Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Innovative Strategies for Automated Water-Body Segmentation: Harnessing Deep Learning in Satellite Image Analysis

Vasamsetty Lakshmi Prasanna
Student, M Tech in Embedded systems
Department of ECE
Aditya College of Engineering, Surampalem, AP
Vasamsettylaxmiprasanna99@gmail.com

Dr.R.Raman
Associate Professor
Department of ECE
Aditya College of Engineering, Surampalem, AP
ramanphdr@gmail.com

ABSTRACT: Automatic segmentation of water bodies from high-resolution satellite images is of great importance in various fields, including environmental monitoring, urban planning, and disaster management. Traditional techniques for water body segmentation often rely on manual or semi-automatic methods, which are time-consuming and prone to human errors. Recently, deep learning approaches have shown remarkable success in computer vision tasks, leading to their adoption for water body segmentation. This study presents a comprehensive investigation into the development of a deep learning framework for automatically segmenting water bodies from high-resolution satellite images. The proposed framework leverages convolutional neural networks (CNNs) to learn distinctive features and achieve precise segmentation outcomes. To assess the framework's performance, a diverse dataset is employed, and the obtained results are compared with state-of-the-art methods. The experimental findings demonstrate the efficacy and efficiency of the proposed approach, offering encouraging possibilities for automating water body segmentation from high-resolution satellite images.

Keywords: automatic segmentation, deep learning, convolutional neural networks, high-resolution satellite images, water-body segmentation.

I. INTRODUCTION

1.1 Background:

In this section, the background of the research topic is presented to provide context and rationale for conducting the study on automatic water-body segmentation from high-resolution satellite images via deep networks. The increasing availability of high-resolution satellite imagery has opened up new possibilities for various applications, including environmental monitoring, urban planning, and disaster management. Accurate identification and segmentation of water bodies within these images are crucial for analyzing changes in water resources, understanding urban development patterns, and responding effectively to natural disasters. Traditional methods for water-body segmentation often rely on manual or semi-automatic techniques, which are labor-intensive and prone to errors. Therefore, there is a need for automated approaches that can handle the analysis of large-scale satellite images efficiently and accurately.

1.2 Problem Statement:

Within this section, a precise articulation of the targeted issue addressed by the research is provided. The problem statement elucidates the constraints and obstacles entailed by conventional methodologies employed in water-body segmentation, emphasizing the necessity for an automated approach utilizing advanced deep learning techniques. The problem statement may encompass factors such as the intricacy inherent in satellite imagery, the imperative for precise demarcation of water boundaries, and the potential repercussions arising from erroneous segmentation in subsequent applications.

1.3 Objectives:

The objectives of the research are outlined in this section. The main objective is to develop a deep learning-based framework for automatic water-body segmentation from high-resolution satellite images. The sub-objectives may include:

1.3.1 Investigating state-of-the-art deep learning techniques for image segmentation.

Image segmentation plays a crucial role in various computer vision tasks, including object detection, image recognition, and scene understanding. With the rapid advancements in deep learning, particularly convolutional neural networks (CNNs), there has been significant progress in achieving accurate and efficient image segmentation. In this section, we delve into the investigation of state-of-the-art deep learning techniques for image segmentation, highlighting their key concepts, methodologies, and contributions. *U-Net, DeepLab, Mask R-CNN, PSPNet, FCN.*

II. LITERATURE REVIEW

2.1 Overview of Water-Body Segmentation:

In this section, an overview of water-body segmentation is provided. It discusses the significance of water-body segmentation in various applications such as

environmental monitoring, urban planning, and disaster management. The section also highlights the challenges involved in water-body segmentation, including the complexity of satellite imagery, variability in water body appearances, and the need for accurate boundary delineation.

2.2 Traditional Methods for Water-Body Segmentation:

This section focuses on the traditional methods that have been used for water-body segmentation. It provides an overview of different techniques, such as thresholding, region-growing, and edge-based methods, which have been commonly employed for water-body segmentation. The strengths and limitations of these traditional methods are discussed, emphasizing their reliance on handcrafted features and the need for manual intervention.

2.3 Deep Learning Techniques for Image Segmentation:

Within this section, attention is redirected towards the utilization of advanced deep learning methodologies for image segmentation. The segment commences with an elucidation of deep learning, with a specific emphasis on convolutional neural networks (CNNs), which have brought about a transformative shift in the realm of computer vision. The foundational principles underlying CNNs, encompassing convolutional layers, pooling layers, and fully connected layers, are comprehensively explicated. Furthermore, prominent architectures employed for image segmentation, including U-Net and Fully Convolutional Networks (FCNs), are thoroughly examined and discussed.

2.4 Deep Learning for Water-Body Segmentation:

This section delves into the application of deep learning techniques specifically for water-body segmentation. It highlights recent research studies and methodologies that have used deep learning for automatic water-body segmentation. The section discusses the advantages of deep learning approaches, such as their ability to learn discriminative features from large amounts of data, their flexibility in handling complex image characteristics, and their potential for achieving high segmentation accuracy. Various strategies employed in deep learning for water-body segmentation, including data augmentation, transfer learning, and ensemble methods, are also explored.

