which the colour black stands for anything other than water and the colour white represents water. The masks were created using the Normalized Water Difference Index (NWDI), which is often used to identify and quantify vegetation in satellite images. However, a higher threshold was used to identify water bodies. The model receives the water body dataset as input. These images were acquired and captured by the Sentinel-2 satellite and the spatial resolution ranges from 10 metres to 60 metres [28, 29]. The dataset consists of two directories. First are the images, and second are the masks. White means the masks contain no water, while black means anything but water. These masks are created using the Normalized Water Difference Index (NWDI) [30]. Normalized Difference Water Index (NDWI) may refer to either of at least two liquid water-related indexes generated by remote sensing: One uses near-infrared (NIR) and short-wave infrared (SWIR) wavelengths to monitor variations in leaf water content, as proposed by Gao in 1996 [31]. This can be represented mathematically as W_{NIR} .

$$NDW1 = \frac{W_{NIR} - W_{SWIR}}{W_{NIR} + W_{SWIR}} \quad (1)$$

where, W_{NIR} and W_{SWIR} is the reflectances of the near infrared band and reflectances of near short-wave infrared reflectances respectively.

$$NDW1 = \frac{W_G - W_{NIR}}{W_G + W_{NIR}} \quad (2)$$

where, W_G is reflectances near green bands. Eqs. (1) and (2) both are used to monitor the changes in the water level. This NDWI creates the digital and visual interpolation of the output image that is -1 to 0 shows bright surface and no evidence of the water content and $+1$ represents the water content.

The evaluation of the proposed approach for water body segmentation utilized a dataset obtained from Kaggle, consisting of Sentinel-2 satellite images of various water bodies. The dataset included a black-and-white mask for each image, where black represented non-water areas and white represented water areas. The masks were generated using the Normalized Water Difference Index (NWDI), a commonly used index for vegetation identification in satellite images. To acquire the data, the Sentinel-2 API was utilized, and the images were pre-processed using rasterio. The masks were generated by applying a higher threshold on the NWDI values obtained from bands 8 and 3 of the satellite data. The dataset contained a total of 5682 images, divided into two directories 2841 images in the image directory and 2841 corresponding

masks in the mask directory. Figure 1 shows the complete Flow chart. As an example, Figure 2 likely displays one of the input images from the dataset, providing a visual representation of the satellite image showing water bodies with their corresponding masks. Table 1 clearly shows the overall dataset of Training, Testing and validation of Images.

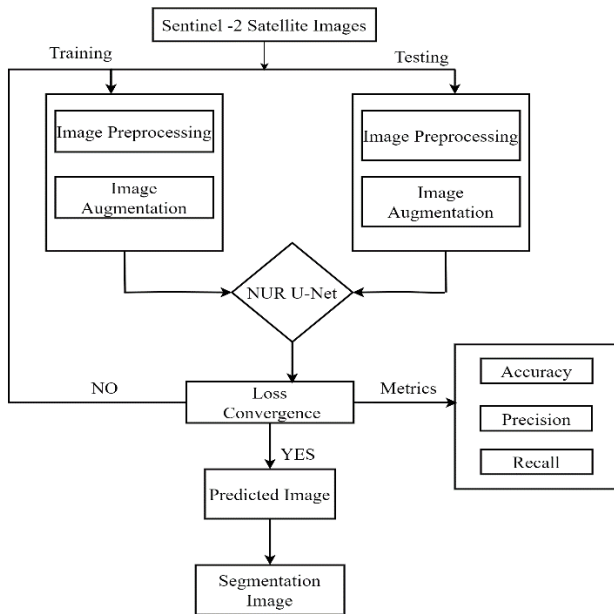


Figure 1. Flow chart of NDR U-Net

Table 1. Overall dataset

Data	Number of Images
Training	2290
Testing	300
Validation	250
Total	2841

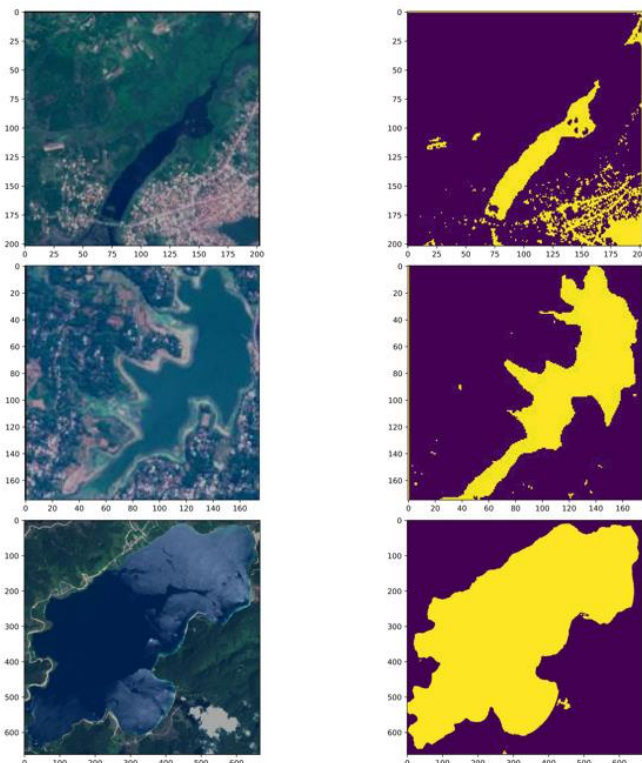


Figure 2. Sample Images with their respective masks

4. METHOD

The data used for the proposed methodology comes from the Water Bodies dataset provided by the study [27]. The dataset consists of masks applied to the corresponding original images.

