Monitoring of Natural Resources from Satellite Images Using Deep Learning

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Application

By

Bazila Farooq

Registration number: 12021125

Supervised By

Dr. Ankush Manocha (24829)
School of Computer Applications
(Assistant Professor)
Lovely Professional University

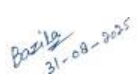
LOVELY PROFESSIONAL UNIVERSITY
PUNJAB
2026

Declaration of Authorship

I, Bazila Farooq, declare that this thesis titled, " MONITORING OF NATURAL RESOURCES FROM SATELLITE IMAGES USING DEEP LEARNIING " and the work presented in it are my own. I confirm that:

- This work is done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis have not previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:



(Dated: 31-1-2026)

Certification

This is to certify that the thesis entitled "Monitoring of Natural Resources from Satellite Images Using Deep Learning", which is being submitted by Ms. Bazila Farooq for the award of the degree of Doctor of Philosophy in Computer Application from the Faculty of Technology and Sciences, Lovely Professional University, Punjab, India, is entirely based on the work carried out by her under my supervision and guidance. The work reported, embodies the original work of the candidate and has not been submitted to any other university or institution for the award of any degree or diploma, according to the best of my knowledge.

Signed:

Dr. Ankush Manocha
Associate Professor
Department of Computer Science
Lovely Professional University
Phagwara, Punjab-144411, India
Date:31-1-2026

Abstract

Remote sensing (RS) technology offers invaluable insights into the Earth's dynamics by allowing a comprehensive view of the planet from a higher vantage point, rendering it an indispensable asset for various applications. With the proliferation of RS data and ongoing urban expansion, there arises a necessity for effective methodologies to harness RS data and its distinct attributes for the surveillance and administration of natural resources. With the launch of earth observation satellites, each capturing images at varying resolutions and revisit periods, the potential for RS to revolutionize field surveys and development is immense. Consequently, employing Deep Learning (DL) for RS applications has attracted much attention over the past few years. DL a powerful technique in Machine learning(ML) has demonstrated remarkable capabilities in numerous computer vision applications. Its potential in remote sensing is yet to be fully explored, but its ability to extract intricate patterns and features from high-resolution imagery holds immense promise for enhancing RS applications. The integration of DL into RS has the potential to revolutionize various aspects of earth observation. Its ability to handle complex datasets and extract meaningful information from high-resolution imagery can significantly improve the accuracy of object detection, monitoring resources, classification, and segmentation tasks. This can lead to more precise monitoring of urban development, water resources, and natural disasters. In addition, DL can be employed to develop algorithms for image fusion, which combines data from multiple sensors to create more comprehensive and detailed images. This could additionally bolster the functionalities of RS applications, furnishing a broader comprehension of the Earth's terrain and its evolving dynamics. The utilization of DL in RS is still in its early stages, but its potential is undeniable. As research progresses and computational resources continue to expand, Dl is poised to play a transformative role in the field of RS, providing a powerful tool for understanding and managing resources. Previous studies have

primarily focused on using high-resolution images to map, classify, and detect objects in urban areas. Nevertheless, the surveillance of water resources stands as a pivotal objective within the United Nations Sustainable Development Goals (SDGs). Field surveys are prohibitively expensive and inaccurate for monitoring water bodies in rural areas due to the lack of transportation. To address the limitations of field surveys in monitoring water resources, we explored the use of RS data and various deep-learning algorithms for analyzing satellite imagery. Our research revealed that among the different Dl approaches, CNNs exhibited superior performance for monitoring tasks. In this manner, two separate frameworks have been proposed for the extraction of small and other bodies of water. Ist framework focuses on extracting small water bodies from satellite images using a combination of multispectral false-color composite Sentinel-2 satellite images and advanced data processing techniques and 2nd one focuses on predicting water areas from spatio-spectral-temporal images. The primary objective is to provide a tool for real-time water monitoring that can assist governments and non-governmental organizations in managing water resources, which are crucial for maintaining a high standard of living. The future scope lies in implementing the learned insights in sustainable resource management, policy formulation, and global collaboration in addressing environmental challenges.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Dr. Ankush Manocha, for his supervision, advice, and guidance from the very first day of this research as well as for giving me extraordinary experiences throughout the work. I am truly very fortunate to have the opportunity to work with him. I found this guidance to be extremely valuable.

I am grateful to my friends and fellow researchers, for their constructive criticism and suggestions.

I would like to show my gratitude to the entire family of Lovely Professional University for providing me with a suitable research atmosphere to carry out my work in the proper time. I would like to thank the Division of Research and Development and the School of Computer Applications for all the support and encouragement throughout the research work.

I am also very much grateful to my mother, Mrs. Mubeena Banoo, and my father, Mr. Farooq Ahmad Ganie, my sister(Asifa Jan) who always listens to my story, my brother(Hashim Farooq) who always makes me smile, and my nieces(Mahira Mudasir , Maisha Mudasir) who always make my day.

Last but not least, I thank To Almighty Allah for sailing me through all the rough and tough times during this research work.

Bazila Farooq **Date = 31-1-2026**

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Description
RS	Remote Sensing
CD	Change Dectection
VHR	Very High Resolution
SAR	Synthetic Aperture Radar
TCW	Tasseled Cap Wetness Index
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
MNDWI	Modified Normalized Difference Water Index
AWEI	Automatic Water Extraction Index
ReLU	Rectified Linear Unit
ML	Machine Learning
DL	Deep Learning
UWI	Urban Water Index
USI	Urban Shadow Index
SVM	Support Vector Machine
DT	Decision Trees
FCN	Fully Convolutional Network
DCU	Decoder Unit
ECU	Encoder Unit
DDU	Dense Down Unit
PRU	Pyramid Unit
DUU	Dense Up Unit

BNU	Bottle Neck Unit
AM	Attention Modules
RGB	Red Green Blue
GPU	Graphical Processing Unit
COAH	Copernicus Open Access Hub

List of Symbols

Symbol	Description
$+$	Plus sign
$\_$	Subscript
$\hat{}$	Superscript
( )	Parentheses
-	Minus / dash
,	Comma
$\cdot$	Multiplication dot
$\times$	Multiplication sign
-	Fraction
$\sum$	Summation
$\odot$	Element-wise multiplication
$\imath$	Dotless i in math
$\| \|$	Norm bars

I would like to dedicate my thesis to my family and especially, one of my beloved people.

Chapter 1

Introduction

Water blankets approximately 71% of the Earth's surface, with only 3% of this substantial reserve categorized as freshwater. Surface water which includes lakes, rivers, wetlands, and other terrestrial water bodies accounts for just 0.3 of this freshwater category as depicted in FIGURE1.1(UPPU, 2020). This is the amount of freshwater found in inland aquatic systems. Despite its widespread disrepute water is an essential resource for both the natural world and human culture. It remains the most abundant of all natural resources having a tremendous impact on our globe and supporting a wide range of ecosystems, economic activities, and cultural functions (Ogidi et al., 2022) . Additionally, as emphasized by the Global Climate Observing System in 2011, it holds a crucial significance in the examination of climate change (Pekel et al., 2014). Water also plays an important role in habitat protection and biodiversity promotion (Palmer et al., 2015). Water plays a vital role in the distribution of nutrients and the global carbon equilibrium on a worldwide scale. Water is used in a variety of activities in people's everyday lives, including agricultural irrigation, potable water supply, energy generation in power plants, leisure and outdoor activities, fishing, ecotourism, transportation, religious rites, and more (Panneer Selvam et al., 2014). In the years ahead, water resources will encounter increasing pressures stemming from a convergence of factors, such as more frequent droughts, urban sprawl, urban population growth, deforestation, intensified use of fertilizers and pesticides, and the spread of invasive species. Additionally, the ongoing shifts in global climate patterns are anticipated to exacerbate the degradation of water quality.

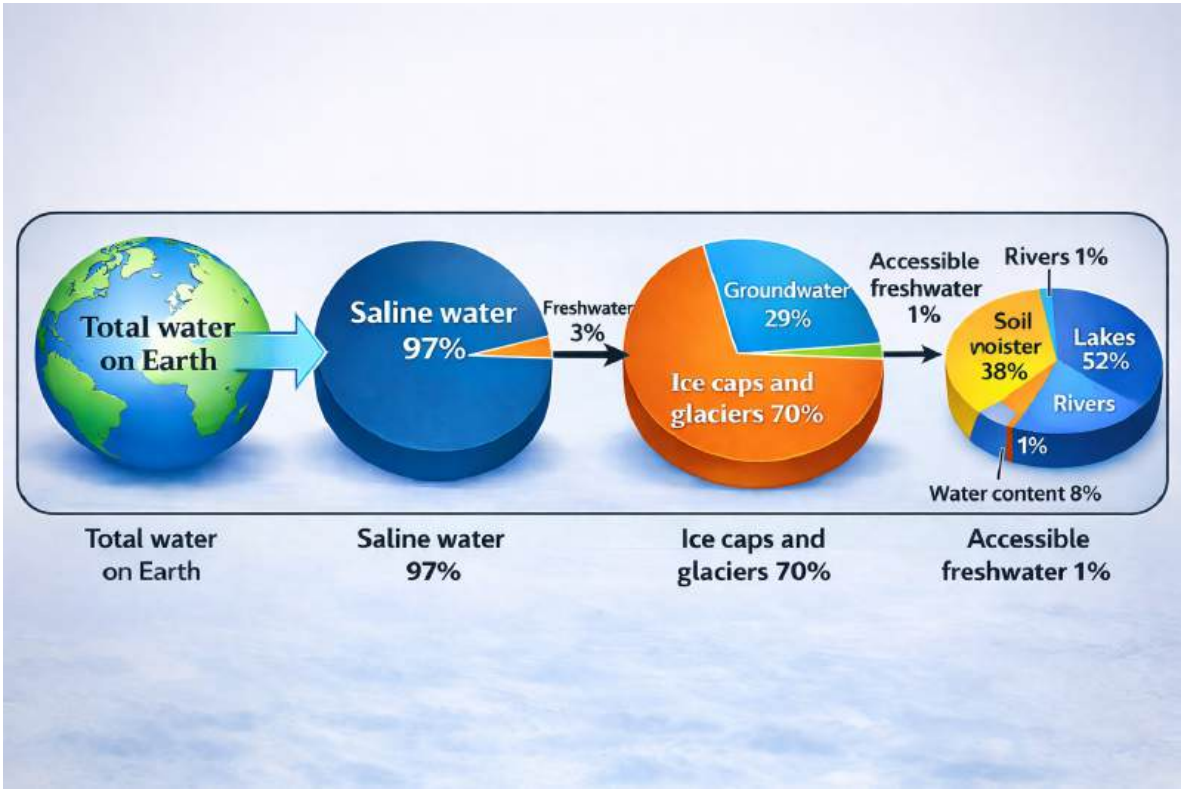


FIGURE 1.1: Earth's Water Distribution(UPPU, 2020)

Surface water bodies inland are typically divided into two primary categories using traditional classification methods: (1) perennial surface water and (2) intermittent surface water. Permanent surface water refers to water bodies that maintain a substantial portion of their volume consistently throughout the year without drying up. The hydro-meteorological cycle, which includes elements such as rainfall, evaporation, and the potential of streams overflowing, has the greatest impact on their changes. Rivers and lakes are common examples of these permanent surface water features. Rivers are defined as flowing watercourses (Kuhn, 2021), whereas lakes are defined as non-moving, enclosed water bodies confined inside a basin that is distinct from the sea (Heydari, 2021). During periods of little or no rainfall, permanent surface water bodies water levels are maintained by contributions from groundwater sources. Grasping the scope and arrangement of permanent

surface water poses a worldwide hurdle due to their varying sizes, spanning from diminutive entities like ponds to immense ones like the Great Lakes. This leads to disparities in detection and cataloging across extensive geographical regions. Moreover, pinpointing permanent surface water holds significance for responding to severe droughts and administering water reserves. These maps can also aid in validating hydraulic modeling (Matgen et al., 2010). Intermittent water comprises sections of water areas that undergo repetitive dry periods of varying lengths. These regions comprise minor ponds, puddles, marshlands, or areas prone to recurrent flooding.

Flood events can be categorized into different types based on their size, duration, and underlying causes. Intermittent surface water refers to stretches of water that experience repeated dry spells of varied lengths. These locations include tiny ponds, puddles, marshes, and areas prone to floods. Flood episodes are classified according to their magnitude, length, and underlying causes. Plain floods are more common in level terrain. They are caused by prolonged periods of rain or the melting of snow and ice, depending on soil saturation. When river embankments fail to withstand flood discharge, extensive areas are often submerged (Hossain et al., 2021). This sort of flood can linger for a long time, impacting large areas. Flash floods, on the other hand, are caused by intense and severe rainfall. These floods are frequently described as unexpected and limited occurrences that affect smaller areas. Flash floods occur in confined hilly basins due to heavy rain, shallow soils, and rapid runoff, sometimes intensified by lahars near volcanoes. Coastal flooding, driven by rising seas, tides, tsunamis, or land subsidence, inundates large coastal areas. Flood impacts vary with geography, erosion, and natural or artificial defenses. When surface waters cause floods, they can endanger human populations, towns, and infrastructure (Feyisa et al., 2014). Floods are often regarded as the most expensive form of hazard owing to property destruction and loss of life (Baci, 2023).

Data obtained from Earth Observation (EO) using optical or Synthetic Aperture Radar (SAR) sensors offers a valuable and advantageous means for mapping and ongoing surveillance of surface water resources in various research scenarios, encompassing both permanent and temporary water

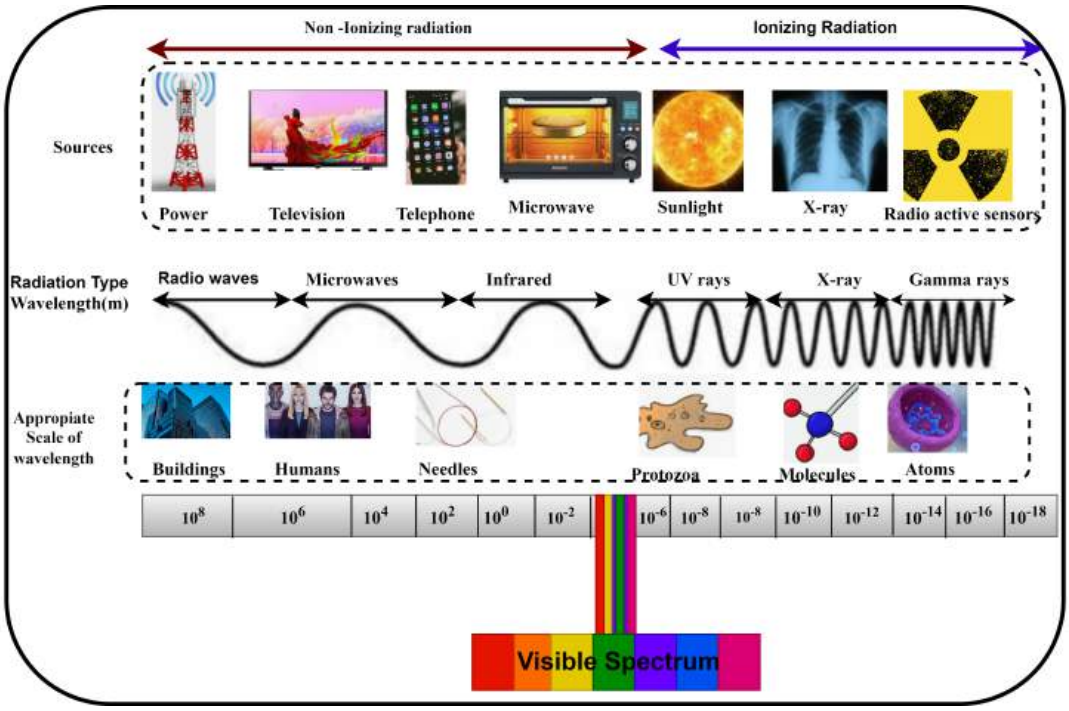


FIGURE 1.2: The Electromagnetic Spectrum

bodies.

1.0.1 Remote sensing of water

Remote sensing is the instantaneous capture of data from distant objects or locations, generally in the form of images, without the requirement for human involvement. Remote sensing imagery is a type of image that is used in the field of remote sensing to address difficulties linked to earth observation. Sensors intended to capture remote sensing images collect data from a distance by using electromagnetic radiation or acoustic energy and are illustrated in FIGURE 1.3. While the electromagnetic spectrum has many wavelengths, not all of them are equally useful for distant sensing FIGURE 1.2. Specific wavelengths from the electromagnetic spectrum are chosen based on the properties of the materials being targeted, allowing for accurate object targeting. The human eye acting as a true distant sensor sees objects by utilizing the visible spectrum. However ultraviolet and infrared wavelengths are

used for the bulk of remote sensing imaging. Sensors are different devices that capture the movement of things by utilizing the electromagnetic spectrum, such as cameras and scanners. In addition, a mode of transportation is used to move the sensors from one site to another. These carriers might take the shape of ground-level platforms, aerial platforms, or space-based platforms.

Furthermore, satellite images are captured utilizing a variety of sensors. Remote sensing images, whether from natural or human activity, provide complete global observations and significant insights into ecosystem health and sustainability. There were around 700 satellites in the year 2020 that were largely dedicated to the duty of Earth observation. The continual improvements in satellite imagery's spatial and temporal resolutions have substantially boosted our ability to distinguish minute details and characteristics of our world. The World View-2 satellite stands out with its remarkable revisit rate of 1.1 days for the Earth's surface, allowing it to capture an extensive daily data coverage of approximately 1 million square kilometers using its 8-band imaging technology. These bands offer a panchromatic resolution of 0.46 meters and a multispectral resolution of 1.84 meters (Xiong et al., 2021). Each day, the Sentinel-2 satellite captures an extensive 6 terabytes of data, resulting in the compilation of a comprehensive image of the entire Earth every five days (Phiri et al., 2020). From the 1970s onward, Multi-Spectral images (MSIs) have been employed in remote sensing applications owing to their abundance of spectral bands.

Concurrently, spectrometry is referred to as Hyperspectral Imaging (HSI), and these technologies are witnessing increasing recognition across a range of remote sensing applications. Unlike multispectral imagery, HSI's typically comprise hundreds or even thousands of bands, each with a much narrower spectral bandwidth, often ranging from 10 to 20 nanometers. In a hyperspectral image, each pixel can be seen as a multi-dimensional vector representing spectral reflectance across numerous closely spaced spectral channels within a specific wavelength range. Modern HSI acquisition systems offer not just impressive spectral precision but also remarkable spatial resolution. MSI and HSI data may both express intricate qualities as well as more detailed

spectral and geographical information. In recent years, earth observation scientists and specialists have identified hyperspectral imaging as the most effective and exact technology for handling varied applications such as identifying unique materials and recognizing things (Pawar et al., 2021,Yang et al., 2020). In recent years, researchers have developed a variety of computer algorithms for recognizing objects in remote sensing data. The primary goal of remote sensing is to automate the collection of demographic data using satellite images, which was formerly done manually. Survey data collection is vital not only in India but also in other countries, acting as a crucial information source for decision-makers. The gathering of survey data needs significant resources and time to acquire critical information, a task that is worsened in places with inadequate communication and transportation facilities.

In recent years, remote sensing has arisen as a solution to these difficulties. Remote sensing technology has found use in a variety of disciplines, including but not limited to classification, change detection, and drought monitoring (Xiong et al., 2021). Nonetheless, using satellite images for monitoring and change detection necessitates specific preprocessing steps. Remote sensing advances especially VHR satellite imagery have streamlined environmental monitoring by improving change detection, segmentation, and classification. High-resolution images enable precise tracking of land use shifts, vegetation changes, and water quality indicators, aiding water resource management and conservation. Despite the challenges of real-time processing, deep learning methods increasingly enhance environmental monitoring and Earth observation.