The detection of water bodies using remote sensing imagery holds significant importance in urban hydrological studies [1]. The field of urban hydrology has emerged as a crucial area of research, aiming to enhance and manage urban water systems in order to address environmental challenges arising from rapid urbanization. It plays a vital role in facilitating effective flood protection planning, water quality control, and ensuring public safety and health [2]. To gain comprehensive insights into water systems within cities, accurate and automated detection of water bodies serves as the initial and fundamental stage, enabling pixel-level identification of water regions [3], [4].

Ever since its inception in 2015, the Sentinel-2 satellite has bestowed the scientific community with unrestricted access to a plethora of multispectral imagery, garnering extensive employment in land-cover applications [5]–[7]. The accessibility of Sentinel-2 data presents one of the most fitting resources for prompt surveillance and examination of urban hydrological processes. This is predominantly owing to its nearly daily refresh rate, surpassing that of more detailed remote sensing data like very high spatial resolution (VHR) imagery [8] and synthetic aperture radar (SAR) data [4]. The utilization of Sentinel-2 imagery enables researchers and hydrologists to leverage its frequent revisit rate, allowing for timely detection and monitoring of water bodies within urban areas. This is particularly crucial in urban hydrological studies, where up-to-date information is essential for understanding the dynamic nature of water systems in response to various factors such as rainfall, land use changes, and infrastructure development. By harnessing the multispectral capabilities of Sentinel-2, which capture information across different parts of the electromagnetic spectrum, accurate identification and delineation of water bodies can be achieved.

Accurate water body detection using Sentinel-2 imagery not only provides valuable insights into urban hydrological processes but also contributes to the development of effective strategies for water resource management, flood risk assessment, and urban planning. The availability of near-daily updates from the Sentinel-2 satellite ensures that the hydrological conditions can be continuously monitored, allowing for the identification of potential risks and the implementation of proactive measures to mitigate flooding and maintain water quality. In this study, our focus is on exploring the potential of 10-meter-resolution multispectral data from the Sentinel-2 satellite for urban hydrological applications that require frequent data updates. Traditional methods for detecting water bodies rely on handcrafted statistical features extracted from multispectral imagery, specifically near infrared (NIR) and short-wave infrared (SWIR) bands. These features include widely known indices such as the Normalized Difference Water Index (NDWI) [9], Normalized Difference Moisture Index (NDMI) [10], modified Normalized Difference Water Index (MNDWI) [11], Automated Water Extraction Index (AWEI) [12], and Pixel Region Index (PRI) [13]. While these methods demonstrate satisfactory performance in controlled datasets, their applicability in real-world conditions for water body detection is limited.

In our investigation, we aim to assess the effectiveness of utilizing Sentinel-2's multispectral data with its 10-meter resolution in urban hydrological applications. The advantage of Sentinel-2 lies in its capability to provide frequently updated data, which is crucial for monitoring dynamic urban water systems. By exploiting the temporal dimension of the data, we can gain valuable insights into the temporal variability of water bodies, capturing changes that occur over time due to rainfall, urban development, and other factors. Our research aims to contribute to the advancement of urban hydrological studies by providing improved methodologies for water body detection. By leveraging the frequent updates and high-resolution multispectral data from Sentinel-2, we anticipate that our proposed approach will enhance the accuracy and reliability of water body detection in real-world scenarios. This will enable better understanding and management of urban water systems, facilitating effective flood protection planning, water quality control, and urban development strategies.

Over the past few years, deep convolutional neural network (DCNN) architectures have garnered significant acclaim for their remarkable proficiency in detecting water bodies [3], [4], [14]–[16]. Prominent examples of such models include fully convolutional networks (FCNs) [17], upsampling pyramid networks [4], and DenseNet [18], which have proven their efficacy in urban hydrological applications by facilitating semantic segmentation of remotely sensed images and precise identification of water bodies. In comparison to conventional water index features, DCNN models possess a distinct advantage of extracting more distinctive and discriminative representations, thereby enhancing the accuracy of water body detection.

The use of multispectral imagery in water body segmentation has the potential to further enhance the performance compared to using only RGB channels. Multispectral data captures information from additional bands across a wider range of the electromagnetic spectrum, providing valuable contextual information. However, recent studies [4], [20] have shown that the use of multiple bands does not always yield significant improvements. For instance, in the Kaggle Satellite Imagery Feature Detection challenge [20], methods incorporating all 20 available channels (including panchromatic, RGB, multispectral, and SWIR bands) achieved only marginal enhancements compared to models utilizing RGB bands alone.

III. PROPOSED MC-WBDN NETWORK MODEL

In this research paper, we present a novel approach called the Multichannel Water Body Detection Network (MC-WBDN). The main objective of MC-WBDN is to leverage the capabilities of multispectral imagery in order to enhance the performance of existing state-of-the-art deep convolutional neural network (DCNN) models for accurate water body segmentation. By incorporating multispectral data, MC-WBDN aims to capture a wider range of spectral information, beyond the traditional RGB channels, to improve the discrimination of water bodies from other land features. This additional spectral information can provide valuable cues for distinguishing water pixels based on their unique reflectance properties in different parts of the electromagnetic spectrum.