The dataset does not contain any additional parameters, only images. Figure 1 shows the flow chart of NDR U-Net. The model is developed incrementally according to the agile development methodology to facilitate the incorporation of future enhancements and maintenance. Seven key high-level phases are used to build the model.

At the beginning, the data set is fed into the model. The exploratory analysis will utilize this data set as its base. Exploratory data analysis entails statistically examining the data to find any irregularities that might have an impact on the model’s performance. This is a precautionary step performed at the outset to ensure smooth performance of the model. As part of the exploratory data analysis, the original images and the masks are first visualized comparatively. The main difference is found in some images. These ground truths seem to contradict each other. Therefore, the images are checked for rotations or padding. In the data preprocessing step, the images are transformed.

The first transformation that must be performed is resizing. Few images that have larger dimensions are difficult to accommodate in memory and differ significantly from other images. So, a threshold of 1500 is applied to both height and width, and all those images having dimensions greater than the threshold are re-scaled. Mask conversion is the next transformation operation that is conducted on the data. In Figure 3, when a histogram of the original images is displayed, the majority of the values are centered around zero. This can be the result of rotation. For those values in the middle, the brightness must be increased. This uses histogram equalization. Masks, on the other hand, must be divided into two distinct values using a threshold and are recorded in jpeg format. The process of mask conversion is employed for this. The Scale values are the third transformation function that is necessary.

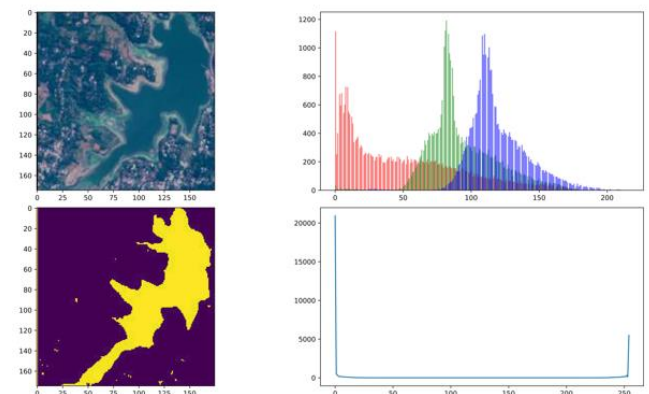


Figure 3. Histogram equalization of sample image and respective mask

The primary objective is to speed up the model’s convergence. To do this, the original image is separated into the [0, 1] range. The range division does not take the masks into account. Padding images is the last transformation operation to be performed on the data. Since the model permits images of any size, specified padding values shouldn’t be

taken into account. Down sampling must be able to split the size of the image depending on the strides. More connections between non-consecutive layers and the identical size of the input and output images are produced by the neural network's specificity. As a result, padding must be used.

The model is trained using NDRUNet in the fourth important phase. NDRUNet performs better in segmentation when compared to other techniques [32]. The normal, relu hidden activation, and sigmoid output activation are used to initialize NDRUNet. Convolution, encoder, and decoder blocks must be built in order to create a model using NDRUNet. There are two crucial considerations for the convolution block's creation. The filter count, which is set to 64 by default, serves as the first input for both the convolution layer and the other one. The suggested model makes use of partial convolution. Encoder and decoder construction is straightforward. The encoder takes in inputs in order from top to bottom. It has two convolutional layers, ReLU, and batch normalization layers for each of them. Each block of the decoder needs inputs, a skip connection, and a count of filters. To get the final skip connection value, the input is up-sampled, and the prior input is additionally concatenated. The methodological structure for the model detection of water bodies utilizing NDRUNet and tensor flow with ECDSA is shown in Figure 4. It gathers the dataset, goes through a number of intermediate phases, and then produces the original image together with the anticipated mask and an ECDSA digital signature. The steps involved to process the proposed model are • Collection of Dataset. • Splitting the Dataset into

train and test groups with a 70:30 ratio. • Data Processing. • Building NDRUNet. • Train the model by defining training parameters. • Testing the model and calculating IOU.

4.1 Data processing

Data processing is the initial stage before training a model. Data must first be pre-processed before it can be utilized as input to a model. In this, we can provide input without resizing it. The data are cleaned up using this technique, which also eliminates trash and unused information. It supports keeping the data set consistent. To suit the input of the model, all photos in this step are scaled to the same value. Additionally, it normalizes the data by reducing the fill bits and scaling the image's pixel values to [0 1]. To get better results during the training and testing steps, every image must be the same size. So it's vital to get rid of these pieces.