1.0.2 Significance of Deep learning in Remote Sensing

Deep learning has been a vital artificial intelligence technology for the past ten years, especially for extracting important data from satellite images. To make it easier to extract useful characteristics, DL has also developed some approaches, such as neural networks, and supervised, and unsupervised techniques. In particular, the automated extraction of hierarchical features

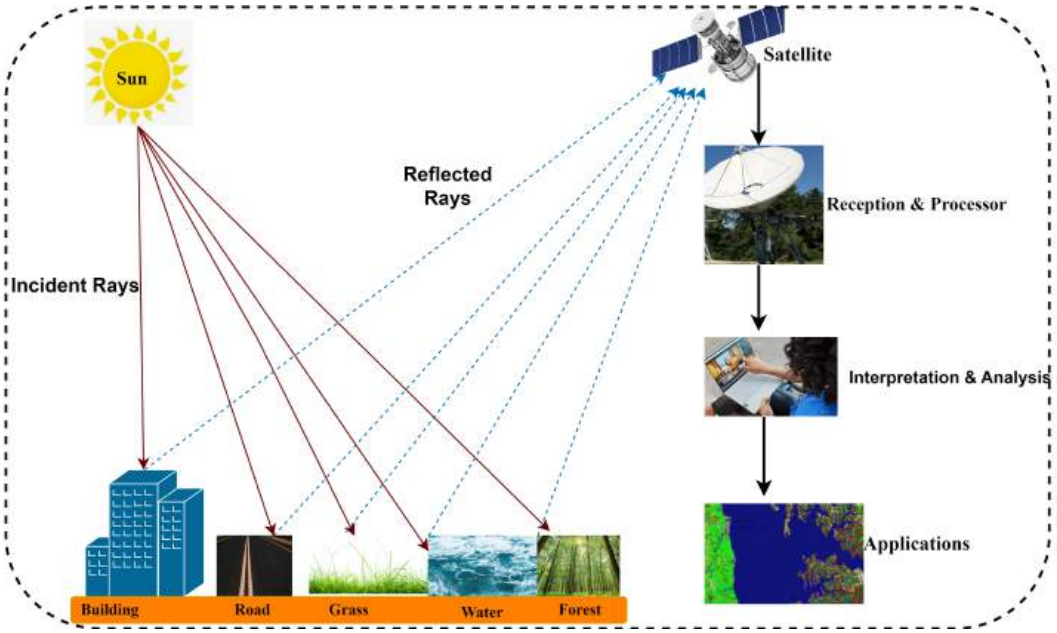


FIGURE 1.3: Remote Sensing Process

has greatly increased its value in RS. In remote sensing, adept interpretation of multispectral images is pivotal, achieved through the utilization of various sensors such as Landsat-8, Sentinel-1, Sentinel-2, Quickbird, Birdseye, and others. This proficiency is essential for effective analysis and understanding of remote sensing data. In this industry, achieving a multispectral resolution of 1.84 meters has become crucial.

Deep learning constantly shows itself as a very promising method for getting a lot of information out of satellite images. Deep learning models excel in precisely estimating complex non-linear correlations between environmental factors because of their multi-layered learning methodology (Zhang et al., 2016). Capturing potential connections between environmental elements, such as information retrieval, data fusion, downscaling, and several other operations, is made possible by this capacity. Furthermore, multi-resolution and multi-level learning applications in particular have shown the profound importance of deep learning in remote sensing applications. The field of image processing and the solving of classification problems have

significantly benefited from deep learning (Mountrakis et al., 2011). Neural network-driven deep learning methodologies have found recent applications in the realm of remote sensing. The remote sensing community, on the other hand generally relied on Support Vector machines (SVM) and related classifiers like Random Forests (RF) before the development of deep learning to handle tasks like image categorization and resource monitoring (Belgiu et al., 2016). Support Vector Machines (SVM), among the classifiers available, have garnered significant attention in remote sensing due to their superior performance in training and ability to manage high-dimensional data effectively. Random Forest (RF) has grown in popularity because of its ease of use in classification applications and excellent accuracy (Vetrivel et al., 2018). DL methods have shown promise in several image processing applications, and the remote sensing field has focused on DL since 2014. These tasks include classifying scenes, identifying objects, classifying land use and land cover (LULC), and natural resource monitoring. This study's main goal is to analyze remote sensing images, which can give decision-makers involved in urban planning and development useful information. Several examples of remote sensing images are shown below in FIGURE 1.4 (Vali et al., 2020,Naushad et al., 2021,Papoutsis et al., 2023,Otukei et al., 2010).

The following are some of the remote-sensing images.

- **Multi-Spectral Images(MSI):** Multispectral imaging, or MSI, refers to a particular kind of satellite image produced by the fusion of three to fifteen multispectral bands. Collectively, these bands accurately portray the MSI scene. Remote sensing applications use multispectral images to solve particular issues in remote sensing, whereas traditional colored images are normally made up of three bands: blue, green, and red. This form of image acquisition is used by several optical remote sensing systems, including well-known ones like Landsat and SPOT. Due to its capacity to offer a lot of useful spectrum information, multispectral mode is frequently favored over panchromatic mode.

- **Hyperspectral Images(HSI):** The term "HSI" which stands for "Hyperspectral Imagery" refers to a particular class of satellite image created

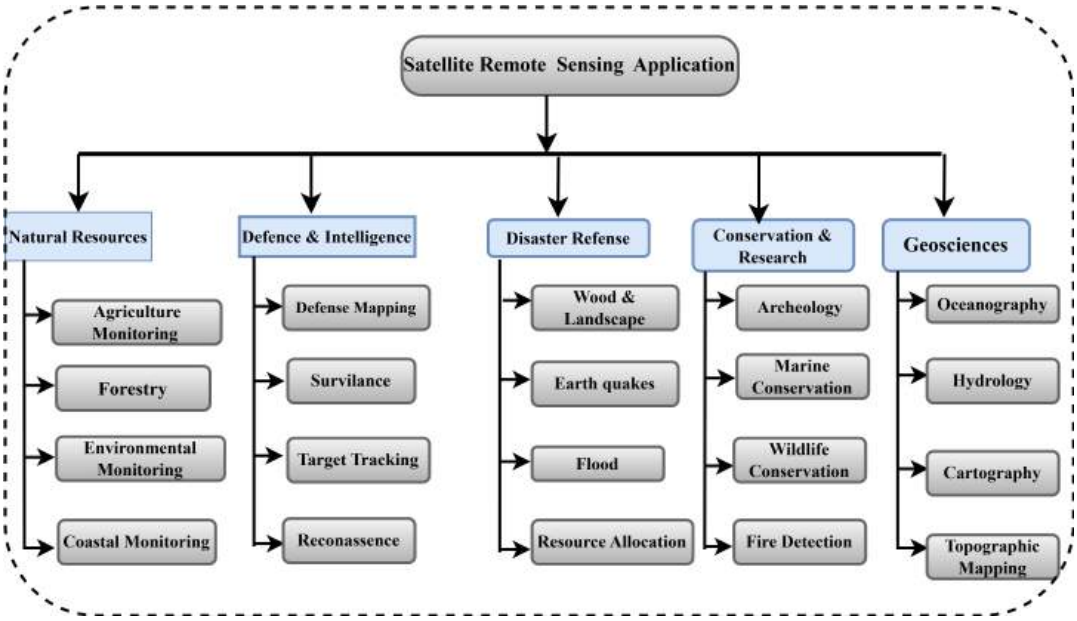


FIGURE 1.4: Different applications of remote sensing

by combining a wide variety of spectral bands with wavelengths between 100 and 1000. Because they have a greater ability to disclose minute features within the imaging, HSI has a distinct advantage over multispectral images. The change detection technique makes considerable use of these images. They gather information using a series of closely spaced bands to acquire data. Since there are so many bands in hyperspectral images, their sensitivity to minor variations in reflected energy is increased, which considerably improves their ability to distinguish differences between land and ocean objects.

- **Panchromatic Images:** Combining data from the visible red (R), green (G), and blue (B) bands results in a high-resolution grayscale image known as a panchromatic image. This amalgamation results in a single composite band devoid of wavelength-specific information. The term "pan-sharpening" pertains to the simultaneous capture of both a panchromatic and a multispectral image over the same geographical area. This can be conceptualized as a particular data fusion challenge. Satellites like Landsat, the Digital Globe satellite series, and the SPOT

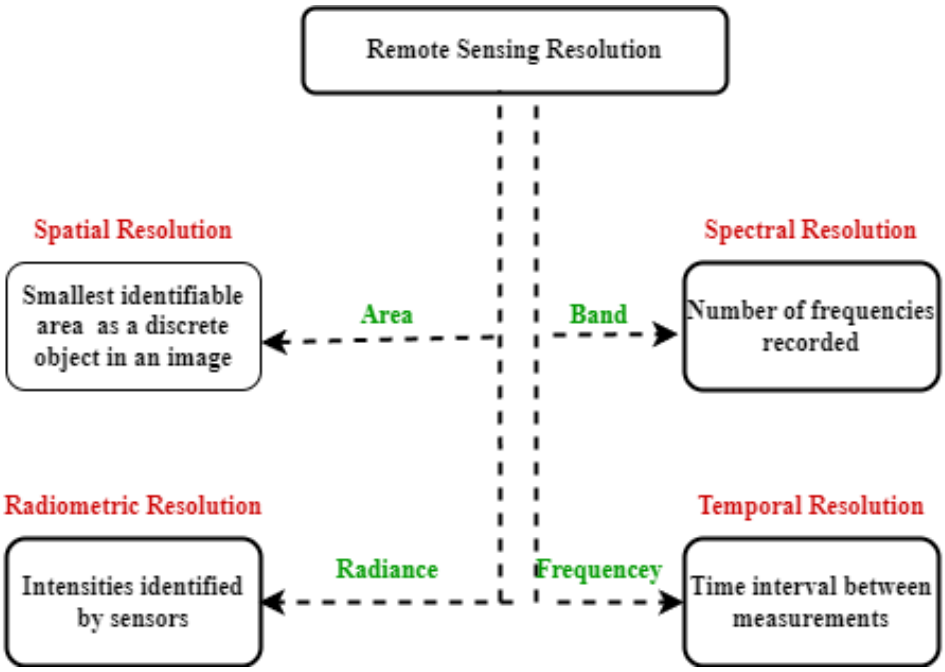


FIGURE 1.5: Different Types of resolutions

constellation possess spatial resolutions of 15 meters and typically generate panchromatic images alongside multispectral ones.

1.0.3 Resolution and its approaches

The amount of detail an image includes is determined by the number of pixels per inch (PPI) that make up a satellite image. The PPI count is closely related to the sophistication and caliber of the images as shown in Table1.1. The classification of image resolution typically comprises three tiers: low, medium, and high resolutions. Low and medium-resolution images reveal details of the image's main aspects, and high-resolution images provide precise details of its tiniest characteristics. In remote sensing data, the primary types of resolution include spatial, spectral, and temporal resolutions. These categories are shown in FIGURE 1.5 and are crucial characteristics for dealing with different distant sensing problems. Applications for remote sensing data include activities like change detection, natural resource monitoring, classification, climate analysis, and evaluation of land cover. However,

before remote sensing images can be processed effectively, many prerequisite actions must be taken, in addition to categorization and change detection. The choice of technique significantly impacts the pre-processing requirements for specific tasks within the field of remote sensing.

TABLE 1.1: Examples of Various Remote Sensing Resolutions.

Type	Sensors	Spatial Coverage	Temporal Resolution	Spatial Resolution
Panchromatic	Landsat, Spot	185 km, 60 to 80 km	16 days, 1 to 5 days	15 m, 1.5 m to 10 m
Multispectral	Landsat, Spot, Sentinel-2	185 km, 60 to 80 km, 290 km	16 days, 1 to 5 days, 2 to 10 days	30 to 120 m, 6 to 20 m
Hyperspectral	AVIRIS, EO-1 Hyperion	12 km, 7.75 km	1 year, 16 days	20 m, 30 m

1.0.4 Applications of Remote Sensing

- **Land Use and Land Cover (LULC)**: Advances in satellite-based remote sensing have boosted high-resolution image acquisition and LULC classification techniques, though data variability remains a challenge. Deep learning now drives major progress, enabling powerful image processing, change detection, and classification to address diverse remote sensing needs.

- **Hydrology**: In response to the pressing issue of water scarcity, the focus within remote sensing has shifted towards the identification and mapping of water resources, taking a prominent position in research and development efforts. Finding water sources that are essential for everyday survival depends heavily on remote sensing data. Furthermore, minor rivers, lakes, artificial wetlands, reservoirs, and other water sources may be automatically identified from satellite images by using complex algorithms on remote sensing data.

- **Agriculture**: Modern agriculture relies heavily on RS data, which provides a variety of applications that improve farming methods. Using satellite and aerial data, agricultural fields may be extracted and mapped, and cutting-edge methods like image segmentation can help discriminate between various land cover types. Additionally, RS makes

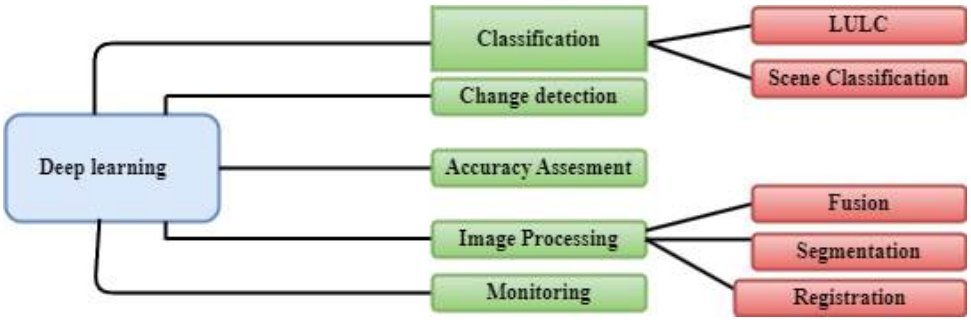


FIGURE 1.6: Different applications of remote seensing in Deep Learning

it possible for continuous crop monitoring using vegetation indices, giving farmers the ability to evaluate the health of their crops, predict yields, and efficiently manage irrigation. By recognizing distinctive spectral characteristics, the system also assists in early illness and pest identification, enabling prompt action. Additionally, RS data supports crop type identification and guides decisions on land use. The value of RS data is further expanded by cutting-edge data processing techniques including deep learning, image segmentation, data fusion, and GIS integration, which promote precision agriculture for improved crop production, resource optimization, and sustainable agricultural and is illustrated in FIGURE 1.6

1.0.5 Image classification using remote sensing

RS classification is a complex procedure that is essential for a variety of applications. Data from sources like satellites or drones are first acquired, then they are preprocessed to increase accuracy. Spectral, spatial, and texture qualities of objects in the data are identified by feature extraction. Known labels are gathered as training data for supervised or unsupervised classification systems. In the classification process, extracted characteristics direct statistical or machine-learning classifiers to place pixels or objects into predetermined classes. Validation checks classed findings against source data

to guarantee correctness. In the end, this classification helps with applications like urban planning, weather forecasting, mineral monitoring, agriculture, and environmental evaluation, providing insightful information for wise decision-making in a variety of industries.

- **Pixel-based classification:**Traditional pixel-based classification in remote sensing relies only on spectral properties, ignoring spatial context. While useful in some cases, it struggles with high-resolution images due to mixed pixels, noise, and boundary issues. Advanced methods like object-based classification and machine learning (e.g., random forests, CNNs) incorporate both spectral and spatial information, delivering more accurate and context-aware results for land cover mapping.

- **Object-based Classification:** Unlike pixel-based methods, object-based classification groups pixels into meaningful "objects" based on spectral and spatial features such as color, texture, and shape. Through segmentation and attribute extraction, these objects are classified by their collective properties, enabling more accurate and context-aware results. This approach is especially effective in high-resolution remote sensing for identifying features like buildings and vegetation. Moreover, classification methodologies are generally categorized into two distinct classification categories:

- **Supervised Classification:** A technique for classifying image pixels based on predetermined training sets is supervised classification. These training sets are made by choosing pixels from the image that have uniform and homogenous properties throughout the image. There are two primary phases to the process: The process of identifying the information categories for the training sets comes first. In this phase, identify the classes or categories you wish to give various areas of the image. Second, the classification algorithm is taught by the training sets how to label each pixel in the incoming image. The parameters of the algorithm, which were established using the training data, are used to guide this labeling. Supervised classification, often called

semi-automated, requires human input to create training sets but offers higher accuracy than unsupervised methods. It is widely used in remote sensing for land cover mapping and object recognition, though it demands more effort and careful data understanding.

- **Unsupervised Classification:** Unsupervised classification, also known as clustering or cluster analysis, is a fully automated method of categorizing images that relies solely on the innate patterns and traits contained in the input image data. For categorizing images, this technique does not require prior information or predetermined categories. It typically consists of two main stages: the first is the process of grouping things based on common traits, where the algorithm groups objects or pixels with similar spectral characteristics or features as determined by statistical patterns in the image data. The second step is to give classes or labels to the newly generated clusters, a process that is difficult because no pre-existing categories are connected to these clusters.Unsupervised classification avoids the need for training sets, offering speed and automation. However, it risks misclassification, is harder to validate, and depends heavily on algorithm choice. It is useful when prior knowledge is lacking but requires careful evaluation of results.

1.0.6 Scope of Study

This study aims to develop an advanced model for accurately identifying water bodies in satellite imagery, with broad applicability in environmental monitoring, disaster management, agricultural planning, and water resource conservation. By leveraging remote sensing technology, the research enhances precision in tracking water bodies, facilitating improved decision-making for sustainable resource management. The model supports flood and drought prediction, irrigation planning, and scientific research in hydrology. Aligning with Sustainable Development Goal (SDG) 13: Climate

Action, the study contributes to climate resilience by monitoring water distribution changes, aiding in climate change adaptation, and ensuring sustainable water use. It provides valuable insights for policymakers to develop evidence-based environmental strategies while strengthening community resilience against climate-induced water challenges. By integrating AI-driven water body detection with remote sensing, this research offers a scalable, data-driven approach to global climate monitoring, promoting sustainable development and environmental sustainability.

1.0.7 Outcome of thesis

The primary objective of this thesis is to develop an intelligent and robust deep learning-based framework for detecting and monitoring water bodies in satellite imagery. The study focuses on four key objectives: analyzing the role of remote sensing in natural resource monitoring, enhancing satellite image quality through advanced pre-processing techniques, developing an efficient segmentation framework for water body identification, and evaluating the model's performance against state-of-the-art approaches using metrics like accuracy, precision, recall, F1-score, and Intersection over Union (IoU). The proposed framework integrates residual refinement modules, attention mechanisms, and deep supervision, offering significant improvements in segmentation accuracy and robustness, especially in challenging scenarios involving small or low-resolution water bodies.As illustrated in the block diagram FIGURE 1.7, each objective aligns with dedicated thesis chapters, ranging from literature reviews on remote sensing and deep learning to the development of the proposed segmentation framework and the final conclusions.

The outcomes of this research demonstrate the model's superiority over existing techniques and its scalability for real-world applications. The framework provides actionable insights for policymakers and environmental agencies to implement AI-driven remote sensing tools for effective water resource

management. Additionally, the thesis contributes to the broader field of natural resource monitoring by showcasing how deep learning and remote sensing can be combined for precise environmental analysis. This work lays a foundation for future innovations in automated satellite image analysis, particularly in supporting sustainable and data-informed decision-making for water conservation efforts.