MC-WBDN builds upon the advancements made in DCNN models and extends them to exploit the rich information available in multispectral imagery. By integrating multiple channels into the network architecture, MC-WBDN enables the model to learn and utilize more comprehensive feature representations, resulting in enhanced accuracy and robustness in water body segmentation. Through the utilization of MC-WBDN, we aim to address the limitations of existing approaches and push the boundaries of water body detection. By leveraging the full potential of multispectral imagery, we anticipate significant improvements in the identification and delineation of water regions, which can have wide-ranging applications in fields such as hydrology, environmental monitoring, and urban planning.

The development and evaluation of MC-WBDN will involve extensive experiments and comparative analyses against state-of-the-art DCNN models and other existing methods. The performance of MC-WBDN will be assessed using various datasets and evaluation metrics to validate its effectiveness and potential for advancing the field of water body segmentation. Overall, the proposed MC-WBDN represents an innovative and promising approach to leverage multispectral imagery for improved water body detection. Through this research project, we aim to contribute to the advancement of accurate and reliable water body segmentation techniques, ultimately supporting various applications in water resource management and environmental studies.

3.1 Data Preprocessing and Data Augmentation

Prior to training the model, several preprocessing and data augmentation techniques are employed to optimize the effectiveness and computational efficiency of the model. The following procedures are applied:

3.1.1. Image splitting

To ensure efficient computation and memory usage while enabling parallel processing, the raster imagery is divided into manageable image blocks. This partitioning approach helps avoid the burden of handling large datasets and facilitates parallelization. Our proposed model operates on input sizes of 512 × 512 pixels for the NIR and RGB channels and 256 × 256 pixels for the SWIR channel.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Rather than directly splitting the entire multispectral image into patches of the required size, we initially divide it into blocks of 1024 × 1024 pixels for NIR and RGB, and 512 × 512 pixels for SWIR. This configuration allows for the implementation of distinct splitting strategies for training and testing purposes. In the training phase, a greater number of samples are necessary to address potential issues of overfitting. Therefore, we introduce an overlapping split of image blocks to generate additional training samples. This overlapping split strategy helps diversify the training dataset and improve the model's ability to generalize.
During the testing phase, patches are extracted from randomly sampled, non-overlapping blocks and used as input to our proposed model.

3.1.2 Cloud filtering and color normalization:
After conducting an initial examination of the spectral information in each band, we determine a heuristic threshold of 3000 to filter out cloudy areas with values exceeding this threshold. To ensure consistency, any values surpassing the threshold are capped.

To normalize the intensities (Xi) in each channel, we employ a normalization technique using the mean (μ) and standard deviation (σ) of the channel. This normalization process aids in standardizing the intensities across different channels, allowing for a more meaningful comparison and analysis. By dividing each intensity value by its corresponding mean and standard deviation, we bring the intensities to a common scale, facilitating subsequent computations and interpretation. The normalization formula can be expressed as follows, considering the intensities Xi in each channel:

$$Normalized\ Intensity\ (Ni) = (Xi - \mu) / \sigma$$

This normalization procedure ensures that the intensities within each channel are centered around zero with a standard deviation of one. By applying this normalization step, we can effectively account for variations in intensity levels and promote more reliable and accurate analysis of the data.

$$X_i = \frac{X_i - \mu(X_i)}{\sigma(X_i)}$$

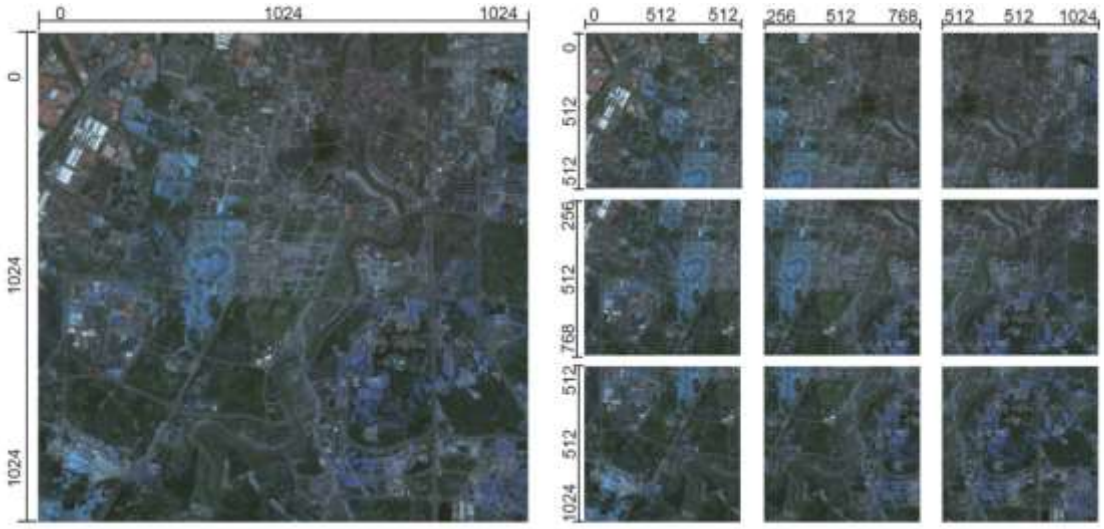


Figure 1: 1024 × 1024 image block (left) and the cropped patches extracted for training (right). The first and third values on each line segment denote the start and end pixel locations in the original image block, while the middle value denotes the length of the line segment