4.2 Building blocks of NDR U-Net

The building blocks of NDRNet is shown in the Figure 4 and the architecture of UNet is shown in the Figure 5. The proposed architecture of NDRUNet shown in Figure 6 is the concatenation of both UNet and NDRNet. The details of the model are shown in the Figure 7. NDRNet is made up with multidilated convolutions that are formed by D2-block [33] in the network model. This network model has 2,006,337 total parameters and 2,006,337 are trainable parameters.

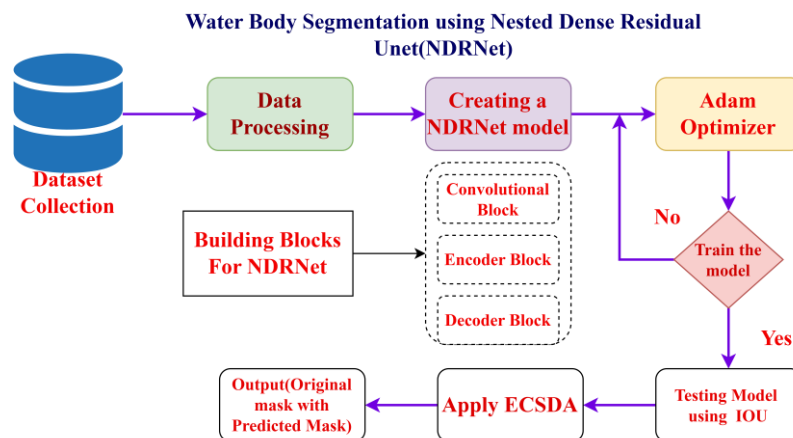


Figure 4. Block diagram of waterbody segmentation using NDR U-Net

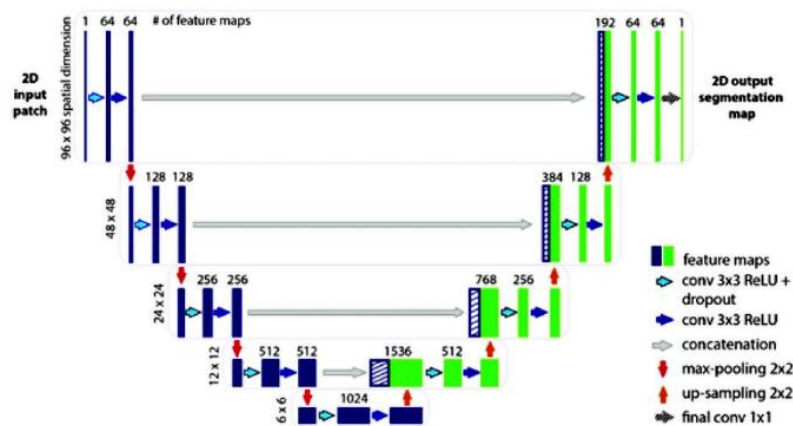
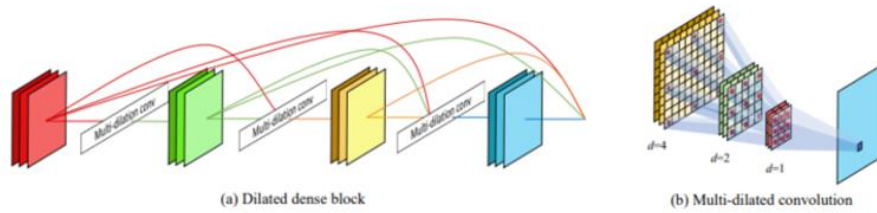
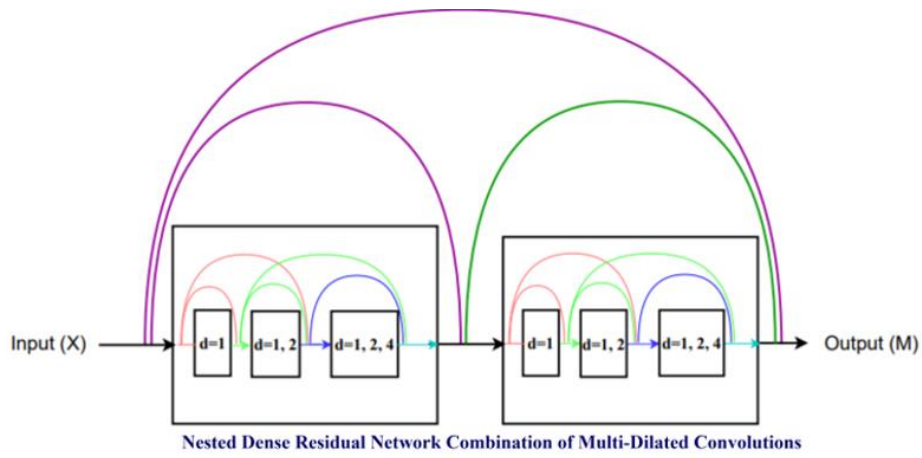


Figure 5. U-Net architecture



Structure of D2 block. (a) The connectivity pattern is the same as in DenseNet except that the D2 block involves the multi-dilated convolution at the third layer. (b) Illustration of the multi-dilated convolution at the third layer.