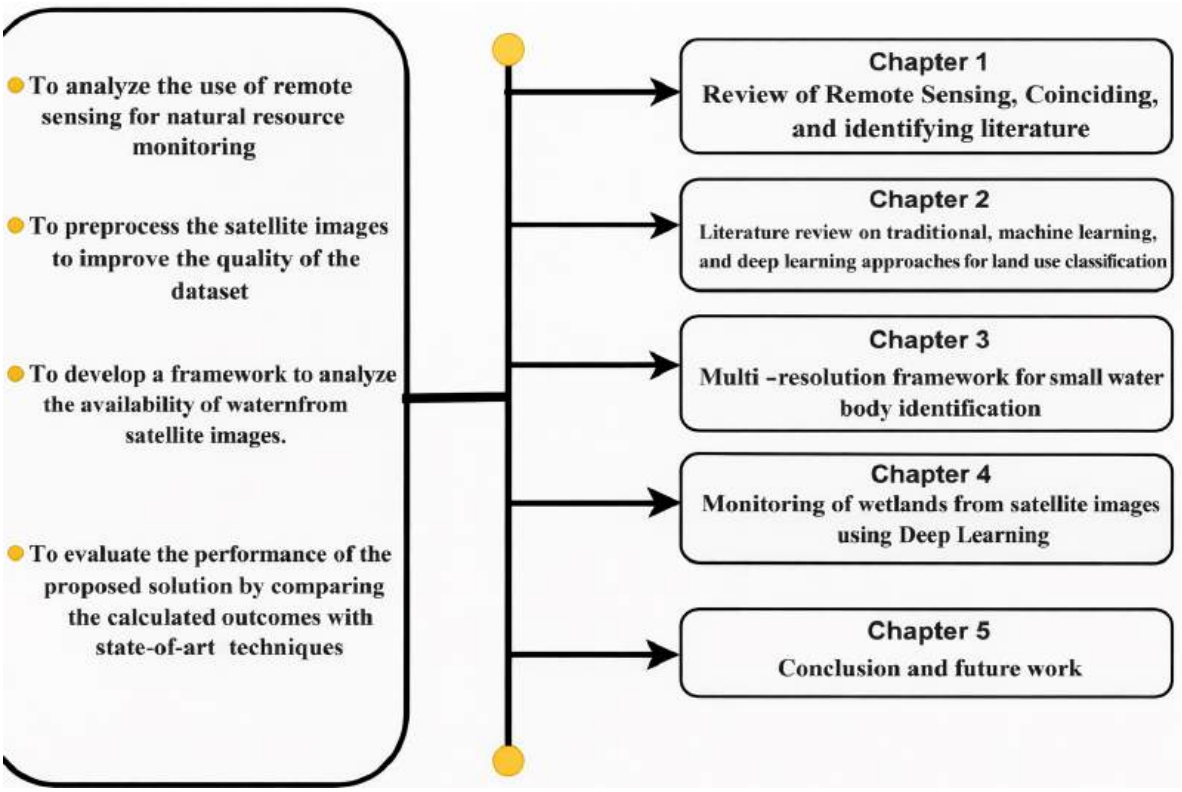


FIGURE 1.7: Thesis Contribution

1.1 Thesis Outline

This thesis is organized into five chapters, each addressing a specific aspect of remote sensing and deep learning–based water body analysis.

Chapter 1 provides an introduction to the study and discusses the fundamental concepts and applications of remote sensing. It covers major application areas such as resource monitoring, image classification, change detection, segmentation, and other related domains, thereby establishing the background and motivation for the present research.

Chapter 2 presents an extensive review of existing literature on deep learning methodologies applied to land use and land cover (LULC) classification using multispectral remote sensing data. This chapter critically analyzes prior studies and highlights the potential of deep learning techniques in efficiently managing large-scale datasets and leveraging the exponential growth in global data availability.

Chapter 3 focuses on the detection of small water bodies using remote sensing data. It discusses the challenges associated with identifying small and fragmented water features and presents the adopted methodology and data preprocessing techniques used in the study.

Chapter 4 introduces an additional deep learning–based methodology for the accurate extraction of large water bodies from satellite images. This chapter details the proposed model architecture, implementation strategy, and performance evaluation, demonstrating its effectiveness in water body segmentation tasks.

Chapter 5 concludes the thesis by summarizing the key findings of the research. It also outlines the limitations of the study and provides directions for future research in the field of remote sensing and deep learning–based water resource analysis.

Chapter 2

Related Work

A comprehensive review of the existing literature is presented, focusing on previous work related to the research area and identifying the research gap.Over the past few years, there has been a surge in the deployment of Earth observation satellites, providing end users with an extensive range of imagery. These include various types such as infrared and SAR, differing in resolution from high to broad coverage, and in spectral bands from monochrome to multispectral, often arranged in periodic temporal sequences. Satellite imagery has proven highly effective in applications like classification, monitoring, change detection, and feature extraction. The processing of remote sensing imagery involves a series of steps with its success largely hinging on the techniques employed (Laignel et al., 2023). In recent times, deep learning techniques have made significant strides, excelling not only in conventional tasks like voice recognition, detection, and text segmentation but also in diverse practical applications across different domains (Ball et al., 2017; Sun et al., 2019). Deep learning techniques draw inspiration from the intricate architecture of the human brain particularly the profound design of the human visual system. The perceptions generated through these techniques manifest multiple phases of absorption (Alzubaidi et al., 2021; Sarker, 2021). In recent years the continuous monitoring and examination of natural resources via satellites have shown a consistent increase (Majumder et al., 2023). Professionals across a range of sectors, including environmental preservation, defense, geography, and cartography, are utilizing satellite imagery for ongoing monitoring to evaluate alterations in particular areas (Chuvieco, 2020).

Despite advancements precisely discerning patterns in satellite images remains a formidable challenge. Previous research has introduced various methods to detect natural and artificial bodies of water. Nevertheless analyzing the dynamic spectral properties of water through satellite images remains a complex undertaking (Baci, 2023). A variety of methods, such as water index calculations, single and multi-band threshold techniques, water mapping approaches, and numerous others, have been developed for the analysis of water bodies using remote sensing technologies (Li et al., 2022; Bijeesh et al., 2020). Various handmade elements have been built with water body features because of classifying pixels as water or no water.

Tasseled Cap Wetness Index : The Tasseled Cap Wetness Index (TCW), introduced by Crist et al., 1984, serves as a tool for discerning between water and non-water elements, leveraging six surface reflectance bands with a threshold set at 0. However, a comparative analysis conducted by Zhou et al in 2017 underscored a notable drawback of the TCW. The study identified instances where non-water pixels were inaccurately labeled as water pixels, revealing a limitation in the TCW's ability to effectively differentiate between water and non-water features. Consequently, this led to occurrences of false positives, as shown by (Zhou et al., 2017). Hence, in specific contexts, the reliability of the TCW may be compromised, prompting users to exercise caution regarding potential misclassifications when employing it to distinguish water from non-water elements.

Normalized Difference Vegetation Index : The Normalized Difference Vegetation Index (NDVI) is recognized for yielding low values in areas covered by water bodies, making surface water areas easily identifiable through their low NDVI values as indicated by (Khand et al., 2017). However, it's important to note that while NDVI is effective in highlighting vegetation, it isn't specifically designed to directly convey information about surface water areas. In essence, although NDVI can be indicative of the absence of vegetation in water bodies due to its low values it doesn't offer a direct measure or detailed information about the extent or characteristics of surface water areas. Therefore while it may be suitable for extracting information about vegetation, its utility is limited when it comes to providing comprehensive

details about surface water areas(Yengoh et al., 2015).

Normalized Difference Water Index: The Normalized Difference Water Index (NDWI), pioneered as the initial iteration of water indices by McFeeters, 1996, employs the Near Infrared (NIR) band to capture the unique characteristics of water (Rad et al., 2021). NDWI has demonstrated effectiveness in effectively distinguishing water bodies from surrounding vegetation. Yet, its utilization faces challenges in precisely distinguishing water from various land cover categories, especially when dealing with soil and developed regions. While NDWI is robust in detecting water features and is widely employed for this purpose, its performance may be less reliable in areas where soil or built-up structures share spectral characteristics with water bodies. This limitation arises from the fact that NDWI relies on the contrast between the reflectance in the NIR band and the visible green band, and certain surfaces, such as bare soil or urban structures, may exhibit spectral similarities to water, leading to potential misclassifications. Therefore, while NDWI is a valuable tool for water body delineation, users should be cautious in its interpretation, especially in landscapes where the spectral signatures of non-water features can overlap with those of water bodies, introducing challenges in accurately classifying soil and built-up areas.

Modified Normalized Difference Water Index: The Modified Normalized Difference Water Index (MNDWI) is created by substituting the Near Infrared (NIR) band in NDWI with the Short-Wave Infrared (SWIR) band, aiming to enhance water detection reliability, particularly in regions dominated by urban development, as the SWIR band is less influenced by sediment concentration in water compared to the NIR band. As per the research of Gautam et al., 2015, MNDWI is designed to offer improved accuracy in identifying water features, especially in areas primarily consisting of urban landscapes. However, there are concerns regarding the spatial resolution of bands in MNDWI. In certain optical data scenarios, the fusion of the SWIR band and the Green band in MNDWI may lead to a reduction in image quality due to differences in spatial resolution between these bands. Additionally, MNDWI encounters challenges in discriminating between water and snow. This difficulty arises from the normalized difference between the

Green band and the SWIR band for snow, which can be comparable to that of water (Huang et al., 2018). Consequently, in specific optical data contexts, MNDWI may face difficulties in accurately distinguishing between water bodies and snow-covered areas.

Kauth-Thomas Transformation: The Wetness component of the K-T (Kauth-Thomas) transformation, proposed by Wang et al., 2013, seeks to pinpoint moist areas within optical images by detecting both surface water bodies and humid regions. This technique involves transforming original spectral bands into new components, with Wetness being one of them. While the Wetness component is sensitive to moisture content, offering utility in highlighting areas with high humidity or water content, it comes with limitations. One notable drawback is its potential to misidentify non-water features, such as buildings, as water in the image. This sensitivity to moisture content may lead to the misclassification of materials or conditions with spectral characteristics similar to water. Particularly in urban areas, where non-water features may share spectral similarities with water in the specific bands used by the transformation, caution in interpretation is crucial. To enhance accuracy and mitigate the risk of false positives, researchers should consider additional validation or complementary methods when utilizing the Wetness component for water identification in optical images.

The Automated Water Extraction Index: The Automated Water Extraction Index (AWEI), a novel index introduced by Jax, 2010, serves as an improved option for water index applications, particularly in regions where the occurrence of shadows and urban surfaces could lead to noisy results. Tailored specifically for Landsat images, AWEI is composed of two indices: AWEI optimized for effective performance in conditions free of shadows, and AWEI, which further enhances the differentiation between water pixels and shadow pixels. AWEI has demonstrated heightened effectiveness in retrieving water-related information from regions prone to noise caused by shadows and built-up surfaces. Significantly, AWEI has been observed to perform more efficiently in identifying water features in urban areas compared to expansive open surface water, as documented by (Rokni et al., 2014).

This highlights AWEI's particular suitability for applications in urban environments, where challenges related to shadows and built-up surfaces are more prominent. This underscores the index's capacity to overcome limitations associated with traditional water indices in urban settings.

Water Index: The Water Index (WI2015), developed by Fisher et al., 2016, employs statistical analysis of surface reflectance across the Visible, Near-Infrared (NIR), and Short-Wave Infrared (SWIR) bands. Tailored specifically for Landsat data, WI2015 benefits from the improved accuracy of surface reflectance input compared to the utilization of Digital Numbers (DN) or top-of-atmosphere (TOA) reflectance. Although surface reflectance provides a more precise representation of Earth's surface reflectance characteristics, the process of converting from DN to surface reflectance involves complex computational procedures. This conversion is crucial for achieving accurate and meaningful outcomes but may present practical challenges due to its computational complexity and time requirements. Consequently, when applying the Water Index (WI2015) to Landsat data, it is essential to carefully consider the trade-off between the enhanced accuracy provided by surface reflectance and the computational demands associated with the conversion process. Thresholding plays a pivotal role in the mapping of surface water, with an array of methods such as density slicing, edge detection, and water indices utilizing thresholding to isolate surface water features. Determining suitable threshold values is challenging and often hard to generalize due to the diverse characteristics of surface water in different images. Incorrectly chosen thresholds can lead to misclassification, mistakenly identifying mountain shadows, urban regions, or other background noise as water bodies.

Therefore, manual adjustments to the threshold are often necessary for accurate results. An alternative approach involves automatic extraction using a single threshold for all data. Automatic thresholding using the Otsu algorithm was implemented by Buma et al., 2018 for mapping the surface water extent of Lake Chad. While the techniques mentioned predominantly rely on pixel-based data for surface water extraction, managing extremely

high spatial resolution images presents difficulties. In such scenarios, Object-Based Image Analysis (OBIA) becomes more suitable, categorizing objects according to distinctive textural attributes Dronova et al., 2015. Nonetheless, OBIA necessitates thorough processing for segmentation, frequently imposing constraints on the image size processed depending on the software utilized (Hahmann et al., 2008). Furthermore, obtaining meaningful outcomes entails proficiency in pertinent features. Despite ongoing endeavors to devise techniques for surface water extraction from optical remote sensing data, encompassing automated approaches, it's essential to recognize their constraints, particularly their vulnerability to cloud cover. Modern techniques: After threshold methods, modern learning methods came into existence

Machine Learning: To overcome the inherent constraints of index-based methods for detecting water bodies in remote sensing imagery, several researchers have pioneered novel methodologies to leverage spatial characteristics and additional parameters for more accurate identification. Qin, 2014 introduced an innovative approach that incorporated nearby input image features, effectively utilizing spatial characteristics to detect water bodies. However, Huang et al., 2015b proposed two separate machine learning frameworks to recognize water bodies in high-resolution remote sensing images. While these frameworks integrated geometric and topographical parameters to classify water body categories following pixel-level recognition, there remained a drawback: manual threshold settings potentially introduced errors during the extraction process. To mitigate such limitations, Yao et al., 2015 proposed an automated urban water extraction system (AUWEM), which integrated a water index with a technique tailored for identifying building shadows. Similarly, Wu et al., n.d. utilized an urban water index (UWI) to identify water bodies, complemented by the urban shadow index (USI) for shadow pixel removal from the detected water bodies. However, both Yao et al., 2015 and Wu et al., n.d. relied on determining thresholds through a support vector machine (SVM) followed by a shadow detection method, which restricted the applicability of these methodologies specifically to urban environments, limiting their universal

adaptability across diverse spectral and spatial properties. Moreover,Chen et al., 2018 highlighted the necessity of additional inputs, such as slope and mountain shadow in the initial band, to enhance the accuracy of these approaches. Nevertheless, integrating these factors escalates the data volume and computational complexities, posing significant challenges in practical implementation. These factors collectively contribute to increased data requirements, computational burdens, and limitations in the generalizability of these methodologies beyond specific environments, hindering their widespread applicability across diverse spectral and spatial characteristics of remote sensing imagery. Thus, while these approaches have made strides in improving water body detection, their limitations underscore the need for more robust and universally adaptable methods in the field.

Deep learning: In recent times, deep learning frameworks have attracted considerable interest within the academic sphere, showcasing remarkable achievements across a spectrum of remote-sensing applications. Notably, Zhu et al., 2017 and Afaq et al., 2021 have utilized fully convolutional networks (FCN) for semantic segmentation, where each decoder module conducts upsampling on its feature map, combined with the corresponding encoder module's feature map. The training process for the extensive FCN involves multiple stages due to its size, incorporating diverse and valuable information through convolution and pooling operations at various levels (Long et al., 2015). Zheng et al., 2015 proposed the CRFRNN technique, integrating conditional random fields (CRF) with recurrent neural networks (RNN) to enhance FCN prediction accuracy. Lin et al., 2017a made efforts to enhance FCN capabilities for ship detection through task partitioning models, specifically designed for semantic labeling at multiple scales in sea-land regions. Lin et al., 2017b employed Fully Convolutional Networks (FCNs) to isolate water bodies from high-resolution images, examining different parameter combinations to pinpoint the most effective models, notwithstanding challenges in detecting narrow rivers and level landscapes.

Addressing misinterpretation challenges of surrounding pixels, Miao et al., 2018 proposed the RRF DeconvNet for water body identification. Isikdogan et al., 2017 introduced DeepWaterMap, capable of differentiating water

bodies from snow/ice, clouds, and shadows using Landsat images. Nevertheless, difficulties emerged when segmenting densely populated urban areas with extremely high-resolution images.

Ronneberger et al., 2015 presented U-Net, an encoder-decoder network employed for biomedical segmentation, using extensive data augmentation like elastic deformation to handle limited training images. Researchers have developed variations of U-Net for specific remote sensing applications, including building identification, road detection, and water body extraction (Rakhlin et al., 2018; Kim et al., 2018; Iglovikov et al., 2018; Yang et al., 2019).

Yang et al., 2019 A novel approach was presented, introducing the Deep Convolutional Encoder-Decoder (DCED) architecture for extracting water bodies from images of remarkably high resolution. Challenges arose in border pixel blurring during the up-sampling process, mitigated by the fully connected CRF (FCCRF) model. TernausNet, proposed by Gonzalez et al., 2019, used knowledge transfer for mapping high-resolution tagged images onto extremely high-resolution images, but encountered difficulties in segmentation due to spectral disparities.

The conspicuous absence of specific details regarding the suggested model leaves room for intrigue, yet the implicit claim of superior performance tantalizes curiosity. By amalgamating insights from Tambe et al., 2021 and Manocha et al., 2023 the model likely incorporates innovative strategies to navigate the challenges posed by the intricate interplay of convolutional and pooling operations within fully connected networks (FCNs). An innovative method integrating multi-level feature fusion is introduced to forecast water patterns about a particular area, enabling analysis of scale and accessibility. It has been noted that this novel approach exhibits greater resilience in pattern analysis compared to existing state-of-the-art methods. One major constraint is the lack of explicit discussion or exploration of potential challenges or issues related to the generalization of the proposed model across diverse environmental conditions or geographical regions. Robustness across varied landscapes, climate conditions, and water body types is crucial for the practical utility of such models. A paper Liu et al., 2023b emphasizes the efficacy of

remote sensing and geostatistics in monitoring urban water resources, offering vital understandings of water quality, availability, and distribution. Successful applications in urban settings showcase their potential to inform water management strategies and decision-making processes. However, challenges remain, including data availability, spatial-temporal resolution, and accuracy limitations. Further research is necessary to address these issues by integrating diverse data sources, improving model validation techniques, and enhancing result interpretability for practical implementation in urban water-resource monitoring. Wieland et al., 2023 presents a semantic segmentation technique for identifying water bodies in high-resolution satellite imagery, achieving noteworthy precision. This approach benefits urban planning, environmental monitoring, and disaster management, improving spatial understanding. Challenges include computational complexity, scalability limitations, and difficulties in accurate delineation in complex urban or vegetated areas. Reliance on high-quality imagery also poses constraints, especially in regions with limited accessibility or frequent cloud cover. Overcoming these challenges is crucial for wider adoption across diverse geographic contexts. Nasir et al., 2023 proposes a deep learning approach for detecting various types of water bodies using optical variables and ensembling techniques, advancing water body classification accuracy and efficiency in remote sensing applications. Sureshkumar et al., 2023 Introducing Denseppmunet-a, a robust deep-learning network designed to precisely segment water bodies from aerial images. It leverages DenseNet and PSPNet architectures for feature extraction and spatial context enhancement, incorporating multi-scale fusion to improve segmentation. With superior accuracy, it holds promise for hydrological modeling and environmental monitoring applications. Pena et al., 2024 introduces a"DeepAqua" method using SAR imagery and deep neural networks to segment wetland water surfaces without manual annotation. By employing unsupervised learning and domain adaptation, the approach accurately delineates water bodies in wetland areas, leveraging SAR features to overcome traditional supervised learning

challenges, and providing an automated solution for wetland water surface segmentation. Filali Boubrahimi et al., 2024 proposes the augmentation of MODIS-Landsat water body data in spatiotemporal dimensions using adversarial networks to enrich the diversity and quantity of training data. This strategy improves model performance by simulating variations in spatiotemporal conditions, enabling more accurate water body detection and monitoring from satellite imagery. Russo et al., 2024 integrates multitemporal data from Sentinel-1 and Sentinel-2 satellites to map water bodies, improving accuracy through the utilization of complementary radar and optical data. This approach improves detection and mapping capabilities, benefiting environmental monitoring and resource management applications. This study Zhang, 2024 suggests water body extraction from remote sensing images using PSPNet, achieving accurate segmentation and enabling hydrological analysis and environmental monitoring. Liu et al., 2024 introduces an innovative deep-learning architecture for extracting lake water bodies from remote sensing imagery. This model achieves accurate lake water body extraction, demonstrating the potential for enhanced environmental monitoring and resource management applications. Tran et al., 2024 Mapping Tri An reservoir's water body areas in the 2023 dry season using Sentinel-1 SAR imagery highlights significant reductions, particularly in the north and near the La Nga bridge, linked to temperature rise. While insightful for climate change impact, future enhancements should consider other factors and integrate actual lake data, alongside advanced methods like Otsu, to improve accuracy. These models analyze spectral and spatial patterns in satellite imagery to accurately map water extent and dynamics, enabling efficient and scalable monitoring for various applications including hydrology, ecology, and resource management. Advanced architectures and techniques, such as U-Net, Mask R-CNN, and data fusion, further improve accuracy and robustness in water body detection and analysis. A literature review on traditional, machine learning, and deep learning approaches for waterbody extraction is provided in this chapter. The traditional techniques are illustrated in Table 2.1, while the machine learning and deep learning techniques are presented in Table 2.2 and Table 2.3, respectively.