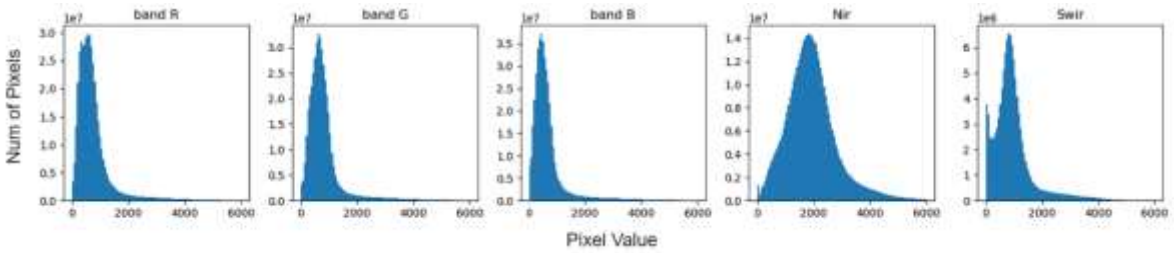


Figure 2: Pixel intensity distributions of the used bands.

3.1.3 Image augmentation

In conjunction with the initial patches acquired from the division stage, we incorporate the subsequent techniques during the training phase to augment the data: 1) a stochastic horizontal or vertical mirroring with a probability of 0.5; 2) a clockwise rotation of 90 degrees with a probability of 0.5; and 3) a fortuitous alteration of the HSV color space within a narrow spectrum, with a probability of 0.25.

3.2 MC-WBDN MODEL

3.2.1 Model Architecture

The architecture of our proposed MC-WBDN (Multichannel Water Body Detection Network) model is visually depicted. Our model takes three input images: the RGB channels, NIR channel, and SWIR channel. These channels are then processed by specific convolution kernels within the multichannel fusion module. This module generates feature maps of the same size for each channel, which are subsequently concatenated. The resulting fused feature maps serve as the input for our backbone encoder-decoder network, responsible for performing pixeXl-level labeling.

To construct our model, we utilize a ResNet-34 model that has been pretrained on the ImageNet dataset [26] as the encoder network. This choice allows us to leverage the prelearned weights and robust feature extraction capabilities of the ResNet architecture. For the decoder network, we employ an enhanced version of the DeepLabV3+ network. This decoder network incorporates fine-grained feature maps generated by the EASPP (Enhanced Atrous Spatial Pyramid Pooling) and S2D/D2S (Spatial-to-Depth and Depth-to-Spatial) modules.

The table outlines the kernel width, kernel height, and the number of kernels in each convolutional layer. Additionally, it presents the output sizes of the feature maps produced by these layers. This information provides crucial insights into the network's structure and serves as a reference for understanding the underlying design choices of our MC-WBDN model.

Our proposed model introduces two distinct features that differentiate it from a traditional backbone encoder-decoder architecture:

Instead of using bilinear upsampling operations, we employ D2S (Depth-to-Spatial) operations in our decoder. This modification enhances the information exchange between channels, drawing inspiration from successful techniques employed in SENet [40] and ShuffleNet [41]. By incorporating D2S operations, we improve the preservation of fine-grained context and achieve more effective upsampling.

In addition to the standard encoder-decoder connections, we incorporate two extra bypasses from lower layers of the encoder. These bypasses are concatenated with dense feature maps provided by an EASPP (Enhanced Atrous Spatial Pyramid Pooling) module. This fusion of bypasses and dense feature maps further enhances the preservation of fine-grained context and allows for more comprehensive feature representation. To ensure numerical stability and enable non-linear representation, we utilize Swish activation functions [42] in both the fusion head and the decoder. Swish activation functions, defined as $f(x) = x \cdot \text{sigmoid}(x)$, exhibit differentiability when handling negative gradients. In contrast, ReLU activation functions ($f(x) = \max(0, x)$) are employed in the encoder section to preserve the representative features transferred from the pretrained deep learning model.

During the testing phase, we employ a sliding window prediction mechanism. This mechanism involves sliding a window along the satellite imagery, and the result is determined by the central area of the window. This approach allows us to make predictions at different spatial locations while efficiently leveraging the model's capabilities. By incorporating these modifications and strategies, our proposed model achieves improved information exchange, preserves fine-grained context, and ensures numerical stability, leading to enhanced performance in water body detection tasks.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