Figure 6. Nested dense residual network (NDRNet) architecture

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, None, None, 3)]	0	[]
conv2d (Conv2D)	(None, None, None, 32)	896	['input_1[0][0]']
conv2d_1 (Conv2D)	(None, None, None, 32)	9248	['conv2d[0][0]']
max_pooling2d (MaxPooling2D)	(None, None, None, 32)	0	['conv2d_1[0][0]']
conv2d_2 (Conv2D)	(None, None, None, 64)	18496	['max_pooling2d[0][0]']
conv2d_3 (Conv2D)	(None, None, None, 64)	36928	['conv2d_2[0][0]']
max_pooling2d_1 (MaxPooling2D)	(None, None, None, 64)	0	['conv2d_3[0][0]']
conv2d_4 (Conv2D)	(None, None, None, 128)	73856	['max_pooling2d_1[0][0]']
conv2d_5 (Conv2D)	(None, None, None, 128)	147584	['conv2d_4[0][0]']
max_pooling2d_2 (MaxPooling2D)	(None, None, None, 128)	0	['conv2d_5[0][0]']
conv2d_6 (Conv2D)	(None, None, None, 256)	295168	['max_pooling2d_2[0][0]']
conv2d_7 (Conv2D)	(None, None, None, 256)	590080	['conv2d_6[0][0]']
layer_normalization (LayerNormalization)	(None, None, None, 256)	512	['conv2d_7[0][0]']
leaky_re_lu (LeakyReLU)	(None, None, None, 256)	0	['layer_normalization[0][0]']
conv2d_8 (Conv2D)	(None, None, None, 32)	8224	['leaky_re_lu[0][0]']
concatenate (Concatenate)	(None, None, None, 288)	0	['conv2d_7[0][0]', 'conv2d_8[0][0]']
layer_normalization_1 (LayerNormalization)	(None, None, None, 288)	576	['concatenate[0][0]']
leaky_re_lu_1 (LeakyReLU)	(None, None, None, 288)	0	['layer_normalization_1[0][0]']

Figure 7. Illustration structure of NDR U-Net

The NDRU-Net employs dilated factors (1, 2, and 4) to enhance receptive coverage. In the NDR U-Net architecture, two Multi Dilated Convolutions (MDC) blocks are integrated, each comprising three hidden layers. Within each MDC layer, convolution is performed with varying dilation factor values. The MDC utilizes a 3×3 kernel size and dilation factors (1, 2, and 4). A total of 13 Skip Connections are implemented for feature reuse. The application of small dilation factors is strategically directed toward extracting local features, while larger dilation factors are employed to capture global features. This strategy ensures a more comprehensive coverage of the input space. To understand the network behaviour for classification we have applied the Gradient Activation Map (GCAM) to images and masks. This is shown in the Figure 8.

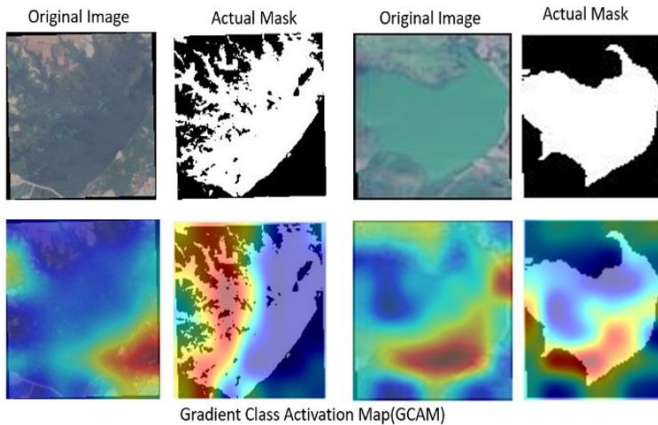


Figure 8. GCAM Applied for both image and actual mask

4.3 Adam optimizer

The algorithm for the gradient descent optimization approach is called adaptive moment estimation. When dealing with a big problem with plenty of data or parameters, the strategy is particularly effective and uses little memory. It combines the RMS algorithm and the Gradient Descent with Momentum method, intuitively. In this model, we have used 30 epochs, 1e-4 learn rate, and a 0.3 learn rate drop factor along with the Adam optimizer to get better performance. This optimizer gives the better performance compare all other optimizers [34].

4.4 ECDSA

The projected mask is given an ECDSA to strengthen the integrity of the specified deep learning model [35]. Rivest, Shamir, and Adleman (RSA) are inferior to ECDSA in terms of power and scalability [36]. It makes use of elliptic curve cryptography-derived keys. With lower key lengths, it accomplishes the same tasks as other digital signature algorithms. By translating the signature key into bytes, the final digital signature is produced. By using NIST192p, which has a length of 24 bytes, one may acquire the signature key. To verify the digital signature on the opposite side, the verification key may also be retrieved from this signing key.

Elliptic Curve Digital Signature Algorithm (ECDSA) is a widely used cryptographic algorithm for ensuring the authenticity and integrity of digital messages. While the underlying mathematics of ECDSA involve elliptic curve mathematics, I'll provide a simplified and intuitive explanation without delving into complex mathematical details.