TABLE 2.1: Remote Sensing-Based Water Mapping Techniques Using Traditional Approaches

Authors	**Journal**	**Remarks**	**Limitations**
Crist et al., 1984	Transactions on Geoscience and Remote Sensing	Simplifies Landsat TM data into brightness, greenness, and soil moisture.	Lacks adaptability for diverse sensors and landscapes.
McFeeters, 1996	Remote Sensing Letters	NDWI effectively maps water bodies, aiding hydrological analysis.	Struggles with distinguishing water from shadows or dark surfaces.
Jax, 2010	Remote Sensing of Environment	AWEI excels in shadowed mountainous regions and outperforms other indices.	Instability in thresholding under varying environmental conditions.
Wang et al., 2013	21st International conference on geoinformatics	Combined MNDWI and K-T transformation achieves 93% classification accuracy.	Lacks robust validation across diverse landscapes and sensor types.
Rokni et al., 2014	Remote Sensing	NDWI-PCs method effectively tracks surface decline over time.	Misclassifications due to spectral similarity and limited spatial resolution of Landsat data.
Fisher et al., 2016	Remote Sensing of Environment	High accuracy for pure water pixels using WI2015, WI2006, and AWEI.	Poor performance on mixed pixels; underestimates water in complex terrains.

Yengoh et al., 2015	SpringerBriefs in Environmental Sciences	NDVI effectively detects vegetation loss and land degradation.	Affected by seasonal variation and coarse resolution.
Gautam et al., 2015	Aquatic Procedia	Combining indices improves urban water body detection and mapping.	Suffers from spectral confusion and seasonal effects; limited accuracy in heterogeneous areas.
Rad et al., 2021	Environmental Modelling and Software	ANDWI and cloud filtering enhance detection accuracy (up to 98%).	Requires validation across different sensors; adds computational complexity.
Wu et al., n.d.	Remote Sensing	TSUWI achieves high precision (Kappa = 0.97) with minimal classification error.	Uncertain generalization to new geographic areas and sensor configurations.

TABLE 2.2: Remote Sensing-Based Water Mapping Techniques Using Machine Learning Approaches

Authors	**Journal**	**Remarks**	**Limitations**
Hahmann et al., 2008	7th European Conference on Synthetic Aperture Radar	Effectively utilizes TerraSAR-X data for flood mapping, capturing spatial variations in water bodies.	Struggles with detecting small water bodies due to speckle noise and clutter in SAR imagery.

Qin, 2014	Journal of Selected Topics in Applied Earth Observations and Remote Sensing	MSVSF method enhances high-resolution image classification, especially in urban environments.	Highly dependent on image characteristics and lacks extensive testing across diverse land cover types.
Huang et al., 2015b	Journal of Selected Topics in Applied Earth Observations and Remote Sensing	Dual-tier ML framework combines pixel and object-level features to achieve 96% accuracy in urban water mapping.	Limited generalization across different cities and complex hydrological structures.
Chen et al., 2018	Remote Sensing	Knowledge decision tree demonstrates good accuracy in extracting water bodies post-disaster.	Model's adaptability is limited across varying disaster scenarios and regional data characteristics.
Sureshkumar et al., 2023	The Computer Journal	Ensemble groundwater prediction model improves accuracy by integrating hydro indices and statistical features.	Suffers from interpretability issues due to the deep learning component and depends on limited data availability.

TABLE 2.3: Remote Sensing-Based Water Mapping Techniques using Deep Learning Approaches

Authors	**Journal**	**Remarks**	**Limitations**
Long et al., 2015	Proceedings of the IEEE conference on computer vision and pattern recognition	Successfully adapts deep architectures (AlexNet, VGG, GoogLeNet) for water segmentation with fast inference.	Tends to lose fine-grained spatial details during up-sampling; prone to dataset bias.
Zheng et al., 2015	Computer Vision Foundation	CRF-RNN integrates CNNs and CRFs for improved semantic segmentation without needing post-processing.	Complex training process, potential overfitting, and limited comparative validation in varied settings.
Ronneberger et al., 2015	Computer Science Book Series	Novel contracting-expanding path network excels in biomedical segmentation tasks like neuronal tracking.	High computational demands and patch-wise processing hinder its real-time applicability.
Lin et al., 2017a	Geoscience and Remote Sensing Letters	FCN provides strong feature representation for ship detection, handling anisotropic shapes effectively.	Performs poorly in densely packed scenes and lacks scalability to large datasets or broader domains.

Lin et al., 2017b	Remote Sensing	Multi-scale FCN architecture preserves fine spatial details in sea-land-ship segmentation.	Struggles in scenes with object overlap, complex environmental variability, or degraded image quality.
Isikdogan et al., 2017	Journal of Selected Topics in Applied Earth Observations and Remote Sensing	DeepWaterMap generalizes well for global water mapping with minimal regional tuning.	Susceptible to overfitting when trained on small datasets; requires high computation and rich data sources.
Huang et al., 2018	Reviews of Geophysics	Summarizes advancements in optical-sensor-based water detection for large-scale monitoring.	Lacks integration of deep learning techniques and offers limited validation depth for proposed solutions.
Zhu et al., 2017	Geoscience and remote sensing magazine	Provides comprehensive overview of DL evolution in remote sensing for environmental applications.	May omit the most recent breakthroughs; identifies difficulty in deep learning adoption in practice.
Afaq et al., 2021	Ecological Informatics	DC-RBM framework yields high precision, particularly effective for low-resolution satellite images.	Needs better focus on operational costs, real-world deployment, and integration with policy frameworks.

Manocha et al., 2023	Neural Computing and Applications	Multi-level feature fusion achieves outstanding 94.51% accuracy in Sentinel-2 image water classification.	Lacks geographic and seasonal robustness; limited model interpretability restricts transparency.

2.1 Research Gap

- Most existing studies prioritize large water bodies, often overlooking smaller yet ecologically vital water sources such as ponds, streams, and small lakes. These water bodies play a crucial role in hydrological cycles, biodiversity, and local water security, yet their monitoring remains inadequate in global water assessments(Manocha et al., 2023).
- Current spectral indices used for water monitoring often suffer from inconsistencies due to noise, shadow effects, and misclassification of outlier pixels. These limitations reduce the accuracy of water body delineation, particularly in complex environments with mixed land cover types.
- Existing research primarily addresses spatial, spectral, or temporal resolution separately. However, an integrated approach combining fine-resolution data across all three dimensions is essential for accurately monitoring water dynamics, especially in regions where small water bodies exhibit high variability due to precipitation and evapotranspiration fluctuations(Liu et al., 2022).

.

2.2 Issues and Challenges

Traditional water resource data collection in rural and remote regions of India largely relies on manual field-based methods such as dipsticks, rain

gauges, piezometers, float techniques, and community surveys. While these approaches are feasible in low-technology settings, they are labor-intensive, time-consuming, and limited in spatial and temporal coverage. Constraints related to accessibility, manpower, and cost often result in infrequent data collection, leading to incomplete representation of water resource dynamics. In addition, human-induced measurement errors and the absence of real-time monitoring reduce data reliability and limit timely decision-making, particularly during extreme hydrological events.

Recent advancements in remote sensing and machine learning have enabled large-scale water monitoring using multispectral satellite data. However, conventional machine learning techniques depend heavily on handcrafted features and are sensitive to data quality and spectral complexity. Although deep learning methods, especially CNN-based models, can automatically learn hierarchical spatial and spectral features, accurate water body extraction remains challenging due to seasonal variability, spectral confusion with surrounding land covers, shadow effects, and scale variations of water bodies.

2.3 Problem statement

Despite progress in remote sensing and deep learning, there is a persistent need for an accurate, scalable, and automated framework for water body extraction from multispectral satellite imagery. Existing manual methods lack efficiency and coverage, while traditional machine learning models suffer from limited robustness and generalization. Furthermore, deep learning–based approaches continue to face challenges related to spectral ambiguity and environmental heterogeneity. Therefore, this research aims to develop a robust deep learning–based framework capable of reliably extracting and monitoring water bodies across diverse geographical and climatic conditions, thereby supporting effective water resource management and disaster mitigation. The primary objective of this research is to develop an efficient

deep learning–based framework for the extraction and analysis of water bodies from multispectral satellite images. To achieve this goal, the specific objectives of the study are as follows:

- To analyze the role of remote sensing techniques in natural resource and water resource monitoring.
- To preprocess multispectral satellite images in order to enhance data quality and improve classification performance.
- To develop a robust framework for analyzing the availability and distribution of water bodies using satellite imagery.
- To evaluate the performance of the proposed approach by comparing its outcomes with state-of-the-art techniques.

Chapter 3

Detection of Small Water Bodies Using Remote Sensing

Water plays a pivotal role across multiple sectors, notably in agriculture, yet the burgeoning demand spurred by industrialization and urban expansion has triggered pollution and posed formidable challenges in administering urban water systems. Insufficient management not only jeopardizes sustainability but also threatens ecosystems. Addressing this predicament necessitates a holistic grasp of urban water systems. Remedial actions encompass the deployment of water treatment technologies, embracing conservation strategies, and erecting resilient infrastructure. Fostering awareness and fostering collaboration among governmental bodies, industries, and communities stand as indispensable requisites for efficacious urban water governance and sustainability. From oceans to rivers, lakes, streams, and reservoirs, various vital water sources sustain diverse ecosystems. However, the rapid pace of urbanization is precipitating alterations and detrimental impacts on extant water reservoirs. Thus, conducting thorough surveys and ongoing monitoring efforts are imperative to amass insights into water bodies, pivotal for judicious conservation and efficient utilization of existing water resources. (Pirot et al., 2000).

While manual surveys provide accurate outcomes, they are immensely laborious and time-intensive. Recent strides in remote sensing have simplified the rapid utilization of data amassed from diverse sensors aboard satellite platforms. This factual data is extensively employed for manifold purposes, including pinpointing specific attributes of water bodies. Remote

sensing imaging primarily targets the identification and differentiation of water bodies from other features. Satellite imagery, by its very nature, encompasses a breadth of information, encompassing man-made structures, forests, snow-covered areas, and arid regions, thereby complicating the precise extraction of water bodies from the imagery. Traditional methods reliant on water body indices encounter challenges in precisely identifying border pixels and accurately categorizing water pixels as non-water pixels, largely owing to the protracted process of selecting thresholds (McFeeters, 1996; Xu, 2006). Deep learning, notably Convolutional Neural Networks (CNNs), has transformed water body surveillance by demonstrating remarkable proficiency in image segmentation and differentiation of water from land. It autonomously discerns features such as texture and color to achieve precise identification. Leveraging supervised learning, deep models effectively classify pixels, facilitating precise mapping. Object detection techniques pinpoint water features, while temporal analysis of time series data monitors changes over time. Transfer learning facilitates adaptation to restricted labeled datasets, and amalgamation with supplementary data sources augments accuracy (Maggiori et al., 2017).Automated data processing streamlines operations by minimizing the need for human involvement in both image processing and the analysis of outcomes. CNNs have gained significant traction in a range of remote sensing fields over the past few years, emerging as a favored tool for many applications (Pena et al., 2024; Guo et al., 2020). In articles (Mullen et al., 2023; Aroma et al., 2024; Li et al., 2018; Isikdogan et al., 2019), CNNs were utilized for the identification of water bodies in satellite imagery. Nevertheless, the diverse spectral attributes exhibited by different bodies of water have presented difficulties in devising versatile methodologies applicable to a range of situations. (Li et al., 2018) focused on landscapes containing both land and sea features, while article (Isikdogan et al., 2019) analyzed raw Landsat images, requiring sophisticated computational resources. Some CNN designs such as (Müller et al., 2021, Badrinarayanan et al., 2017) produced unclear outputs and struggled to differentiate objects, resulting in subpar segmentation. These patterns involve a considerable

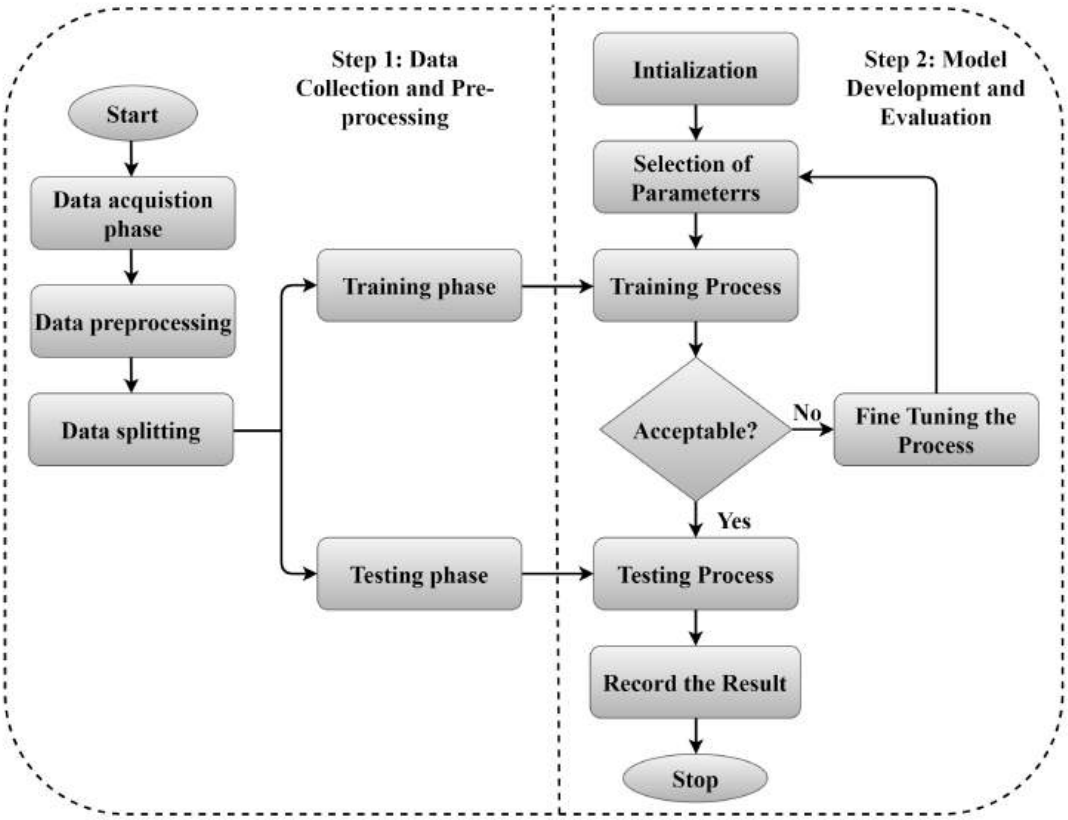


FIGURE 3.1: Integrated Conceptual Framework

amount of adjustable parameters. Water extraction methods employ standard procedures to acquire data on water availability using remote sensing images. In line with the need to analyze water resources, the proposed solution utilizes remote sensing (RS) imagery, as illustrated in FIGURE 3.1. To fulfill the objectives of this study, the key contributions of the proposed work are summarized as follows:

- A comprehensive architecture is designed to facilitate the extraction of water bodies through an end-to-end approach. The design of WaterNet simplifies the segmentation of non-water and water pixels.
- The expanded routes (decoders) and contracted routes (encoders) both incorporate dense layers, which utilize asymmetric convolutions to merge data gathered at different scales.

- The number of features available for training significantly impacts the overall complexity of the network in terms of computation. To address this asymmetric convolutions are used to reduce complexity and create more refined and less dense architecture.
- These modules have a dual function of improving expected images and assessing the smoothness and irregularity of pixels that outline the borders of water Impressive Results
- WaterNet demonstrates its efficacy and resilence by exhibiting outstanding performnace across a variety of datasets.

3.1 Literature Review

A variety of methodologies have surfaced for the remote sensing analysis of water bodies, enc ompassing techniques such as water index methodologies, threshold methods applied in both single and multiple-band scenarios, strategies for water monitoring, and several others (Bagwari et al., 2023). (McFeeters, 1996) pioneered the introduction of the Normalized Water Difference Index (NWDI) method for examining particular subjects. However, a drawback is evident in this approach concerning the analysis of regions between shadows and water bodies. (Xu, 2006) suggested a solution to tackle the previously mentioned challenge by integrating a mid-infrared band for normalization, leading to notable improvements in water body extraction outcomes. Moreover, a range of methodologies, including Support Vector Machines (SVM),k-means clustering, and k-nearest neighbor, have been developed for forecasting water availability (Liu et al., 2023a). Primarily employed for images with inadequate spatial resolution, these methods combine texture, shape, and spectral attributes fusion, playing a pivotal role in handling feature extraction complexities (Chen et al., 2020a).

In recent times, Deep Learning Frameworks (DLF) have attracted considerable interest from researchers owing to their efficacy in diverse remote sensing applications (Himeur et al., 2022). (Sharma et al., 2024) utilized

Fully Convolutional Networks (FCN) for semantic segmentation. Upsampling was employed in each decoder to integrate feature maps and deliver valuable outcomes. (Cai et al., 2024) devised Conditional Random Fields (CRF) and Recurrent Neural Networks (RNN) to improve the prediction capabilities of FCNs. However, this method requires extensive fine-tuning on large datasets (Arya et al., 2021). Various adaptations and alternative techniques have been proposed, such as task division for ship identification (Ren et al., 2022), multi-scale semantic labeling for sea-land and ship segmentation (Gharbia, 2023), and FCNs tailored specifically for water body extraction (Aroma et al., 2024). However, many FCN networks encounter difficulties in accurately segmenting border pixels of water bodies. To address this challenge, (Isikdogan et al., 2019) introduced DeepwaterMap and its refined iteration, DeepwaterMapV2. The Respective Region Field DeconvNet (RRF DeconvNet), a limited domain deconvolutional network, was proposed by (Miao et al., 2018).

The Structured Edge Network (Segnet) was initially introduced by (Gharbia, 2023) for land segmentation. It was later refined by (Badrinarayanan et al., 2017) with a significant decrease in trainable parameters. In remote sensing applications like water body extraction, numerous variations of Unet have been as well (Isikdogan et al., 2019). (Feng et al., 2018) introduced a Deep Convolutional Encoder-Decoder (DCED) architecture aimed at mitigating issues associated with the blurring of border pixels. This was achieved by employing techniques such as Fully Connected CRF (FCCRF) and Regional Restriction (RR). (Gonzalez et al., 2019, Tambe et al., 2021) introduced TernausNet, a model that integrates knowledge transfer approaches to handle high-resolution to very high-resolution images. While these frameworks exhibited impressive performance, they often contained numerous trainable parameters, leading to increased computational costs. Extensive research efforts have been dedicated to addressing the limitations outlined in Table 3.1. In response, the proposed WaterNet incorporates the concepts introduced by (Jonnala et al., 2023), effectively rivaling existing methods and surpassing them in performance.