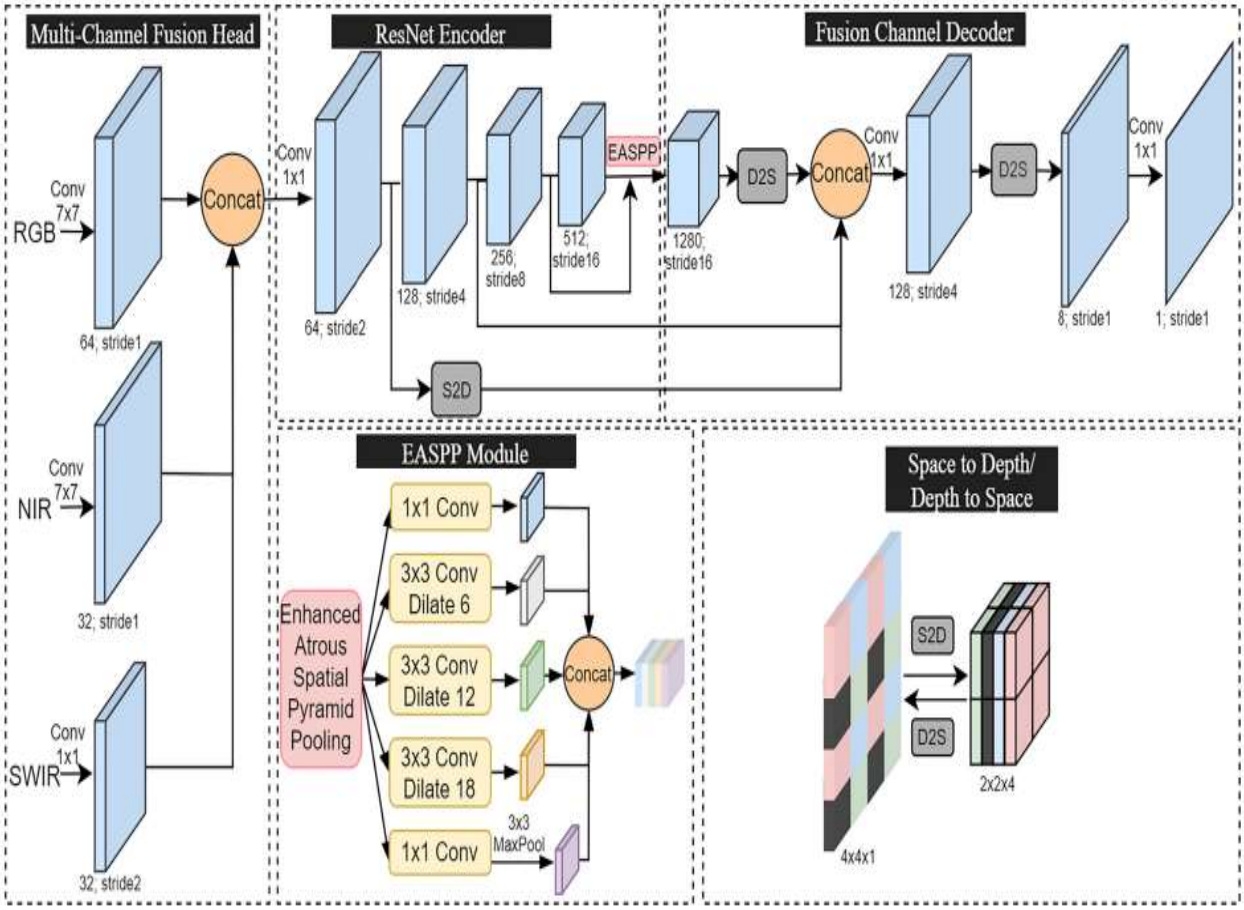


Figure 3: Proposed MC-WBDN network architecture, which adopts the popular encoder–decoder structure for semantic segmentation with a fusion head and works fully end-to-end. The baseline MC-WBDN model replaces S2D and D2S with corresponding pooling sub sampling and bilinear up sampling. Residual connections between each convolutional block are omitted. The number below each block denotes the channels of feature maps, whereas "stride2" indicates two times down sampled resolution compared to original input. The output is a single-channel feature map of the same size as the input.

Table 1: Layer configuration of employed network architecture

layer	MC-WBDN w/o S2D/D2S	MC-WBDN w/ S2D/D2S	output size		K_{width}	K_{height}	$K_{filters}$	non-linearities
fusion module	RGB		(256,256,64)		7	7	64	BN,Swish
	NIR		(256,256,32)		7	7	32	BN,Swish
	SWIR		(256,256,32)		1	1	32	BN,Swish
Concatenate			(256,256,128)					
Conv2d			(256,256,64)		1	1	64	BN,Swish
ResNet34-Block1			(256,256,64)		(3,3)	(3,3)	$64 \times 2 \times 3$	BN,ReLU
ResNet34-Block2			(128,128,128)		(3,3)	(3,3)	$128 \times 2 \times 4$	BN,ReLU
ResNet34-Block3			(64,64,256)		(3,3)	(3,3)	$256 \times 2 \times 6$	BN,ReLU
ResNet34-Block4			(32,32,512)		(3,3)	(3,3)	$512 \times 2 \times 3$	BN,ReLU
EASPP module			(32,32,1024)		(1,3,3,3,1)	(1,3,3,3,1)	512	BN
Upsample-4x		D2S+S2D	(128,128,1024)	(256,256,128)				
Concatenate			(128,128,1024+128)					
Conv2d			(128,128,128)		1	1	128	BN,Swish
Upsample-4x		D2S	(512,512,128)	(512,512,8)				
Conv2d			(512,512,16)	(512,512,1)	1	1	16	BN,Swish
Conv2d			(512,512,1)		1	1	1	

BN = Batch normalization.