Key Generation: ECDSA involves the generation of a pair

of cryptographic keys: a private key and a public key. The private key is kept secret, while the public key can be freely shared.

Elliptic Curve Properties: ECDSA relies on the mathematical properties of elliptic curves. These curves have special mathematical structures that make certain operations, like scalar multiplication, easy to perform in one direction but computationally infeasible in the reverse direction.

Digital Signature Generation: When a user wants to sign a message, they use their private key and the message as input to the ECDSA algorithm. The algorithm performs computations using the elliptic curve, generating a digital signature.

Signature Verification: To verify the authenticity of the signature, anyone can use the signer's public key, the received message, and the signature as input to the ECDSA algorithm. If the signature is valid, the verification process will succeed.

5. RESULTS AND DISCUSSION

We utilized the NDR U-Net model for semantic segmentation on a desktop within Google Collab, featuring 32GB of RAM, a 1TB SSD, and NVIDIA GTX graphics. The dataset employed for training and evaluation consisted of 5,682 images, comprising 2,841 satellite images along with their respective masks. These images were fed into the NDR U-Net model as input for the semantic segmentation task. The training and testing curve of the model is shown in the Figure 9. Figure 10 shows the output of the proposed model and shows the original images, respective water source is clearly visible and differentiated with other objects present in the images. The performance metric of the model is tabulated in the Table 1. After classifying we have calculated the Intersection Over Union (IOU) using the mathematical relation:

$$IOU = \frac{TP}{TP + FP + FN} \quad (3)$$

where, TP is true positive, FP is false positive, TN is true negative, and $F - N$ is false negative. We have achieved a good accuracy of 93% of IOU.

GCAM enhances the interpretability of deep neural networks by offering a visual explanation of where the model focuses when making predictions. It helps practitioners, researchers, and stakeholders understand and trust the decision-making process of complex models. The darkened color portion is shown that that portion is more needful for detection and more features are used for the classification. Lightened color portion of image shown that that portion is not more useful for classification and minimum number of features are used for classification. Finally, to validate our proposed model, we compared our model to state-of-the-art models.

Decreasing Training Loss: In the initial stages of training, it's expected to observe a decrease in the training loss. The model learns to map input data to the target outputs, and the training loss measures the disparity between the predicted and actual values. As the model iteratively updates its parameters during training, the training loss typically decreases. **Validation Loss:** The validation loss is computed on a separate dataset that the model has not seen during training. Initially, both training and validation losses tend to decrease. However,

it's essential to monitor the behaviour of the validation loss over epochs. Convergence Indicators: Convergence is often indicated by both the training and validation losses reaching a stable or plateau-like state. If the training loss continues to decrease while the validation loss starts to increase or remains stagnant, it may suggest overfitting. Overfitting occurs when the model learns the training data too well but fails to generalize to new, unseen data. "Utilizing small kernels is employed for extracting local features, while large kernels are

employed to extract global features. However, many CNN and U-Net based models opt for a fixed kernel size for both, resulting in a deficiency of receptive coverage for feature extraction and a significant loss of information. Some approaches based on 'densenet' aim to enhance the receptive area by exponentially processing kernel sizes. Nevertheless, these methods face challenges such as the aliasing effect and an increase in the number of trainable parameters, leading to a notable loss of efficiency."