TABLE 3.1: A comprehensive evaluation of the suggested method with previous research

Authors	Definition	Types of Bands	Satellite Imagery	Technology
Feyisa et al., 2014	An automated water extraction index proposed seeks to enhance the precision of water extraction by accounting for various types of environmental interference.	NIR	Yes	Traditional
McFeeters, 1996	The proposition involves employing the NDWI method for identifying open water characteristics.	NIR	Yes	Traditional
Huang et al., 2015a	An initial assessment is performed on VIR images to detect surface water.	NIR	Yes	Traditional
Xu, 2006	The suggestion is to use a modified NDWI, referred to as MNDWI, to improve the identification of open water characteristics.	NIR	Yes	Traditional
Kang et al., 2016	The hybrid approach is introduced to recognize water reservoirs during periods of low precipitation.	NIR	Yes	Traditional
Acharya et al., 2019	Evaluate machine learning algorithms for defining the water's surface in a Landsat-8 image taken in Nepal.	Landsat-8 OLI	Yes	Modern
Yang et al., 2017	This paper contributes to high-resolution city waterways using imagery ZY-3 spectral using K-means and SVM.	ZY-3 multispectral bands	Yes	Modern
Xiaobiao et al., 2021	Proposes a unique approach, the Two-Step Urban Water Index (TSUWI), for analyzing water with high resolution.	L-spectral bands	Yes	Modern
Himeur et al., 2022	The introduction of the CNN framework aims to identify various water sources within Landsat-8 imagery.	Multispectral bands	Yes	Modern
Chen et al., 2018	An innovative solution, assisted by deep learning, is introduced to extract water bodies within urban environments.	False color composition	Yes	Modern
Manocha et al., 2023	Leveraging a deep learning-inspired feature fusion technique to demarcate water bodies from Sentinel-2 imagery.	False color composition	Yes	Modern
Proposed	A novel framework, named WaterNet, is created to examine the patterns of water resources within a specific region.	False color composition with optical images	Yes	Modern

3.2 Proposed Work

This section offers a thorough structural outline of the proposed methodology. It also delves into the initial layers that enhance efficiency and reduce computational complexities within the network.

3.2.1 Architectural design of WaterNet

In this section, the segmentation of water bodies is conceptualized, necessitating the utilization of both lower-level and higher-level features (Kalla et

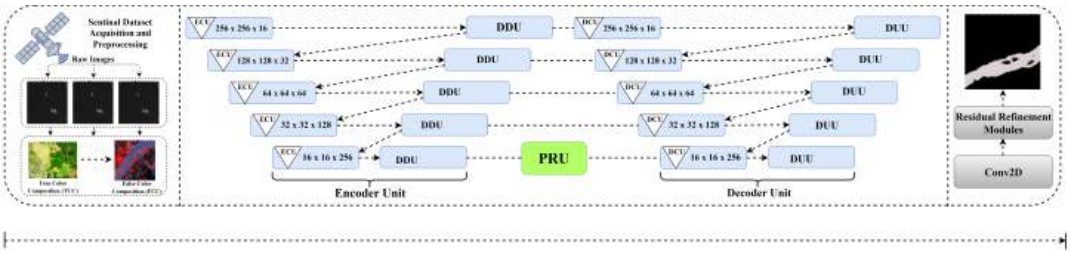


FIGURE 3.2: An overview of the proposed framework (WaterNet)

al., 2023). FIGURE 3.2 illustrates the overall design of the proposed WaterNet, which operates akin to an encoder-decoder architecture, conducting both convolution and deconvolution. WaterNet demonstrates computational efficiency by integrating features across a hierarchy derived from various convolutional layers. As illustrated in Fig.3.2, there exists a symmetrical relationship between the Decoder Unit (DCU) and Dense Up Unit (DUU), as well as the Encoder Unit (ECU) and Dense Down Unit (DDU). The network's design is structured in a W-shape, emphasizing its emphasis on water body extraction, hence earning the name "WaterNet". The dimensions of the imagery are denoted as W $\times$ H $\times$ C (Height, Width, and Channels). Each ECU and DCU is connected by black lines to its corresponding dense unit (DDU and DUU), while solid-dash black lines indicate additional skip connections established from the DDU to DCU. Through the pyramid unit (PRU), the fifth DDU is linked to the corresponding DCU. The input image undergoes down-sampling by the ECU, followed by a DDU after each ECU. The PRU generates feature maps with dimensions matching its input. Conversely, the DCU performs up-sampling on the input image, with a DUU block added after each DCU to maintain consistent feature map sizes throughout the network. The final output of the DUU undergoes a sigmoid operation, followed by residual refinement modules to generate the ultimate binary output. Two refining modules have been incorporated to enhance the quality of the predicted image.

3.2.2 Topologies of ECU, DDU, and PRU

FIGURE 3.3 provides a detailed breakdown of the structural elements of ECU, DDU, and PRU. Each ECU comprises two convoluted layers of size 3×3, followed by a pooling layer with a stride of two and a 2×2 filter, as illustrated in FIGURE 3.3(a). In this architecture, every convolutional layer integrates Rectified Linear Unit (ReLU) Batch Normalization (BN), and Convolutional Operations (conv2D) (Sener et al., 2024, Huang et al., 2022 to mitigate internal covariate shifts and mitigate overfitting (Huang et al., 2022). The Rectified Linear Unit (ReLU) activation function, denoted as $f(x) = \max(0, x)$, is utilized to augment non-linear operations, avoid saturation, and promote sparsity throughout the learning process (Pham et al., 2021). The pooling layer reduces the size of the input image by a factor of two and conducts spatial pooling using 2×2-pixel filters. Following each ECU, the number of filter channels is doubled. Previous research utilized VGGNet to illustrate variations of Unet (Farhadi et al., 2024) for image segmentation, which possessed simpler architectures but higher memory and computational demands (Iglovikov et al., 2018). To mitigate computational expenses, (Lian et al., 2020) aimed to decrease the overall number of trainable parameters by integrating dense blocks (DDU and DUU) into WaterNet after each ECU and DCU. The proposed WaterNet comprises two distinct dense units: five DDUs corresponding to five ECUs and five DUUs corresponding to DCUs, as shown in FIGURE 3.2. Each ECU produces a feature map, which serves as input for the associated DDU.

The structure of the DDU is extensively detailed in FIGURE3.3(b). Each convolutional layer within the DDU incorporates ReLU and conv2D. Spatial pooling is achieved through a max pooling layer with a 3×3 filter and a stride of one, followed by a convolutional layer with a 1×1 filter. Additionally, asymmetric convolutions are utilized to spatially decompose the 3×3 filter conv2D, splitting it into two stages: 1×3 and 3×1 filter conv2D. As depicted in FIGURE 3.3(b), two 3×3 filter convolutions are employed instead of a single 5×5 filter convolution, and each 3×3 convolution is subsequently

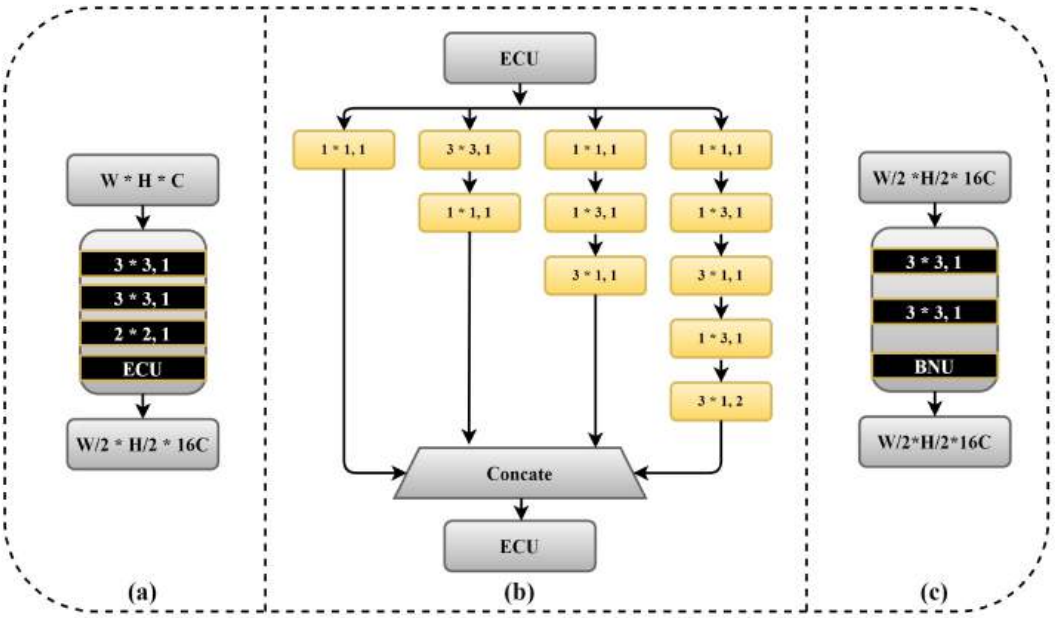


FIGURE 3.3: Illustrations of (a) ECU, (b) DDU, and (c) PRU

subjected to depth-wise separable convolutions. The incorporation of depth-wise separable transformations alleviates the computational burden on the entire network, resulting in fewer trainable parameters and faster network training (Szegedy et al., 2016). In this investigation, there are 10.33 million trainable parameters, representing a notable reduction from previous studies (Isikdogan et al., 2019, Liu et al., 2019). The 1×1 convolution serves dual purposes: applying non-linear activation (ReLU) and diminishing the parameter count to enhance computational efficiency before employing asymmetric convolutions. This empowers WaterNet to comprehend more intricate functions. The fifth DUU generates a feature map with a single output, which then serves as input for the subsequent ECU by amalgamating feature maps originating from various branches operating at different scales. The PRU receives the output of the fifth DDU, and the organizational structure of the PRU is precisely illustrated in FIGURE 3.3(c). Comprising two stages of 3×3 filter convolution with consistent spatial precision, the PRU follows the first DCU.

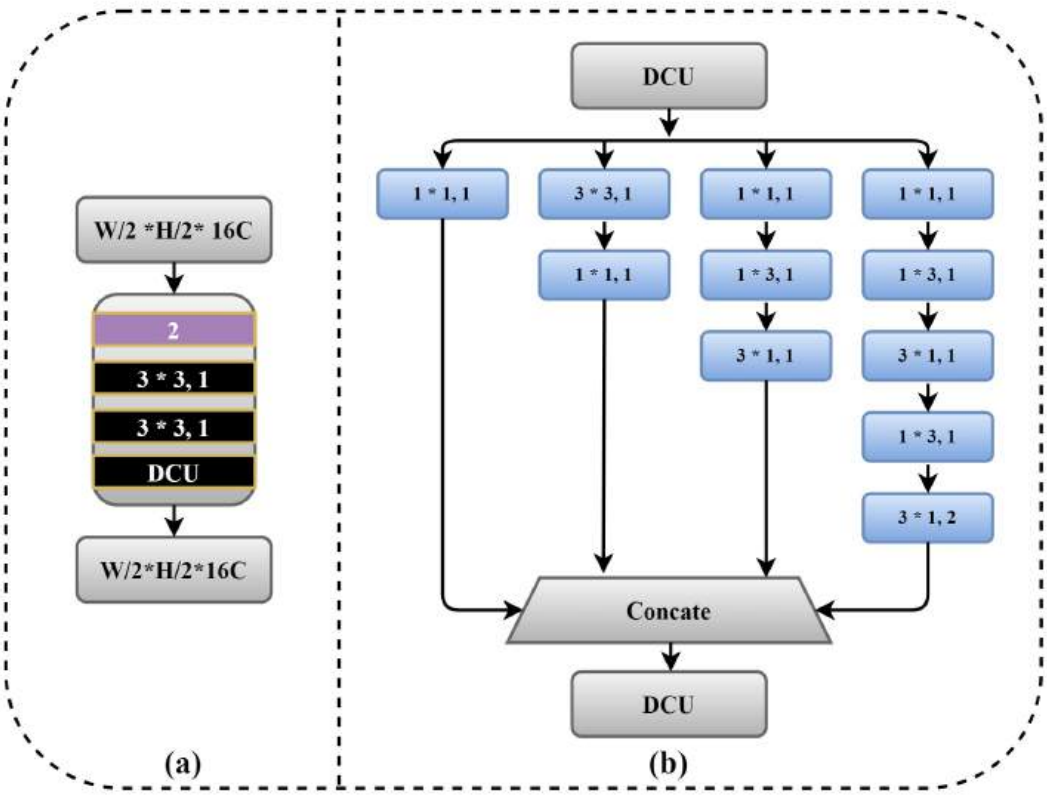


FIGURE 3.4: A demonstration of (a) DCU and (b) DUU

3.2.3 Topologies of DCU and DUU

The specific arrangement of the DCU is depicted in FIGURE 3.4(a). To maintain the input image's aspect ratio, each DCU enlarges the feature map initially. Next, a transposed convolutional layer is utilized with a 3×3 filter. Within each transposed convolutional layer, ReLU activation and a convolutional operation (conv2DT) are employed. The up-sampled feature map then moves to the DUU before progressing to the next DCU. As depicted in FIGURE 3.4(b), the primary distinction between DDU and DUU lies in the inclusion of a transposed convolutional layer instead of a conventional convolutional layer. The fifth DUU generates the final feature map with dimensions identical to the input image, utilizing a sigmoid activation function and a 1×1 filter conv2D, which is then employed in the output convolutional layer.

3.2.4 Dilated Convolutions

The aggregated parameters resulting from various convolutions using filters of different sizes are illustrated in FIGURE 3.5. After applying diverse filters (f) with a step size (s), the input images, originally sized 256 $\times$ 256 $\times$ 3, are resized to 128 $\times$ 128 $\times$ 3. In FIGURE 3.5(a), a 5 $\times$ 5 convolution generates 700 learnable elements. The formula $((P \times q \times r + 1)s)$ is utilized to compute the learnable parameters, where p and q represent the dimensions of the filter, indicating its width and height, respectively. Additionally, each filter includes a bias term represented by 1, and the numbers r and s denote the number of filters employed in the preceding and succeeding layers. The network's capability to produce distinct parameters using appropriate convolution factorization enhances training speed. Simply augmenting the network's depth does not guarantee optimal performance. Therefore, achieving equilibrium between the quantity of filters at each stage and the network's depth is essential for achieving overall improved performance. To enhance the computational efficiency of the network, FIGURE 3.5(b) illustrates the utilization of two 3$\times$3 convolutions instead of the conventional 5$\times$5 filter convolution, thereby expanding the network width-wise.

Notably, each 3$\times$3 convolution reduces the number of learnable elements to 280, resulting in a total of 560 parameters. This demonstrates a substantial reduction in trainable parameters through convolution factorization. Furthermore, asymmetric convolutions such as 1$\times$3 and 3$\times$1 convolutions are employed to factorize each 3$\times$3 convolution in FIGURE 3.5(c). With asymmetric convolution, each convolution comprises 90 trainable parameters, totaling 360 trainable parameters. This architecture reduces variables by approximately 50.63% compared to previous setups using 3$\times$3 and 5$\times$5 convolutions.

3.2.5 Residual Refinement unit

Two techniques are employed to improve the quality of the predicted images: Contrast stretching is implemented to enhance visual quality, while Laplacian of Gaussian (LoG) filtering is utilized in the first residual module.

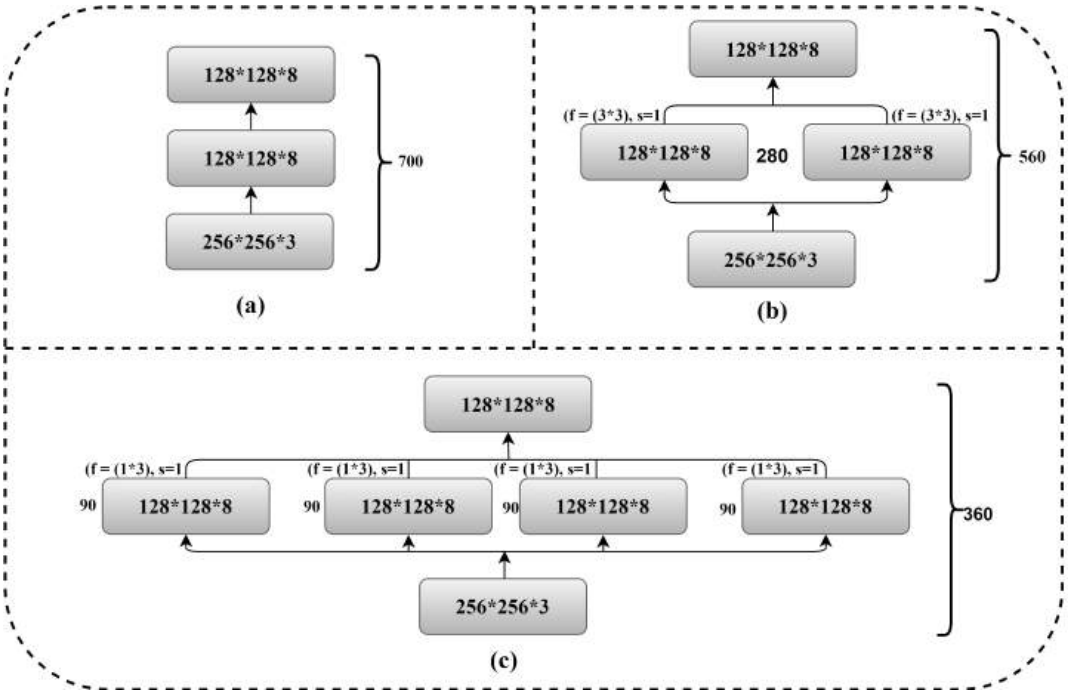


FIGURE 3.5: Influence of different filter sizes

The second residual module extracts edge features via Canny edge detection (Canny, 1986). In FIGURE 3.6(a), (b), and (c), the initial image, ground truths, and edge masks are displayed. FIGURE 3.6(d) illustrates the predicted image produced by WaterNet, along with the output of refinement module one, which generates the filtering result depicted in FIGURE 3.6(e). All test images are adjusted with a constant sharpness factor of 50.0, and the enhancement factor for refinement is set to 4.0. These enhanced images are subsequently smoothed using filters of two sizes: three by three and five by five. FIGURE 3.6(f) exhibits the edge output images created by refinement module two, which receives the filtering output. Accurately delineating the boundary pixels of the water body is a crucial step. In the second refinement module, less prominent pixels are identified as faint border pixels, while prominent ones are classified as robust border pixels to precisely outline the water body's perimeter. The refinement modules offer the advantage of enhancing the image's aesthetics and sharpness, facilitating the

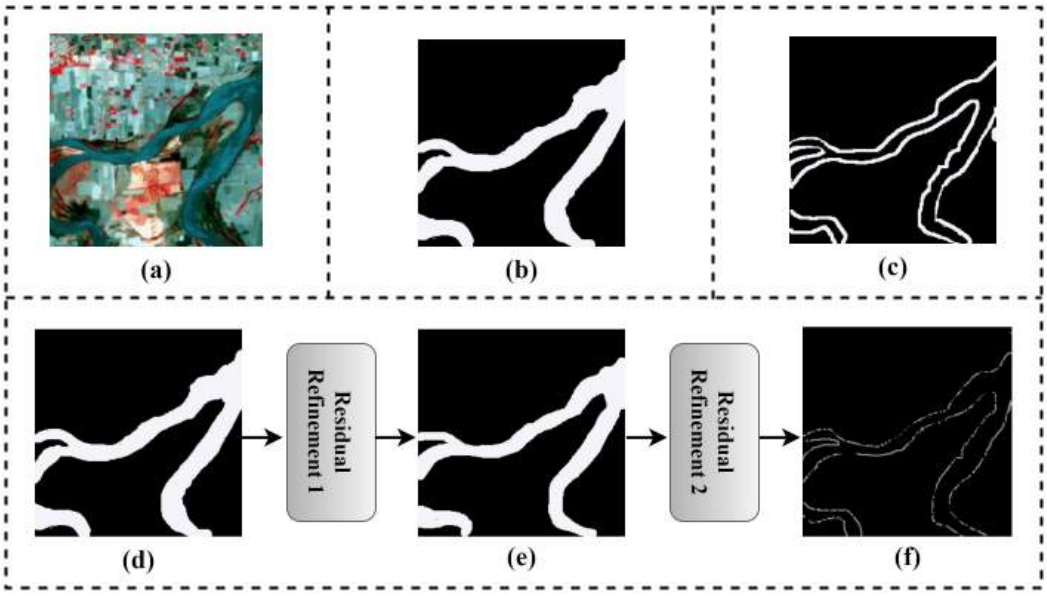


FIGURE 3.6: (a) Input image, (b) Ground truth, (c) Border mask, (d) Predicted image, (e) Refined output, (f) Second residual refinement output

analysis of edge information. This aids in the easy identification and examination of incorrectly categorized boundary pixels in the edge input images. Additionally, this paper also calculates the normalized skip cuts $N_{Skipcut}$ as a segmentation criterion as depicted in Eq.(3.1)

$$N_{\text{Skipcut}_C(W)} = \sum_{c=1}^{C} \frac{\text{cut}(A_c, W - A_c)}{\text{assoc}(A_c, W)} = \sum_{c=1}^{C} \frac{\sum_{u \in A_c, w \in W - A_c} m(u, w)}{\sum_{u \in A_c, t \in W} m(u, t)} \tag{3.1}$$