3.2.2 Multichannel Fusion Head:

To integrate the RGB channels with the NIR and SWIR channels, we incorporate a fusion step at the

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

initial stage of our processing pipeline. Specifically, for the RGB and NIR channels that share the same resolution, we utilize 7x7 convolution kernels to expand the receptive field. Additionally, we employ a stride of 2 to ensure the output size aligns with that of the SWIR channel. In the case of the lower-resolution SWIR band, we employ 1x1 convolution kernels to increase the density of its feature maps. The outputs from these operations are then concatenated, followed by a 1x1 convolutional operation. This final convolutional operation yields the combined channel representation that is utilized by the context encoder module. By performing these steps, we effectively fuse the RGB, NIR, and SWIR channels, enabling the subsequent stages of the model to leverage the complementary information from each channel.

3.2.3 EASPP Module

In order to capture distinctive characteristics from a variety of receptive fields, we propose an enhancement to the Atrous Spatial Pyramid Pooling (ASPP) technique [36]. Our innovative approach, termed Enhanced ASPP (EASPP) module, introduces a modification to the original ASPP method. Instead of relying on upsampling operations, we employ 1x1 convolution operations followed by local max pooling on the feature maps derived from preceding layers. This modification establishes a direct connection from the previous layers, thereby amplifying the effectiveness of the trainable weights. By incorporating the EASPP module, we successfully extract dense features from diverse scales within the input feature maps [36]. This is achieved through the utilization of individual dilated convolutions [35], [43], [44] at varying scales, each scale representing a distinct region size that can be activated within the feature maps. The hierarchical arrangement of the receptive fields facilitates the aggregation of contextual information from the input, resulting in a comprehensive and informative feature representation. Lastly, the multiscale feature pyramid is concatenated and refined through a 1x1 convolution operation, generating output feature maps that can be further processed.

3.2.4. Space-to-Depth and Depth-to-Space

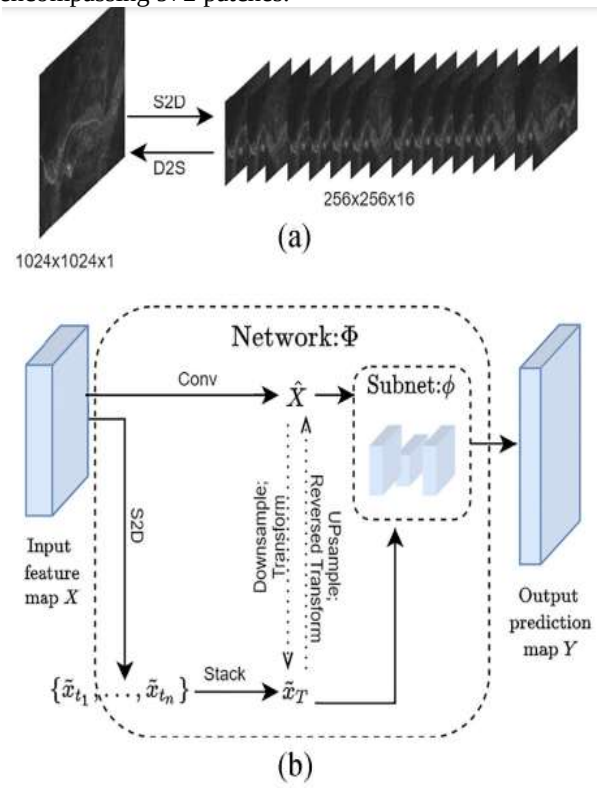
In the traditional workflow of a DCNN, the inclusion of pooling operations in the encoder and upsampling operations in the decoder is prevalent. Nonetheless, these operations entail certain limitations. Pooling operations have a tendency to discard fine-grained feature responses, while upsampling operations lack trainability. Although transposed convolution operations can serve as an alternative to upsampling, they significantly augment the parameter count within the DCNN [45]. To overcome these challenges, the adoption of S2D (Spatial-to-Depth) and D2S (Depth-to-Spatial) operations has been proposed as a viable solution [46], [47].

The S2D operationinvolves moving pixels from their original spatial locations to the channel dimension. This operation enables the preservation of more local features for the subsequent decoder process. Additionally, the S2D operation can be viewed as an intramodel augmentationproviding different views of inputs with varying pixel shifts. On the other hand, the D2S operation serves as an alternative to transposed convolutions for upsampling and offers two advantages. Firstly, it is parameter-free while retaining all the responses from the previous layers. Secondly, it facilitates information fusion across feature map channels, enabling effective feature exchange instead of focusing solely on individual channels as in the case of transposed convolutions.

IV. SIMULATION RESULS

4.1 Experimental Setup

To ensure an impartial assessment of the proposed methodology, the dataset is partitioned into distinct training, validation, and test sets. Initially, the satellite imagery undergoes division into 441 blocks, each with dimensions of 1024 × 1024. Consequently, nine patches are derived from every 1024 × 1024 block, yielding a cumulative count of 3969 image patches. For the purpose of training, a subset of 300 blocks, measuring 1024 × 1024, is employed, thereby resulting in 2700 training patches. The remaining 141 blocks are allocated for the formation of a validation set, comprising 33 blocks and encompassing 297 patches, while the remaining 108 blocks form the test set, encompassing 972 patches.



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, June : 2023

Figure 4. (a) Visualization of S2D and D2S. (b) Role of S2D in a network. Dashed arrows demonstrate the relationships between slices after S2D and the original branch but do not actually take effect when processing.