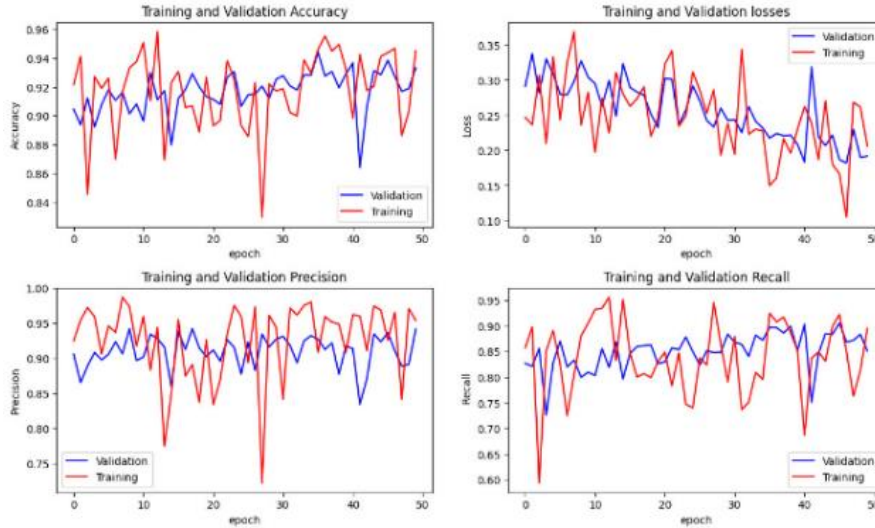


Figure 9. Training and testing curves of accuracy, precision, loss and recall

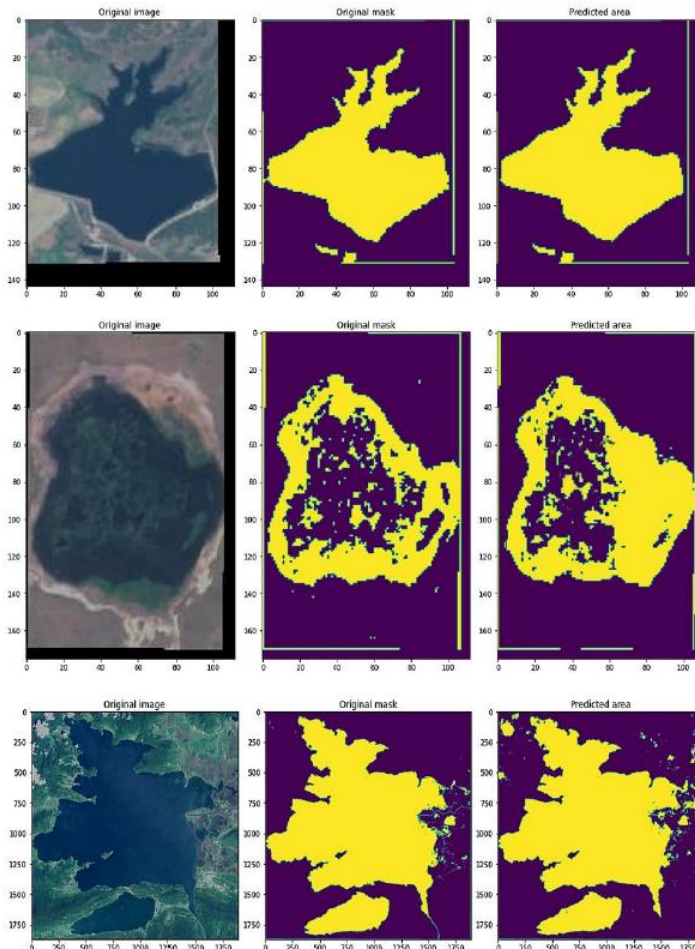


Figure 10. Output images with predicted mask and original mask

"To address challenges such as limited receptive area coverage, information loss, and aliasing effects, we introduce the NDR U-Net method with the following key strategies:

Mitigating Information Loss: We employ residual skip connections to alleviate the loss of information. These connections facilitate the reuse of features in each set of the method, enhancing the information flow within the layers.

Avoiding Aliasing Effects: To counter aliasing effects, we incorporate Nested Multi-Dilated Connections. These connections effectively control the dilation factor rate, contributing to a reduction in aliasing artifacts. **Enhancing Receptive Coverage Area:** We utilize dilated factors (i.e., 1, 2 and 4) to improve the receptive coverage area. Small dilation factors are strategically applied to extract local features, while larger dilation factors are employed for capturing global features. This approach ensures a more comprehensive coverage of the input space."

Table 2. Performance metrics

Parameter	Training	Testing
Loss	0.3115	0.1374
MAE	0.1214	0.1018
Accuracy	0.9699	0.9560
Precision	0.9777	0.9088
Recall	0.9681	0.8690
AUC	0.9914	0.9870
IOU	0.9536	0.9327

*MAE: Mean average Error

Table 3. Comparison of proposed method with previous state-of-the-art models

Model	Method	Backbone	IOU (%)
[37]	U-Net	—	85.58
[3]	RefineNet	Resnet-101	86.21
[38]	DeeplabV3+	Resnet-101	86.50
[39]	DANet	Resnet-101	87.90
[6]	CascadePSP	DeeplabV3+&Resnet-101	87.00
[16]	MECNet	MEC + MPF + DSFF	90.64
Proposed	NDRUNet	D3Net+Unet	95.36

Table 2 shows the performance metrics based on parameter and Table 3 shows the comparison of our model with other models. In the study [37], they used basic UNet architecture to segment the satellite images with higher accuracy of 85.58%. Chen et al. [38] used RefineNet with Resnet101 pretrained convolutional neural network (CNN) for classification and achieved the IOU of 86.21%. DeeplabV3 with Resnet101 is used for segmentation and prediction in the study [39] and they achieved an 86.50% of mean IOU. In DANet [39], utilizing the Resnet101 CNN model resulted in a mean Intersection over Union of 87.90%. CascadePSP has achieved an accuracy of 87.00% that is combination of DEEPLABV3 and Resnet101 CNN [6]. Zhang et al. [16] used MECNet with combination of MEC+MPF+DSFF for classification and achieved 90.64% of mean IOU. Our proposed work used the combination of UNet and NDRNet for the segmentation and classification achieved highest IOU of 95.36%, 96.99% of accuracy, 97.77% of precision, and 96.81% of recall.