In this context, A_c represents the group of pixels within segment C, and W denotes the set containing all pixels. The variable M indicates the weight between two pixels. However, because of the non-differentiable nature of the argmax product, finding the relevant gradient during reverse propagation is not feasible. Instead, a differentiable version of the Normalized $N_{Skipcut}$ loss is utilized, allowing for gradient updates during reverse propagation as depicted in Eq.(3.2).

$$
\begin{aligned}
K_{\text{soft}} - N_{\text{skipcut}}(W, C) &= \sum_{c=1}^{C} \frac{\text{cut}(A_c, W - A_c)}{\text{assoc}(A_c, W)} \\
&= C - \sum_{c=1}^{C} \frac{\text{assoc}(A_k, A_k)}{\text{assoc}(A_k, A_k)} \\
&= C - \sum_{c=1}^{C} \frac{\sum_{u \in W, v \in V} (w(u,v) p(u = A_k) p(v = A_k))}{P_{u \in A_c, t \in W} (w(u,t) \cdot p(u = A_c))} \\
&= C - \sum_{c=1}^{C} \frac{\sum_{u \in W} (\sum(u = A_c)) \sum_{u \in W} (w(u,w) \cdot p(w = A_c))}{\sum_{u \in W} (p(u = A_w)) \sum_{t \in W} (w(u,t))}
\end{aligned} \tag{3.2}
$$

The encoder rapidly computes $p(u = A_k)$, providing an estimate of the probability that node u belongs to class A_k. Train U_{ENC} to minimize the $K_{soft} - N_{Skipunit}$ loss. This might lead to a reduction in the overall standardized separation between groups while enhancing the standardized connection within the groups. The loss is denoted in Eq.(3.3).

$$
J_{\text{loss}} = \|kX - \text{UDec}(\text{UEnc}(X, M_{\text{Enc}}), W_{\text{Dec}})\|_2^2. \tag{3.3}
$$

In this context, M_{Enc} represents the encoder parameters, M_{Dec} denotes the decoder parameters, and X signifies the input image. WaterNet is trained to minimize J_{Loss} between new images and original inputs. Additionally, to further minimize the cut, U_{Enc} is trained. Within the encoder layer, U_{Enc} is trained to minimize $j_{soft} - N_{Skipcut}$, thereby enhancing segmentation association and reducing segment disassociation. The network endeavors to find a harmony between the precision of fresh images and stability in the encoded representation layer by repeatedly employing j_{Loss} and $j_{soft} - N_{Skipcut}$. While deep CNNs with maximally pooled layers have been effective in capturing high-level feature information from inputs, increasing consistency and expanding receptive fields may lead to reduced localization accuracy. The absence of smoother constraints could result in inadequate item separation,

particularly in tasks necessitating pixel-level categorization. WaterNet addresses this issue by employing a soft-established cut loss and skipping multiple layers, thereby improving object border localization. Moreover, integrating responses at the final U_{Enc} layer using a residual refinement module enhances segmentations with finely delineated borders. The residual refinement module utilizes the energy value in the following manner as denoted in Eq.(3.4).

$$E(X) = X_u \Phi(u) + X_{u,w} \Psi(u, w) \tag{3.4}$$

The variables u and w represent pixels within the given dataset X. The unary probability $\phi(u) = -\log p(u)$ is determined by the softmax layer in U_{Enc}, where $p(u)$ denotes the predicted probability of label annotation. The pairwise potential $\Psi(u, w)$ calculates the relative penalties for assigning different labels to two identical pixels and involves the application of two Gaussian kernels in distinct feature spaces. The evaluation of individual pixels along the original over-segmented division borders is conducted by amalgamating local signals at various scales and global boundary metrics utilizing spectral clustering methods and is mathematically denoted in Eq.(3.5)

$$gPb(x, y, \theta) = \sum_s \sum_i \beta_{i,s} G_{i,\sigma(s)}(x, y, \theta) + \gamma_s Pb(x, y, \theta) \tag{3.5}$$

In this scenario, s denotes the boundaries, i represents the channels containing features like hue, saturation, and appearance, and $G_{i\sigma(s)}(x, y, \theta)$ quantifies the dissimilarity between the two halves of an object with a diameter of $\sigma(s)$ situated at (x, y) in channels i at an angle θ. The mPb signal encompasses all boundaries in the image, while the sPb signal captures the organization of the image's most pronounced curvature (g). Subsequently, a weight matrix W is established for $J_{Soft} - N_{Skipcuts}$ represented in Eq.(3.6).

$$w_{ij} = e^{-\frac{\|F(i)-F(j)\|^2}{2\sigma_F^2}} I^* \begin{cases} e^{-\frac{\|X(i)-X(j)\|^2}{2\sigma_X^2}}, & \text{if } \|X(i) - X(j)\|^2 < r \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

Here, $X(i)$ and $G(i)$ denote the distance between the pixel position.

3.3 Experimental Detail

To showcase the effectiveness and capabilities of the newly introduced WaterNet, a comprehensive study was undertaken, encompassing both qualitative and quantitative analyses using Sentinel-2 images. In this section, the model's optimal performance is demonstrated utilizing the Sentinel-2 image collection. During the evaluation phase, the system setup consists of the following specifications: Central Processing Unit (CPU): RYZEN 5 2.8 GHz, Operating System: Ubuntu 18.4 LTS, Graphics Processing Unit (GPU): NVIDIA GEFORCE GTX, Programming Language: Python. The entire process is outlined in FIGURE 3.2 and divided into three stages: (i) data retrieval, (ii) preprocessing, and (iii) prediction of water resources. WaterNet is trained using the freely available TensorFlow open-source framework (Agrawal et al., 2019). The adjustments made to the model's parameters include setting the input image dimensions to $(256 \times 256 \times 3)$ and the ground truth to $(256 \times 256 \times 1)$. A batch size of different sizes is utilized, along with 100 epochs, and the Adam optimizer (Kingma et al., 2014) is employed with a learning rate. During the initial training phase, data augmentation techniques such as image translation, rotation, and flipping are applied to each patch (Krizhevsky et al., 2012). The augmented data, sized $(256 \times 256 \times 3)$, is used for training WaterNet. Weight initialization for the DDU and DUU convolutional and deconvolutional layers follows a random approach based on (Tan et al., 2021). Unlike many existing networks that focus on semantic segmentation with multiclass labeling (Arya et al., 2021, Isikdogan et al., 2017). (Isikdogan et al., 2019, Xu et al., 2018), and (Xiang et al., 2023) concentrated on distinguishing two classes, although they utilized datasets different from Sentinel-2. Due to these variations, WaterNet was successfully trained without relying on pre-trained models.

3.3.1 Data acquisition

Multiple networks underwent training and evaluation to validate their effectiveness using diverse data collected from various sensors. (Isikdogan

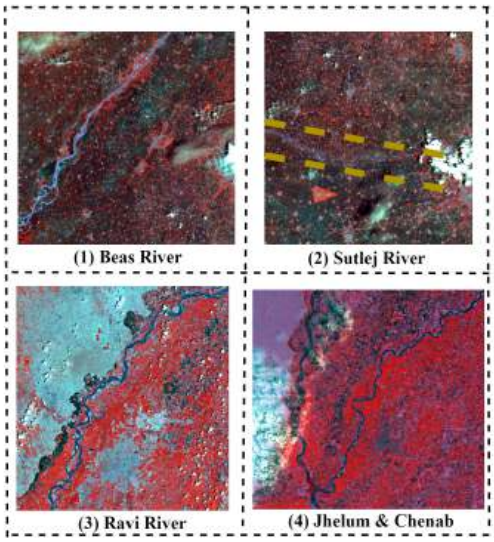


FIGURE 3.7: Highlighted water bodies in Sentinel-2 images

et al., 2019) utilized Landsat-8 images for testing and training their model, while (Chen et al., 2020a) employed natural-color images from Google Earth, and (Gonzalez et al., 2019) utilized the second Sentinel data. The variety of dataset sources makes it challenging to compare multiple models on a consistent dataset. To streamline the comparison, the proposed model utilized images extracted from specific locations from the Copernicus Open Access Hub (COAH). These images consist of Sentinel-2 false-color composites depicting various water bodies in different regions. Although the initial quality of Sentinel-2 images was 10981×10981, they were resized to 256×256 to align with the model's input requirements. This study specifically focuses on four bands of the Sentinel-2 images: VNIR, red, blue, and green. The proposed approach evaluates specific land and water body labels within the dataset. Table 3.2 offers a comprehensive summary of the data. In FIGURE 3.7, particular attention is given to small water bodies in the study to conduct a close analysis of the water bodies.

TABLE 3.2: Detailed description of the image in the dataset

Type of Dataset	Satellite Imagery	Pixel	Category	Data Range
Sentinel-2 Image Dataset	False-color composite images	256*256 pixel density (VNIR, blue, green, and red)	Small water bodies	January 2021 - February 2023

3.3.2 Image Pre-processing

The selected Sentinel-2 images must undergo normalization as the collection comprises red, VNIR, blue, and green bands. The objective of choosing the normalization technique is to align the images with the system's input requirements. Several procedures were employed to preprocess the images, including radiometric correction (Cai et al., 2024), atmospheric condition adjustment (Arya et al., 2021), and geometric correction (Ren et al., 2022). The study employs radiometric correction methods, including histogram alignment, recognizing intersecting areas among images, and utilizing pseudo-invariant characteristics. The challenges presented by atmospheric conditions, including distortion and the necessity for object removal, necessitate appropriate solutions. The proposed research employs adaptive methods to address severe atmospheric conditions. Geometric corrections are pivotal in accurately positioning an image, particularly in scenarios involving multiple geometric registrations such as co-registration, geo-referencing, and orthorectification. To increase the number of images and mitigate overfitting, a common issue in deep learning (Gharbia, 2023,Isikdogan et al., 2017, Cheng et al., 2016), data augmentation techniques are utilized, including rotation, clipping, distortion, shifting, and flipping, as illustrated in FIGURE 3.8. Additionally, the ground truth for water segmentation is annotated using Label Studio, a versatile software application renowned for its data annotation capabilities. Conversely, 24,255 non-overlapping clipped patches and their corresponding ground truths are generated using the Python tool Anaconda.

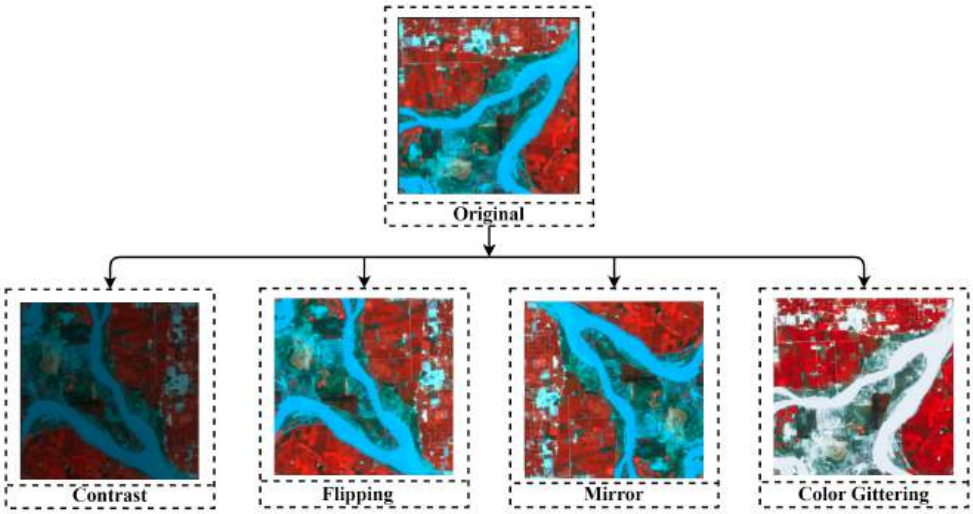


FIGURE 3.8: Illustrations of various data augmentation techniques

TABLE 3.3: Performance of the proposed framework across varying data ratios.

Images	Proportion	Training	Test
Original Dataset	70:30	7702	1680
	75:25	7982	1400
	80:20	8262	1122
Augmented Dataset	70:30	8967	4655
	75:25	9571	4851
	80:20	10430	3992

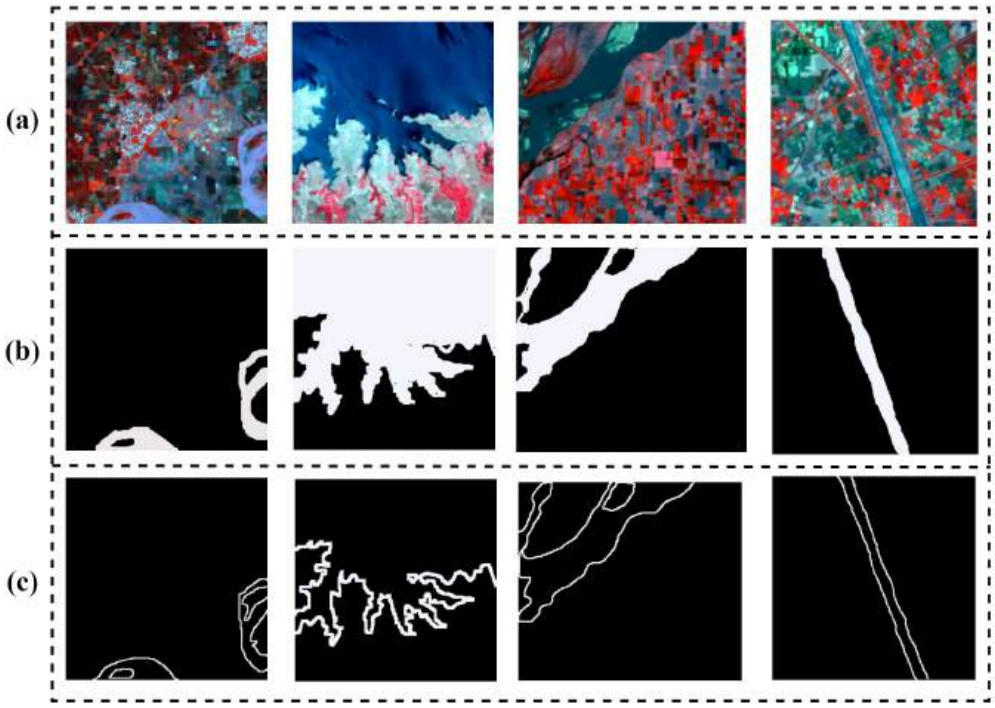


FIGURE 3.9: (a) Input images, (b) Manually annotated ground truths, and (c) Sample of edge masks.

FIGURE 3.9 showcases examples from the representative dataset, encompassing a diverse range of samples featuring river-like formations with varying widths, alongside non-river features like sea-land landscapes, islands, and river beds. The meticulous annotation of ground truths is evident in the second row of FIGURE 3.9, while the third row illustrates the creation of edge masks based on these ground truths, employing the methodology outlined in (Canny, 1986). Initially, the collection comprises 9,382 meticulously selected images. To fulfill the high data requirements for training deep learning algorithms, various data augmentation techniques were applied to expand the original dataset to 14,423 augmented images.

3.3.3 Assessment Metrics

WaterNet's performance was assessed by applying it to a carefully curated dataset of water bodies. To gauge WaterNet's proficiency in semantic segmentation, three commonly acknowledged metrics were employed: mean intersection over union (IoU), class average accuracy (C), and global accuracy (G) (Zhang et al., 2016). Class average accuracy (C) portrays the accuracy of predictions for all classes, indicating the percentage of correctly predicted pixels as per Eq. (3.7), (3.8), and (39).

$$G = \frac{\sum_i m_{ii}}{\sum_i n_i} \tag{3.7}$$

$$C = \frac{1}{mc} \sum_i \frac{m_{ii}}{n_i} \tag{3.8}$$

$$\frac{I}{U} = \frac{1}{mc} \sum_i \frac{m_{ii}}{n_i} + \sum_i m_{ji} - m_{ii} \tag{3.9}$$

In Eq's. (3.7)–(3.9), the indices i and j refer to the spatial location of the input feature map in the vertical and horizontal dimensions, respectively. The range of values for these indices is defined as $i = 0, 1, \ldots, H-1$ and $j = 0, 1, \ldots, W-1$, where H and W denote the height and width of the input feature map.This range ensures that the convolutional kernel is applied to every valid spatial position within the feature map, covering the entire input domain. The indexing starts from 0, which is standard in most deep learning frameworks and ensures compatibility with 0-based array access patterns. In this scenario, m_{ij} denotes various categories, n_i signifies the total number of pixels in category i anticipated to be in category j, and m_{ii} represents the count of pixels in category i, encompassing both false positives (FP) and true positives (TP). Furthermore, the assessment of semantic segmentation entails three commonly utilized metrics derived from Eq. (3.10), (3.11),(3.12) and (3.13). These metrics hold particular significance in the realm of water body extraction.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.10}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.12}$$

$$F - measure = \frac{2 \cdot P \cdot R}{P + R} \tag{3.13}$$

The terms TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, correspondingly. Eq's. (3.8) and (3.10) represent two forms of accuracy used to evaluate segmentation performance. Eq. (3.8) computes the overall pixel-wise accuracy by measuring the ratio of correctly classified pixels to the total number of pixels, including both true positives and true negatives. This metric offers a general view of model performance; however, it may provide inflated values when class imbalance exists, as the dominant class heavily influences the result.On the other hand, Eq. (3.10) calculates the mean accuracy by averaging the classification accuracy for each class individually. This metric ensures equal importance is given to all classes, regardless of their frequency in the dataset. It is particularly valuable in remote sensing tasks where one class (e.g., water bodies) may be underrepresented. By evaluating per-class accuracy, Eq.(3.10) offers a more balanced and interpretable performance assessment, complementing the global accuracy presented in Eq.(3.8).

3.3.4 Performance Evaluation

Moreover, the datasets were manually partitioned to create distinct testing and training sets, with proportions of 70:30, 75:25, and 80:20. Table 3.3 presents a summary of the accuracy measures for both datasets.