To assess the robustness of the proposed method, the training and validation sets are further divided into three folds. Each fold comprises 300 blocks for training and 33 blocks for validation. This division allows for testing the proposed method on different subsets of the data and provides insights into its performance across multiple scenarios. The mIoU (mean Intersection over Union) results, expressed in terms of average and standard deviation, are reported on the test set when trained using the three trained models. This comprehensive evaluation provides a reliable measure of the model's performance and its consistency across different training instances.

Figure 5: Input Image

Figure 6: Binary Image

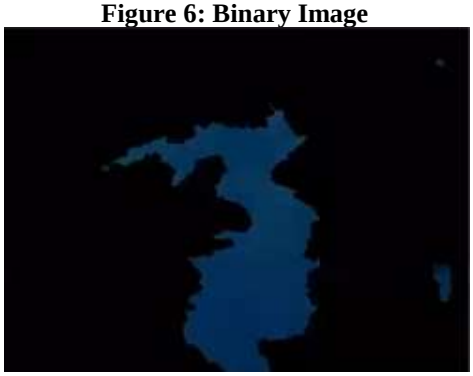


Figure 7: Output segmented image

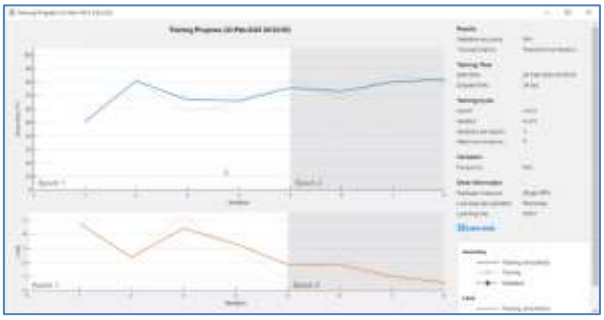


Figure 8: Training progress

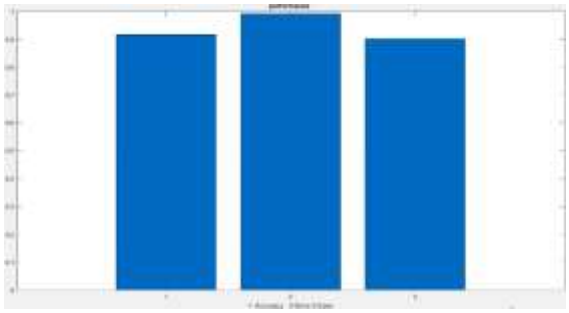


Figure 9: Performance metrics

V. CONCLUSION

Driven by the remarkable achievements of deep learning techniques in various fields, including remote sensing, we propose a novel approach for satellite-based water body extraction. In this article, we introduce an effective DCNN model that incorporates several key contributions. While the RGB channel has been widely utilized in remote sensing applications, we demonstrate the value of incorporating additional wavelength bands, such as NIR and SWIR, to enhance the segmentation accuracy. By leveraging the multispectral information provided by Sentinel-2 satellite data, we effectively exploit the unique characteristics of each band to improve the network's ability to identify water areas accurately.

In our future work, we aim to further explore the practical applications of our proposed method in hydrological studies. By leveraging the strengths of our MC-WBDN model, we anticipate enabling more accurate and reliable analysis of water bodies in various hydrological contexts.

REFERENCES

[1] Z. Shao, H. Fu, D. Li, O. Altan, and T. Cheng, “Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation,” Remote Sens. Environ., vol. 232, 2019, Art. no. 111338.

[2] X. Wang and H. Xie, “A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management,” Water, vol. 10, 2018, Art. no. 608.

[3] Z. Miao, K. Fu, H. Sun, X. Sun, and M. Yan, "Automatic water-body segmentation from high-resolution satellite images via deep networks," IEEE Geosci. Remote Sens. Lett., vol. 15, no. 4, pp. 602–606, Apr. 2018.

[4] J. Zhang et al., "Water body detection in high-resolution SAR images with cascaded fully-convolutional network and variable focal loss," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 1, pp. 316–332, Jan. 2021.

[5] A. B. Hamida et al., "Deep learning for semantic segmentation of remote sensing images with rich spectral content," in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2017, pp. 2569–2572.

[6] M.Wurm, T. Stark, X. X. Zhu,M.Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," ISPRS J. Photogrammetry Remote Sens., vol. 150, pp. 59–69, 2019.

[7] P. T. Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," Sensors, vol. 18, no. 1, 2018, Art. no. 18.

[8] G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 5, pp. 1633–1644, May 2018.

[9] S. K. McFeeters, "The use of normalized difference water index (NDWI) in the delineation of open water features," Int. J. Remote Sens., vol. 17, pp. 1425–1432, 1996.

[10] B.-C. Gao, "Normalized difference water index for remote sensing of vegetation liquid water from space," Proc. SPIE, vol. 2480, pp. 225–237, 1995.

[11] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," Int. J. Remote Sens., vol. 27, no. 14, pp. 3025–3033, 2006.

[12] G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud, "Automated water extraction index: A new technique for surface water mapping using landsat imagery," Remote Sens. Environ., vol. 140, pp. 23–35, 2014.

[13] Y. Zhang, X. Liu, Y. Zhang, X. Ling, and X. Huang, "Automatic and unsupervised water body extraction based on spectral-spatial features using GF-1 satellite imagery," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 6, pp. 927–931, Jun. 2019.