6. CONCLUSIONS

"To address challenges such as limited receptive area

coverage, information loss [40], and aliasing effects, we introduce the NDR U-Net method. Even for tiny and hazy satellite Images, this model performs well. Additionally, NDRUNet's architecture ensures that it produces the best results for water bodies that are close to a land boundary. The model also makes use of an Adam optimizer. With this optimizer, computation times are comparatively swift. Additionally, since they affect how well the model works and performs, actions like scaling, conversion to masks, and padding are closely followed while doing data analysis. Multiple channels and feature maps are used in the proposed NDRUNet architecture. Additionally, kernel initializers, an output activation feature, and concealed activation features are utilized. To resolve the dimensionality of the images and improve speed, the model employs max-pooling. The following model makes better use of satellite images to pinpoint the expected region. Additionally, it performs best on images taken under poor weather conditions. The model is superior in that it secures the identified water area using ECDSA and generates a secured signature to keep it safe from unwanted access. This model can eventually be expanded to accept video input. The cloud may be used to store the observed water locations. This model achieved higher IOU of 93.27%, accuracy of 95.60%, precision of 90.88%, and recall of 86.90%.

DATA AVAILABILITY

The dataset used in this work is available on Kaggle as Satellite images of water bodies <https://www.kaggle.com/datasets/franciscoescobar/satellite>.

REFERENCES

- [1] Liu, Y., Zhang, Q.N., Wang, F.P., Kiang, C.T., Pang, L.K., Zhang, H.C., Lu, C., Lu, G.J., Nam, L. (2021). Adaptive weights learning in CNN feature fusion for crime scene investigation image classification. *Connection Science*, 33(3): 719-734. <https://doi.org/10.1080/09540091.2021.1875987>
- [2] Ye, A., Zhou, X., Miao, F. (2022). Innovative hyperspectral image classification approach using optimized CNN and ELM. *Electronics*, 11(5): 775. <https://doi.org/10.3390/electronics11050775>
- [3] Lin, G., Milan, A., Shen, C., Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1925-1934. <https://doi.org/10.48550/arXiv.1611.06612>
- [4] Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M. (2019). Basnet: Boundary-aware salient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7479-7489.
- [5] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N. (2018). Learning a discriminative feature network for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1857-1866. <https://doi.org/10.48550/arXiv.1804.09337>
- [6] Cheng, H.K., Chung, J., Tai, Y.W., Tang, C.K. (2020). Cascadepsp: Toward class-agnostic and very high-