Utilizing the proposed model on the augmented dataset, as opposed to

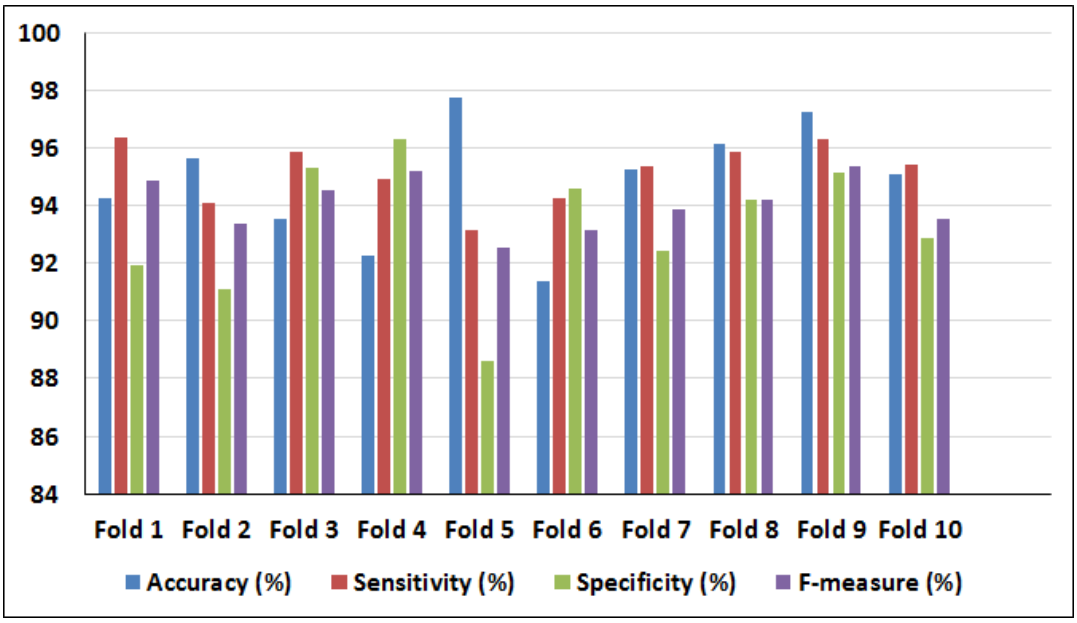


FIGURE 3.10: The assessment of WaterNet performance on Sentinel-2 images at a resolution of 256×256 pixels

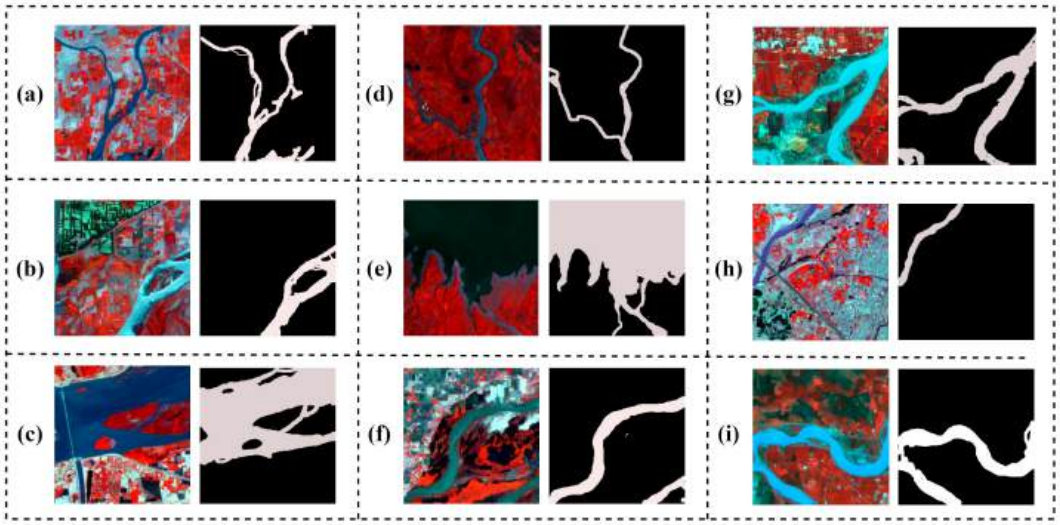


FIGURE 3.11: The identification of water bodies

the original dataset, yielded improved results, particularly with an 80:20 ratio. This ratio was chosen to thoroughly assess the proposed model's performance, as outlined in Table 3.4.

TABLE 3.4: Performance analysis monitoring of WaterNet.

Model	Ratio	Accuracy	Sensitivity	Specificity	F-measure
Original Dataset	70:30	89(%)	87(%)	84(%)	86(%)
	75:25	90(%)	89(%)	86(%)	85(%)
	80:20	91(%)	90(%)	88(%)	89(%)
Augmented Dataset	70:30	92(%)	89(%)	86(%)	88(%)
	75:25	93(%)	91(%)	89(%)	87(%)
	80:20	97(%)	94(%)	91(%)	92(%)

Visual representations of the computed outputs from the proposed method can be observed in FIGURE 3.10 and FIGURE 3.11. The results demonstrate the method's ability to accurately detect various types of water bodies in diverse locations, including wetlands and obstacles. To mitigate biases in the dataset, a range of large lakes and ponds were deliberately selected, with images captured at different time intervals. Moreover, the proposed model's effectiveness is evaluated through a 10-fold leave-one-out testing strategy to ascertain predictive accuracy. Each iteration involves training with 9 out of 10 folds while holding one fold to assess prediction accuracy, ensuring that each fold serves as the test set in rotation. FIGURE 3.10 illustrates the average prediction outcomes for water identification across ten folds. It is worth noting some misclassifications, particularly in distinguishing between rivers and lakes, where 4% of wetlands are erroneously classified as lakes. This misclassification issue may stem from insufficient training samples, which could be addressed by expanding the training dataset for the relevant class. Consequently, the proposed framework achieves an average prediction accuracy of 0.97% using the k-fold cross-validation approach.

3.3.5 Sensitivity Parameter Analysis

Several experiments were conducted to evaluate the effect of different learning rates while maintaining a fixed number of epochs. Three different learning rates such as 0.001, 0.01, and 0.1 were employed across different batch

sizes, as shown in Tables 3.5, 3.6, and 3.7. The results demonstrate that both inadequate and excessive learning rates hinder the task of remote image segmentation. In particular, a low learning rate results in slow convergence and the loss of significant image features, while an excessively high learning rate can cause the model to miss important patterns, leading to poor generalization and overfitting. When the learning rate is at a medium level of 0.01, the most consistent and reliable results have been observed across various batch sizes. This learning rate facilitated balanced convergence, preventing both underfitting and overfitting. It allowed the model to effectively capture important image features while avoiding excessive sensitivity to noise or irrelevant patterns. As a result, optimal accuracy was achieved at this medium learning rate, enabling precise extraction during the segmentation task. Thus, the medium learning rate of 0.01 strikes the best balance between model performance and learning stability.

TABLE 3.5: Model performance with a learning ratio of 0.001.

Epochs	Batch size	Accuracy	Sensitivity	Specificity	F-measure
100	8	60(%)	61(%)	59(%)	60(%)
100	16	68(%)	64(%)	66(%)	65(%)
100	32	70(%)	72(%)	69(%)	71(%)
100	64	76(%)	78(%)	75(%)	77(%)
100	128	80(%)	79(%)	81(%)	80(%)

TABLE 3.6: Model performance with a learning ratio of 0.01.

Epochs	Batch size	Accuracy	Sensitivity	Specificity	F-measure
100	8	77(%)	70(%)	75(%)	72(%)
100	16	84(%)	79(%)	82(%)	80(%)
100	32	90(%)	83(%)	88(%)	85(%)
100	64	95(%)	85(%)	92(%)	89(%)
100	**128**	**97%**	**94%**	**91%**	**92%**

To evaluate the robustness of the WaterNet model, noise ratios of 10%, 20%, and 30%, were introduced incorporating Gaussian noise, salt-and-pepper noise, and label noise to simulate real-world imperfections. The model was trained using fixed epochs, varying batch sizes, and different learning rates.

TABLE 3.7: Model performance with a learning ratio of 0.1

Epochs	Batch size	Accuracy	Sensitivity	Specificity	F-measure
100	8	90(%)	79(%)	85(%)	82(%)
100	16	93(%)	81(%)	89(%)	85(%)
100	32	94(%)	84(%)	90(%)	87(%)
100	64	96(%)	85(%)	92(%)	89(%)
100	128	96(%)	86(%)	93(%)	90(%)

Performance metrics such as Accuracy, Sensitivity, Specificity, and F-measure were measured at each noise level. At a 10% noise ratio, the model maintained a strong accuracy of 97%, demonstrating its stability under low-noise conditions. Even as the noise ratio increased to 30%, the accuracy only slightly decreased to 96%, showcasing the model's reliability in handling higher noise levels. Similarly, the F-measure followed a consistent pattern, further confirming the model's resilience against moderate noise. These results highlight WaterNet's ability to manage noise-induced disruptions effectively, proving its robustness for image segmentation tasks in noisy environments. A key contributor to this robust performance is the inclusion of residual refinement modules in the model architecture. These modules enhance image clarity by utilizing two types of filters: the Canny filter and a Gaussian filter. The Canny filter efficiently detects edges, while the Gaussian filter helps to reduce noise in the output images, allowing for more precise segmentation. This dual-filter system works to refine the results, ensuring that noise from the input data does not significantly impact the overall performance. FIGURE 3.12 summarizes the model's sensitivity to different configurations, illustrating that WaterNet remains reliable and effective even in challenging, noisy environments.

3.3.6 Performance Comparative Analysis

WaterNet was compared against NDWI(McFeeters, 1996), Unet (Müller et al., 2021), Deepunet(Li et al., 2018), Segnet(Badrinarayanan et al., 2017), DeepwaterMap (Isikdogan et al., 2017), and MFFA(Jonnala et al., 2023), all of

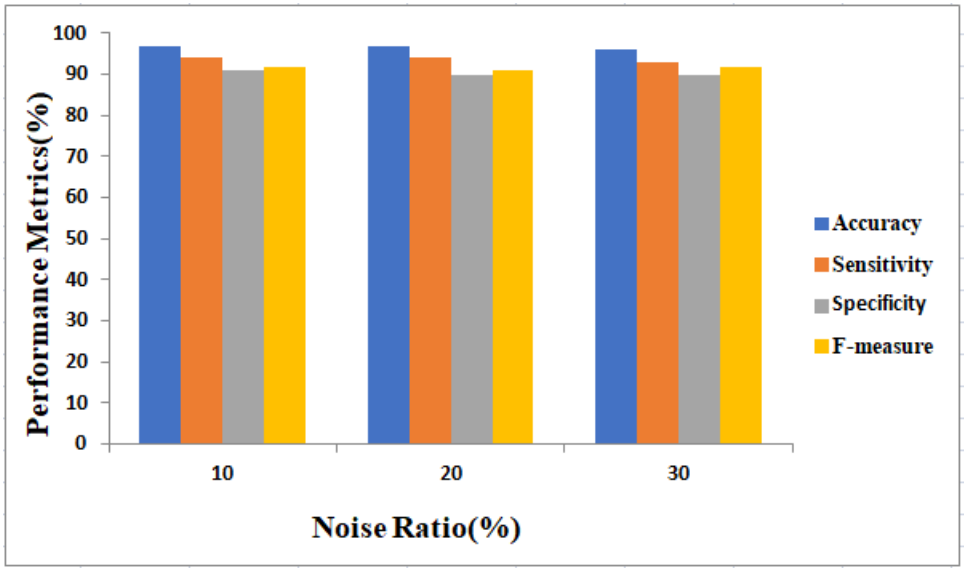


FIGURE 3.12: Model sensitivity to varying noise ratio.

which were fine-tuned using water body datasets. WaterNet showcased substantial performance advantages over these alternative architectures, as illustrated in FIGURE 3.13. NDWI-produced images tended to exhibit blurriness compared to other methods. Although Unet performed slightly better than NDWI, it remained prone to pixel misclassifications. Notably, Unet showed improved performance in scenarios involving non-river structures, resulting in fewer pixel misclassifications compared to river-like scenes. Deepunet, originally designed for detection tasks, effectively segmented river-like structures but struggled to differentiate images with non-river features. This indicates that lower-level layers captured localized characteristics with more confined receptive fields, whereas more complex layers led to an increased rate of false positives (non-water) in instances displaying non-river patterns. DeepWatermap excels in handling images with characteristics other than rivers but struggles when dealing with river-like features. It encounters challenges in distinguishing between dark areas and steep mountain ranges. In such scenarios, pixels representing shadows and mountain ridges are erroneously classified as either water or non-water pixels, with shadows often mislabeled as non-water pixels. In regions devoid of rivers, Deepwatermap erroneously identifies land pixels as water pixels. Despite its effectiveness in

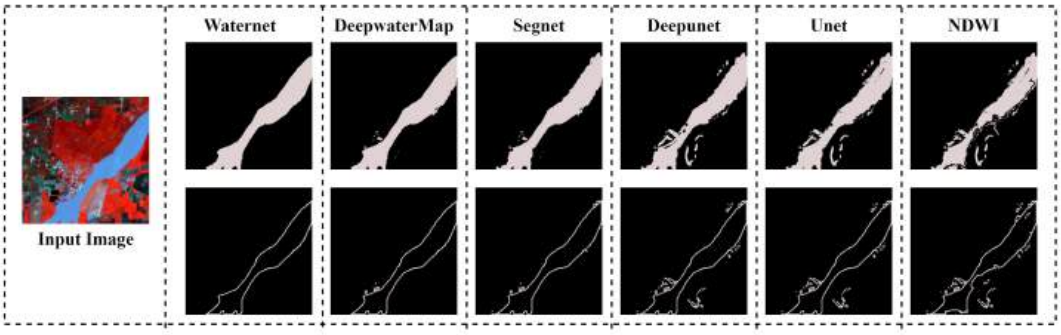


FIGURE 3.13: Results on sample input image The first column shows filtering outputs (residual refinement module one) and 2nd row shows edge data(residual refinement module two). Each row represents (a) WaterNet, (b) MFFA, (c) DeepwaterMap, (d) Segnet, (e) Deepunet, (f) Unet, and (g) NDWI model

various aspects, MFFA faces limitations in accurately extracting small water bodies from false-color composite Sentinel-2 images. In contrast, WaterNet demonstrates accurate discrimination between ground and non-water data, as depicted in FIGURE 3.13(b). The results shown in FIGURE 3.13 illustrate the filtering and edge outputs after refinement one and two have been applied. In this analysis, a single representative image randomly chosen from the dataset is examined. The first row shows filtered images, representing the result of refinement one (filtered output), while the second row displays the edge data obtained after refinement two has been applied.

In FIGURE 3.13, the first row (labeled as (a)–(g)) displays the outcomes of NDWI, Unet, Deepunet, Segnet, DeepwaterMap, MFFA, and the proposed WaterNet. It's apparent that after the initial refinement, Unet effectively reduces blurriness. DeepwaterMap, MFFA, and WaterNet generate refined images with clear and sharp boundaries. However, applying the initial refinement to Unet and NDWI results in an increase in misclassifying non-water pixels as water, leading to more false negatives. The second refinement module is then utilized to capture edge details, as illustrated in the second row of images in FIGURE 3.14

Table 3.8 offers a comparison of the convolution count, total trainable

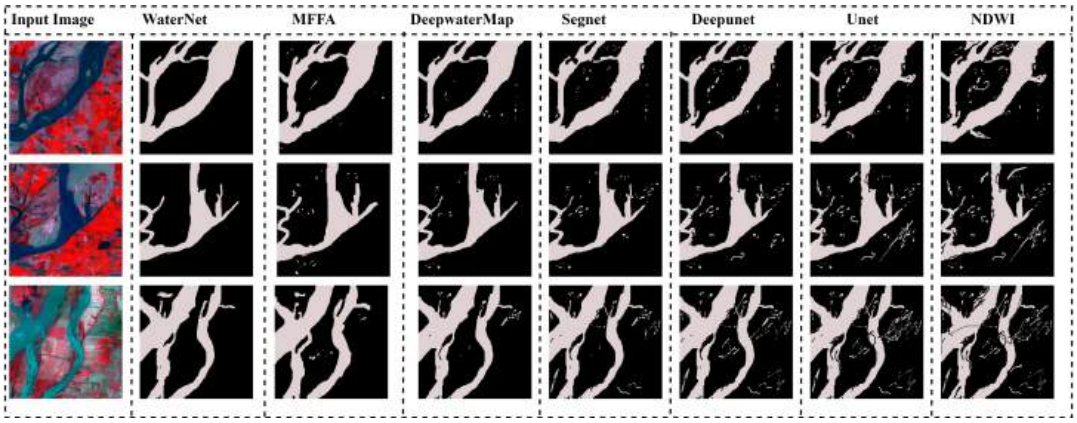


FIGURE 3.14: Results on sample input image The first column shows filtering outputs (residual refinement module one) and 2nd row shows edge data(residual refinement module two). Each row represents (a) WaterNet, (b) MFFA, (c) DeepwaterMap, (d) Segnet, (e) Deepunet, (f) Unet, and (g) NDWI model

TABLE 3.8: The assessment of comparison results among different networks, with emphasis on highlighting the optimal values in the table

Metrics	WaterNet	MFFA	DeepwaterMap	Segnet	Deepunet	Unet	NDWI
Accuracy	**97%**	96%	95%	94%	91%	90%	86%
IOU	**93%**	92%	89%	86%	86%	84%	80%
Classes	**97%**	95%	94%	91%	92%	90%	89%
Sensitivity	**88%**	87%	86%	84%	81%	80%	76%
Specificity	**89%**	88%	87%	86%	85%	84%	82%

parameters, and the corresponding processing duration. In affirming the model's efficacy, it's essential to underscore its noise ratio performance. Our comprehensive assessments demonstrate that the model achieves a noise ratio as minimal as 15%, marking a substantial advancement compared to modern state-of-the-art models. This reduction in noise ratio not only signifies enhanced accuracy but also underscores the model's ability to distinguish images from noise, thereby strengthening its utility in real-world applications. By minimizing the influence of extraneous factors, the model demonstrates unparalleled fidelity, ensuring reliable performance across varied contexts and exhibiting a stronger generalization capability in complex

TABLE 3.9: The duration for training and testing various models for predicting water bodies from Sentinel-2 images

Models	Layers	Trainable Time	Testing Time
WaterNet	**23**	**200m**	**50.81m**
MFFA	25	250m	80.81m
DeepwaterMap	20	260m	89.33m
Segnet	25	280m	179m
Deepunet	32	300m	186m
Unet	23	380m	205m

scenarios and across various cloud types.

3.3.7 Model Efficiency Assessment

Table 3.9 provides comprehensive information on the training and testing processes for each advanced model. The results indicate that Unet requires a longer time for both training and testing compared to other models. In contrast, the proposed model demands significantly less time for both processes. Specifically, the proposed model completes training in 200 minutes, while Unet, Deepunet, Segnet, Deepwatermap, and MFFA take 380, 300, 280, 260, and 250 minutes respectively. Therefore, based on both qualitative and quantitative evaluations, it is clear that the innovative WaterNet surpasses alternative methods. Furthermore, the time complexity of the proposed solution is characterized by O(n3), indicating that the execution time increases cubically with the input image size. Additionally , the space complexity is denoted as O(nlogn) suggesting the memory usage of the proposed solution. These complexity metrics collectively validate the efficiency and performance of the proposed solution.

3.3.8 Dataset-based performance Analysis

To validate the effectiveness of our proposed model, we thoroughly reviewed the experimental data obtained from the CIFAR-10 (Cukierski, 2013), AquaSat(Ross et al., 2019), and COCO-10 (Lin et al., 2014) datasets. This meticulous examination aimed to confirm the reliability of the model derived from our

experiments. Table 3.10 displays the test outcomes for these three widely acknowledged public datasets. As illustrated in Table 3.10, the proposed model consistently exhibits exceptional performance in both classification and recognition tasks across these datasets. Remarkably, in the classification task using the COCO-10 dataset, the model achieves an impressive recognition accuracy of 95%. The applied research methodology adeptly captures and extracts crucial feature information from the datasets through a range of diverse experiments. This method notably enhances the accuracy of image classification tasks, demonstrating robust performance across multiple datasets. Particularly noteworthy is the model's resilience and versatility, as evidenced by its strong compatibility with various types of data. This highlights the capacity and adaptability of our research model in tackling diverse challenges within the domain.