[14] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 11, pp. 2351–2355, Nov. 2015.

[15] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric SAR image classification using deep convolutional neural networks," IEEE Geosci. Remote Sens. Lett., vol. 13, no. 12, pp. 1935–1939, Dec. 2016.

[16] W. Feng, H. Sui, W. Huang, C. Xu, and K. An, "Water body extraction from very high-resolution remote sensing imagery using deep U-net and a superpixel-based conditional random field model," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 4, pp. 618–622, Apr. 2019.

[17] L. Li, Z. Yan, Q. Shen, G. Cheng, L. Gao, and B. Zhang, "Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks," Remote Sens., vol. 11, no. 10, 2019, Art. no. 1162.

[18] G. Wang, M. Wu, X. Wei, and H. Song, "Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks," Remote Sens., vol. 12, no. 5, 2020, Art. no. 795.

[19] K. Makantasis, A. Doulamis, N. Doulamis, and A. Voulodimos, "Common mode patterns for supervised tensor subspace learning," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2019, pp. 2927–2931.

[20] Kaggle, "DSTL satellite imagery feature detection." Accessed: Mar. 25, 2020. [Online]. Available: https://www.kaggle.com/c/dstl-satelliteimagery-feature-detection.

[21] J. Mukherjee, J. Mukherjee, and D. Chakravarty, "Automated seasonal separation of mine and non mine water bodies from landsat 8 OLI/TIRS using clay mineral and iron oxide ratio," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 7, pp. 2550–2556, Jul. 2019.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3431–3440.

[23] J. Guo, H. Zhou, and C. Zhu, "Cascaded classification of high resolution remote sensing images using multiple contexts," Inf. Sci., vol. 221, pp. 84–97, 2013.

[24] M. Wang, Y. Wan, Z. Ye, and X. Lai, "Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm," Inf. Sci., vol. 402, pp. 50–68, 2017.

[25] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, arXiv:1404.1869.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.

[27] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous convolution for semantic image segmentation," 2017, arXiv:1706.05587.
[28] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoderdecoder with Atrous separable convolution for semantic image segmentation," in Proc. Eur. Conf. Comput. Vis., 2018, pp. 833–851.
[29] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár, "Learning to refine object segments," in Proc. Eur. Conf. Comput. Vis., 2016, pp. 75–91.
[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention, 2015, pp. 234–241.
[31] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 5168–5177.
[32] L. Chan, M. S. Hosseini, and K. N. Plataniotis, "A comprehensive analysis of weakly-supervised semantic segmentation in different image domains," Int. J. Comput. Vis., vol. 129, pp. 361–384, 2021.
[33] A. Tao, K. Sapra, and B. Catanzaro, "Hierarchical multi-scale attention for semantic segmentation," 2020, arXiv:2005.10821.
[34] C. Peng, Y. Li, L. Jiao, Y. Chen, and R. Shang, "Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remotesensing image semantic segmentation," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 8, pp. 2612–2626, Aug. 2019.
[35] B. Yu, L. Yang, and F. Chen, "Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 9, pp. 3252–3261, Sep. 2018.
[36] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 6230–6239.
[37] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," ISPRS J. Photogrammetry Remote Sens., vol. 145, pp. 60–77, Nov. 2018.
[38] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2016, pp. 680–688.
[39] W. Sun and R. Wang, "Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM," IEEE Geosci. Remote Sens. Lett., vol. 15, no. 3, pp. 474–478, Mar. 2018.
[40] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 7132–7141.
[41] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 6848–6856.
[42] P. Ramachandran, B. Zoph, and Q. Le, "Searching for activation functions," 2017, arXiv:1710.05941.
[43] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in Proc. 4th Int. Conf. Learn. Representations, ICLR, San Juan, Puerto Rico, May 2016. [Online]. Avilable: http://arxiv.org/abs/ 1511.07122
[44] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global context based automatic road segmentation via dilated convolutional neural network," Inf. Sci., vol. 535, pp. 156–171, 2020.
[45] H. Ravishankar, R. Venkataramani, S. Thiruvenkadam, P. Sudhakar, and V. Vaidya, "Learning and incorporating shape models for semantic segmentation," in Proc. Int. Conf. Med. Image Comput.-Assisted Intervention, 2017, pp. 203–211.
[46] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," in Proc. 5th Int. Conf. Learn. Representations, ICLR, Toulon, France, Apr.2017. [Online]. Avilable: https://openreview.net/forum?id= HkpbnH9lx
[47] T.-J. Yang et al., "DeeperLab: Single-shot image parser," 2019, arXiv:1902.05093.
[48] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," in Medical Image Computing and Computer-Assisted Intervention. New York, NY, USA: Springer, 2017, pp. 674–682.
[49] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovasz-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4413–4421.
[50] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," in Nature, Nature Publishing Group, vol. 323, no. 6088, pp. 533–536, 1986.
[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Representations, ICLR, 2015. [Online]. Avilable: http://arxiv.org/abs/1412.6980

[52] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 192–194.

[53] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. Int. Conf. Neural Inf. Process. Syst., 2019, pp. 8024–8035.

[54] W. Zhu et al., "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," Med. Phys., vol. 46, no. 2, pp. 576–589, 2019.