- resolution segmentation via global and local refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8890-8899. <https://doi.org/10.48550/arXiv.2005.02551>
- [7] Yu, L., Wang, Z., Tian, S., Ye, F., Ding, J., Kong, J. (2017). Convolutional neural networks for water body extraction from Landsat imagery. *International Journal of Computational Intelligence and Applications*, 16(1): 1750001. <https://doi.org/10.1142/S1469026817500018>
- [8] Miao, Z., Fu, K., Sun, H., Sun, X., 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. <https://doi.org/10.1109/LGRS.2018.2794545>
- [9] 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. <https://doi.org/10.3390/rs11101162>
- [10] Duan, L., Hu, X. (2019). Multiscale refinement network for water-body segmentation in high-resolution satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, 17(4): 686-690. <https://doi.org/10.1109/LGRS.2019.2926412>
- [11] Guo, H., He, G., Jiang, W., Yin, R., Yan, L., Leng, W. (2020). A multi-scale water extraction convolutional neural network (MWEN) method for GaoFen-1 remote sensing images. *ISPRS International Journal of Geo-Information*, 9(4): 189. <https://doi.org/10.3390/ijgi9040189>
- [12] Wang, L., Bie, W., Li, H., Liao, T., Ding, X., Wu, G., Fei, T. (2022). Small water body detection and water quality variations with changing human activity intensity in wuhan. *Remote Sensing*, 14(1): 200. <https://doi.org/10.3390/rs14010200>
- [13] Yang, F., Feng, T., Xu, G., Chen, Y. (2020). Applied method for water-body segmentation based on mask R-CNN. *Journal of Applied Remote Sensing*, 14(1): 014502-014502. <https://doi.org/10.1117/1.JRS.14.014502>
- [14] Lira, J. (2006). Segmentation and morphology of open water bodies from multispectral images. *International Journal of Remote Sensing*, 27(18): 4015-4038. <https://doi.org/10.1080/01431160600702384>
- [15] Lalchhanhima, R., Saha, G., Sur, S.N., Kandar, D. (2021). Water body segmentation of synthetic aperture radar image using deep convolutional neural networks. *Microprocessors and Microsystems*, 87: 104360. <https://doi.org/10.1016/j.micpro.2021.104360>
- [16] Zhang, Z., Lu, M., Ji, S., Yu, H., Nie, C. (2021). Rich CNN features for water-body segmentation from very high resolution aerial and satellite imagery. *Remote Sensing*, 13(10): 1912. <https://doi.org/10.3390/rs13101912>
- [17] Shang, Y. (2022). Efficient and secure algorithm: The application and improvement of ECDSA. In 2022 International Conference on Big Data, Information and Computer Network (BDICN), pp. 182-188. <https://doi.org/10.1109/BDICN55575.2022.00043>
- [18] Yang, X., Liu, M., Au, M.H., Luo, X., Ye, Q. (2022). Efficient verifiably encrypted ECDSA-like signatures and their applications. *IEEE Transactions on Information Forensics and Security*, 17: 1573-1582. <https://doi.org/10.1109/TIFS.2022.3165978>
- [19] Jing, J., Wang, Z., Rättsch, M., Zhang, H. (2022). Mobile-Unet: An efficient convolutional neural network for fabric defect detection. *Textile Research Journal*, 92(1-2): 30-42. <https://doi.org/10.1177/0040517520928604>
- [20] Wang, Q., Ihme, M., Chen, Y.F., Anderson, J. (2022). A TensorFlow simulation framework for scientific computing of fluid flows on tensor processing units. *Computer Physics Communications*, 274: 108292. <https://doi.org/10.1016/j.cpc.2022.108292>
- [21] Sharma, J., Soni, S., Paliwal, P., Saboor, S., Chaurasiya, P.K., Sharifpur, M., Khalilpoor, N., Afzal, A. (2022). A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India. *Energy Science & Engineering*, 10(8): 2909-2929. <https://doi.org/10.1002/ese3.1178>
- [22] Erfani, S.M.H., Wu, Z., Wu, X., Wang, S., Goharian, E. (2022). ATLANTIS: A benchmark for semantic segmentation of waterbody images. *Environmental Modelling & Software*, 149: 105333. <https://doi.org/10.1016/j.envsoft.2022.105333>
- [23] Rajyalakshmi, C., Mohan Rao, K.R., Rao, R.R. (2022). Compressed high resolution satellite image processing to detect water bodies with combined bilateral filtering and threshold techniques. *Traitement du Signal*, 39(2): 669-675. <https://doi.org/10.18280/ts.390230>
- [24] Lin, H.Y., Hsieh, M.Y., Li, K.C. (2022). A fast fault-tolerant routing with Ecdsa signature protocol for internet of vehicles. In Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, pp. 517-529. https://doi.org/10.1007/978-981-16-2102-4_48
- [25] Chen, F., Wang, N., Yu, B., Wang, L. (2022). Res2-Unet, a new deep architecture for building detection from high spatial resolution images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 1494-1501. <https://doi.org/10.1109/JSTARS.2022.3146430>
- [26] Kaplan, G., Avdan, U. (2017). Object-based water body extraction model using Sentinel-2 satellite imagery. *European Journal of Remote Sensing*, 50(1): 137-143. <https://doi.org/10.1080/22797254.2017.1297540>
- [27] Images and masks of satellite images of water bodies for image segmentation. (2019). <https://www.kaggle.com/datasets>
- [28] Caballero, I., Ruiz, J., Navarro, G. (2019). Sentinel-2 satellites provide near-real time evaluation of catastrophic floods in the west mediterranean. *Water*, 11(12): 2499. <https://doi.org/10.3390/w11122499>
- [29] Jiang, W., Ni, Y., Pang, Z., Li, X., Ju, H., He, G., Lv, J., Yang, K., Fu, J., Qin, X. (2021). An effective water body extraction method with new water index for sentinel-2 imagery. *Water*, 13(12): 1647. <https://doi.org/10.3390/w13121647>
- [30] Campos, J.C., Sillero, N., Brito, J.C. (2012). Normalized difference water index have dissimilar performances in detecting seasonal and permanent water in the Sahara-Sahel transition zone. *Journal of Hydrology*, 464-465: 438-446. <https://doi.org/10.1016/j.jhydrol.2012.07.042>
- [31] Gao, B.C. (1995). Normalized difference water index for remote sensing of vegetation liquid water from space. In *Imaging Spectrometry*, 2480: 225-236. <https://doi.org/10.1117/12.210877>
- [32] Ding, Y., Chen, F., Zhao, Y., Wu, Z., Zhang, C., Wu, D. (2019). A stacked multi-connection simple reducing net

- for brain tumor segmentation. *IEEE Access*, 7: 104011-104024.
<https://doi.org/10.1109/ACCESS.2019.2926448>
- [33] Takahashi, N., Mitsufuji, Y. (2020). D3net: Densely connected multidilated densenet for music source separation. arXiv preprint arXiv:2010.01733. <https://doi.org/10.48550/arXiv.2010.01733>
- [34] Şen, S.Y., Özkurt, N. (2020). Convolutional neural network hyperparameter tuning with Adam optimizer for ECG classification. 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey, pp. 1-6. <https://doi.org/10.1109/ASYU50717.2020.9259896>
- [35] Ch, A., Ch, R., Gadamsetty, S., Iwendi, C., Gadekallu, T. R., Dhaou, I.B. (2022). ECDSA-based water bodies prediction from satellite images with UNet. *Water*, 14(14): 2234. <https://doi.org/10.3390/w14142234>
- [36] Oh, H., Nam, K., Jeon, S., Cho, Y., Paek, Y. (2021). MeetGo: A trusted execution environment for remote applications on FPGA. *IEEE Access*, 9: 51313-51324. <https://doi.org/10.1109/ACCESS.2021.3069223>
- [37] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [38] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-decoder with Atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818. <https://doi.org/10.48550/arXiv.1802.02611>
- [39] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H. (2019). Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3146-3154.
- [40] Jonnala, N.S., Gupta, N.(2024). SAR U-Net: Spatial attention residual U-Net structure for water body segmentation from remote sensing satellite images. *Multimed Tools Appl*, 83: 44425–44454. <https://doi.org/10.1007/s11042-023-16965-8>.