TABLE 3.10: Results of the research technique described in this work on three available datasets

Dataset Name	Total Images	Accuracy	Classes	IoU	Sensitivity	Specificity	F-measure
CIFAR-10 (Buma et al., 2018)	30,000	0.86(%)	0.89(%)	0.85(%)	0.81(%)	0.82(%)	0.83(%)
AquaSat (Dronova et al., 2015)	10,000	0.92(%)	0.93(%)	0.89(%)	0.85(%)	0.86(%)	0.91(%)
COCO-10 (Hahmann et al., 2008)	29,000	0.95(%)	0.95(%)	0.89(%)	0.84(%)	0.85(%)	0.94(%)

3.4 Ablation Study

The innovative architecture of WaterNet introduces specialized units known as DDU (Dense Down Unit) and DUU (Dense Up Unit). In this comparative analysis, three variations of WaterNet are pitted against the baseline Deepunet. The first variation, WaterNet+, integrates DDU in the encoder path while excluding DUU in the decoding path. The second variation incorporates DUU in the decoding path but excludes DDU (DownDense Unit) in the encoder path. Lastly, the third variant, WaterNet++, features both DUU in the decoding path and excludes DDU in the encoder path. This thorough

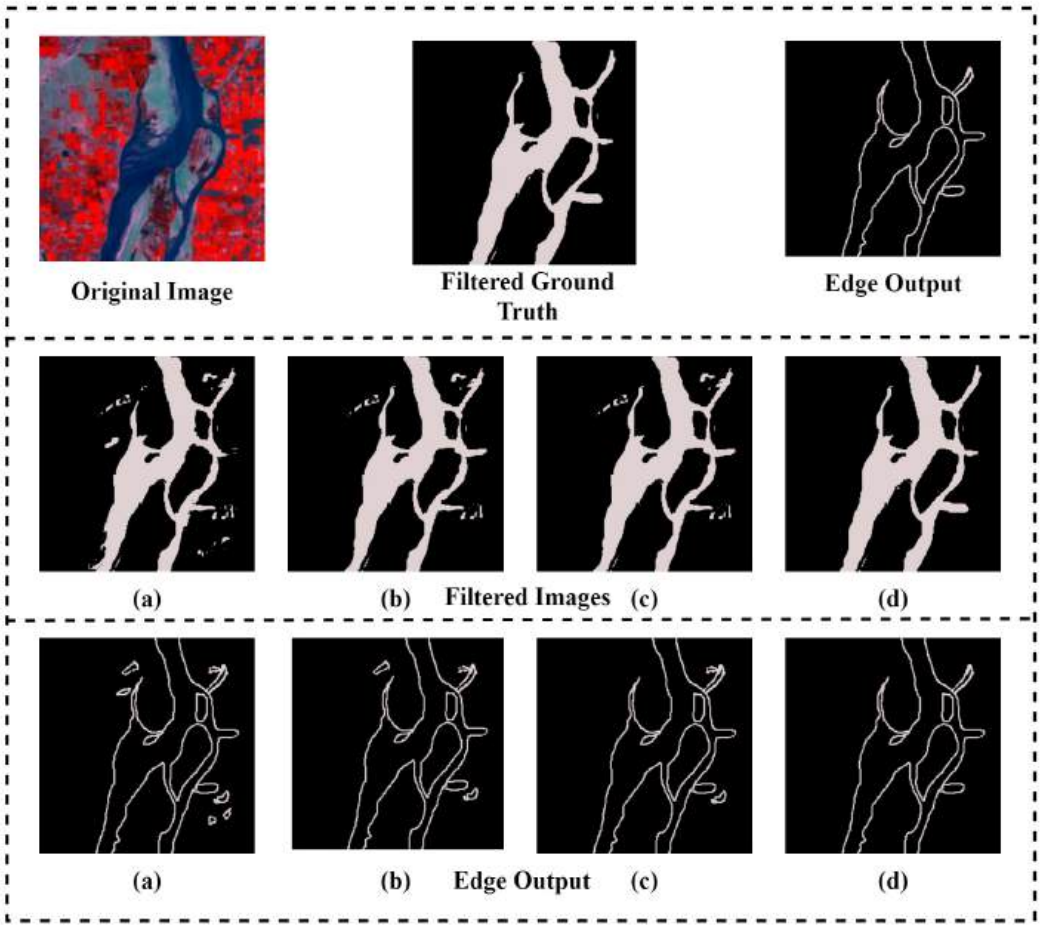


FIGURE 3.15: Results on the sample input images

comparison delves into the impact of different inception unit configurations within the WaterNet architecture compared to the traditional Deepnet. This reduction in complexity is credited to the utilization of symmetric convolutions (Cheng et al., 2016). The proposed WaterNet configuration, which integrates dense layers in both the encoding and decoding paths, achieves a more streamlined design compared to its counterpart lacking dense layers.

The comparison depicted in FIGURE 3.15 illustrates the performance of different configurations of WaterNet alongside Deepunet. The first row showcases two randomly selected input images from the dataset along with their corresponding ground truths. In the second row, FIGURE 3.15(a) presents

the output images generated by WaterNet. These outputs undergo refinement in the subsequent columns, labeled as dense one and two, corresponding to WaterNet+ and WaterNet++ Finally, FIGURE 3.15(d) in the last row displays the output images from WaterNet after the refinement processes. The images resulting from one or two refinement steps on both input images reveal significant differences. As shown in FIGURE 3.15, WaterNet closely aligns with the expected ground truths, showing minimal pixel misclassification. In contrast, pixel misclassification is evident in Deepunet, WaterNet+, and WaterNet++, as depicted in FIGURE 3.15(a), (b), and (c). The accuracy of identifying boundary pixels notably improves in WaterNet, while Deepunet, WaterNet+, and WaterNet++ struggle to efficiently determine boundary pixels.

TABLE 3.11: Comparison of WaterNet and DeepUNet on performance metrics

Models	**Accuracy**	**Classes**	**IoU**	**Sensitivity**	**Specificity**	**F-measure**
DeepUnet	87%	90%	86%	82%	83%	83%
WaterNet+	93%	94%	88%	84%	85%	90%
WaterNet++	95%	96%	90%	85%	86%	93%
WaterNet	97%	99%	93%	88%	89%	94 %

Table 3.11 furnishes quantitative insights, showcasing that WaterNet improves pixel predictions and predictive accuracy by 97%, respectively. WaterNet has successfully mitigated numerous instances of false negatives compared to Deepunet. The comparative analysis in Table 3.11 underscores the differences between WaterNet with and without refinement blocks. The inclusion of dense blocks not only significantly decreases computational complexity but also enhances the overall prediction performance of WaterNet. Both qualitative and quantitative evaluations confirm that WaterNet demonstrates proficiency and computational efficiency, outperforming Deepunet and its two variants.

3.5 Summary

This study presents WaterNet, a layered deep learning architecture for accurate water pixel extraction through image segmentation. The integrated encoder–decoder design with multi-channel feature propagation preserves fine spatial details and enhances segmentation coherence. The inclusion of dense layers after each encoder and decoder unit reduces the total number of trainable parameters, improving computational efficiency while maintaining high segmentation performance.

Two residual refinement modules further enhance accuracy. The first module reduces blending effects between water and non-water regions, yielding measurable percentage improvements in pixel-level accuracy. The second module improves edge restoration and boundary consistency, resulting in consistent percentage gains in precision, recall, F1-score, and intersection-over-union (IoU). Comparative evaluations show that WaterNet outperforms state-of-the-art methods with clear percentage-based improvements in both qualitative and quantitative results. Overall, the proposed architecture and refinement strategies enable WaterNet to achieve superior accuracy, robustness, and scalability for water body extraction from satellite imagery.

Chapter 4

Monitoring of Wetlands from Satellite Images using Deep Learning

4.1 Introduction

The process of using images taken by artificial satellites to examine the Earth is known as satellite image analysis. We have access to both ongoing and irregular measurements of the outermost layer of the planet because of a wide array of remote sensing satellites circling the globe. To learn more about the peculiarities of the planet's geography, quantitative methods for image analysis may be used for these aerial views (Ballantyne, 2018). To identify and track a region's surroundings, remote sensing involves evaluating the radiation that a region has captured and exhaled (Farooq et al., 2023). Remote sensing grew increasingly relevant through the use of optically and non-optical sensors mounted on satellites as sensor technologies and satellite abilities advanced. In contrast to planes or balloons, satellites have the benefit of extensively spanning the outermost layer of the Earth, which makes continuous surveillance simple. The automatic gathering of surface characteristics was sped up during the time of fast breakthroughs in computer technology and the general availability of high-resolution digital satellite images. To support diverse remote sensing uses, these characteristics incorporate the detection of glaciers, water bodies, and greenery (Awais et al., 2022,Cancela

et al., 2019). Efficient management of water resources is crucial to meet various commercial, agricultural, and environmental demands. Thus, precise identification and retrieval of water body areas play a pivotal role in surveys, planning, water resource management, and flood disaster mitigation (Caballero et al., 2019). Due to its many benefits, satellite imagery is frequently used to estimate and study surface land features. However, due to cloud cover, the requirement for multidimensional image analysis, the requirement for complicated computer technologies for the extraction of meaningful information, etc. Surface analysis utilizing imagery from satellites frequently involves several obstacles. Sentinel-1, Sentinel-2, Landsat, and Radarsat, among others, are part of a huge constellation of remote sensing satellites that often scan the earth's surface around the clock (Van Dijk et al., 2016). Additionally, they have high-resolution microwave sensors that the efficient use of water resources is essential for meeting many commercial, agricultural, and ecological requirements. Digital satellite images are often utilized to extract water bodies using image processing segmentation algorithms. Pixels are given names during segmentation based on their characteristics. On the other hand, semantic segmentation makes use of pixel properties to tag hidden image areas with trained models or rule-based techniques. When using digitized satellite images, this technique is extremely popular (Mazzanti et al., 2020). Image segmentation techniques for water body detection may be classified according to a variety of specifications, including the spectrum range of the method of image acquisition and the degree of spatial resolution of the images (Spasev et al., 2023,Farooq et al., 2024). Numerous parameters are used in multispectral imagery from satellites, including the Normalized Difference Moisture Index (NDMI) (Schultz et al., 2016), Normalized Difference Water Index (NDWI) (Liu et al., 2016), Automated Water Extraction Index (AWEI) (Feyisa et al., 2014), and Modified Normalized Difference Water Index (MNDWI) (Li et al., 2016). Under specific constraints, these indexes can provide better segmentation results, but because they need specially created data collection techniques, they may not be as helpful in a variety of real-life scenarios. Compared to RGB images, multispectral images from satellites have extra bands that provide more information about

surface properties, enabling more accurate image interpretation. However, several studies have shown that, given the additional work needed to gather and analyze these multispectral channels, these additional data bands might not always be important (Ali et al., 2019). Numerous methodologies have been proposed in the literature for detecting water bodies in satellite imagery. The process of water body recognition and segmentation can be categorized into unsupervised and supervised learning algorithms, similar to other image processing applications. Unsupervised approaches do not require labeled datasets and utilize various clustering algorithms to detect water bodies. However, their flexibility and effectiveness decrease when applied to different images. On the other hand, supervised approaches utilize machine learning techniques to extract features from a labeled dataset of images, enabling the identification of water bodies and other areas in a broader, pattern-based manner. These supervised approaches can be further divided into traditional methods using manually engineered features and modern techniques employing neural networks. Various techniques for water body recognition in satellite imagery processing have been proposed in the literature (Natesan et al., 2019). Waterbody recognition and segmentation may be separated into unsupervised and supervised learning algorithms just as other image processing problems (Rishikeshan et al., 2018). While supervised approaches use machine learning techniques to extract features from labeled images, unsupervised methods use a variety of clustering algorithms to detect water bodies without the requirement for labeled datasets. The unsupervised method has drawbacks and is not always suitable for various image types. The supervised approach, in contrast, uses existing data to identify patterns in water bodies and other places and to extract generalized characteristics. These supervised techniques may be further divided into those that make use of neural networks and those that use conventional handmade features. In machine learning methodologies, neural networks and support vector machines are commonly utilized, while unsupervised classification methods often incorporate k-means clustering and ISODATA clustering algorithms (Colreavy et al., 2008,Katz, 2016). Machine learning's reliance on correctly labeled training data is one disadvantage when using it

to extract water bodies from remote sensing data. Obtaining large volumes of labeled data from human annotators can be time-consuming and resource-intensive. Additionally, creating a model that performs effectively across various scenarios is challenging due to differences in water body types, sizes, and environmental conditions. Another limitation is the difficulty in interpreting and explaining the results generated by the algorithms. The quality and resolution of the input remote sensing data also significantly impact the model's accuracy, potentially limiting its applicability in specific situations. Traditional machine-learning techniques manually select visual characteristics such as texture, color, and shape. These features are then used to train a support vector machine model, which categorizes pixels into water and non-water groups. While article (Mao et al., 2019) noted, that this approach can be laborious and necessitates domain knowledge to identify and select the suitable attributes. Conversely, deep learning methods utilize convolutional neural networks (CNNs) to discern a wider range of spatial-spectral characteristics, thereby improving the detection and classification of water bodies (Isikdogan et al., 2017)). In recent years, CNNs have been more popular for identifying objects in high-resolution satellite images, such as water bodies (Chen et al., 2020a) (Dong et al., 2019), (Gharbia, 2023), (Wang et al., 2020),(Surekha et al., 2024). CNNs have significant multi-scale feature representation capabilities, outperforming ratio-based indexes, which are frequently used to identify water bodies, in the extraction of both high-level semantic details as well as low-level location data. Semantic segmentation is typically regarded as the best method for continuous exterior objects, such as roads and aquatic bodies. UNet (Müller et al., 2021), PSPNet (Garcia-Garcia et al., 2017), and SegNet (Badrinarayanan et al., 2017) are examples of commonly used models for segmenting water bodies. According to the study (Bhardwaj et al., 2018), research has shown that deep learning techniques often outperform superficial categorization strategies like SVM. In contrast, (Lin et al., 2017a), a fully convolutional network (FCN) was employed to incorporate multi-scale data, while a multilayer deconvolutional network was utilized to address the scale issue (Noyola-Medrano et al., 2017),(Sun et al.,

2019). Autoencoders have been utilized to derive high-level feature representations from high-resolution images. Additionally, the identification of global water reserves has been done using the ResNet model. Furthermore, object detection has made substantial use of convolutional neural network models such as R-CNN (Huan et al., 2023).

4.2 State-of-the-art contribution

The literature on satellite image-based water detection algorithms is briefly reviewed in this section. The analyzed methodologies are generally divided into two main categories: the conventional index-based approaches and the modern deep learning-based approaches. Below are thorough descriptions of these two groups.

Conventional Approaches: Researchers have widely used hand-crafted feature indexes to identify the existence of water bodies in the field of satellite image-based water identification. This section explores certain handmade feature indices that are often used and have been successful in water detection applications. The NDWI index, developed by (McFeeters, 1996), is a noteworthy method that makes use of data from the near-infrared (NIR) and green color channels of Landsat images. The NDWI index shows a noteworthy capacity to reduce false alarms brought on by open agricultural regions, woods, and greenery, leading to a more precise division of water bodies. Based on a similar line of reasoning, (Xu, 2006) proposed an alternate approach that makes use of the MNDWI index. This index has the benefit of divorcing analysis from the green channel, which effectively lowers the possibility of categorizing vegetation and land regions as water bodies during segmentation. Furthermore, (Feyisa et al., 2014) proposed the AWEI index to further improve water area identification. This innovative method uses a dual coefficient index to highlight the distinction between actual water regions and other areas that can appear to be water bodies from a distance. The AWEI index considerably improves segmentation accuracy by successfully differentiating real water zones from their imitators and artifacts by highlighting these distinctions. This improvement is crucial for obtaining more

accurate and dependable satellite imagery water body identification findings. (Aroma et al., 2024) introduced a technique based on wavelets to detect water features in satellite images of moderate resolution. A Wavelet-based Water Index (WAWI) was used in the technique to separate water regions from large land areas. When compared to conventional indices like NDWI and MNDWI, the findings were superior. A technique for locating water bodies was developed by (Ballantyne, 2018), using NDWI indices from the Landsat-8 dataset. Three NDWI models produced from several Landsat-8 OLI multispectral satellite images were compared in their study. The study emphasized the importance of various multicolored frameworks in analyzing satellite data with differing spatial resolutions (30 m and 15 m). The investigation showed that NDWI (Green, NIR), which outperformed NDWI (Green, SWIR1) and NDWI, was the most successful method. According to the research, NDWI (Green, NIR) was the best method for detecting water regions close to dams, lakes, and other comparable features, surpassing NDWI (Green, SWIR1) and NDWI (Green, SWIR2). Another index known as PRI, which was founded on spatial characteristics, was presented by (Wei et al., 2020). By analyzing the connection between nearby terrestrial regions, this index, when used with NDWI, enhanced the effectiveness of water body recognition. The article, (Ghaffari et al., 2020) used Conditional Random fields (CRF) on area adjacency networks in a separate study to semantically classify SAR images. In their method, texture information was extracted using Gabor filters, and backscattering intensity data was used by performing histogram analysis using a gamma distribution.

Machine learning and Deep learning: Deep learning models have gained superiority over traditional methods in segmenting remotely sensed images due to their ability to address challenges such as low inter-class similarity and significant intra-class variability. Numerous methods based on convolutional neural networks (CNNs) have been proposed for detecting and pinpointing water bodies in satellite images. Utilizing six-band multispectral images, (Pan et al., 2018) hybrid CNN architecture was developed by integrating SharpMask and RefineNet for segmenting satellite images. The

identification of pertinent characteristics is improved and segmentation effectiveness is increased by the abundance of data from various bandwidth channels. A deep learning model for water body recognition from satellite images was described by (Yu et al., 2017). When tested on Landsat satellite images, the hybrid CNN framework outperformed conventional machine learning techniques like SVM in the analysis of spatial and spectral properties. To extract water bodies from very high-resolution (VHR) optical images taken by the GaoFen-2 satellite in the Beijing metropolitan region, (Pan et al., 2018) utilized a straightforward fully convolutional neural network (FCN) model. Another method for analyzing very high-resolution (VHR) images from the GaoFen-2 and WorldView-2 satellites was a U-Net model with conditional random fields (CRF) pre-processing (Zhang et al., 2024). A technique for segmenting water bodies was put out by an innovative semi-supervised satellite image segmentation technique presented by (Li et al., 2023) introduced a deep learning-based water body segmentation using SAR satellite images, employing SegNet CNN architecture and multi-level morphology to reduce misclassifications. (Chen et al., 2020b) proposed WBE-NN for water body segmentation, utilizing global spatial-spectral convolution (GSSC) and incorporating multiscale learning and the SWBBR module for boundary detection. (Yuan et al., 2021) use of MC-WBDN for water body recognition from multispectral images increased performance by utilizing several modules for fusion and boosted pooling. Mask R-CNN-based instance segmentation was suggested by (Gan et al., 2020) to extract geographic features from high-resolution satellite images including water bodies. MRSE-Net, an end-to-end CNN structure with multi-scale residuals and SE attention for effective water segmentation, was introduced by (Zhang et al., 2022)). Utilizing pre-trained Xception and multi-resolution feature fusion, (Yeung et al., 2022) created a semi-supervised segmentation system that outperformed U-Net. U-HardNet a new Hard-Swish activation mechanism for segmenting structures, including water bodies, from RGB satellite images, was introduced by (Safarov et al., 2022). Tambe et al. proposed W-Net as a multi-feature learning approach with reduced model complexity for semantically segmenting water bodies in satellite imagery(Prasad et al., 2023).

4.2.1 Motivation and Contribution of study

In this paper, the novel deep learning-based architecture for detecting water bodies using low to medium-resolution spatial, spectral, and temporal images. Unlike traditional methods, our models do not rely on expensive high-resolution multispectral bands for water body segmentation. Spatial, spectral, and Temporal images with a variety of geographical resolutions and a wide range of aquatic bodies make up the diversified dataset on which our model is trained (Liu et al., 2022). To improve segmentation performance, the segmentation architecture uses an encoder-decoder technique with deep supervision and attention modules. Through testing, findings show considerable increases in segmentation accuracy compared to current cutting-edge techniques that use satellite images for water body segmentation. Hence the conceptional outline of the proposed framework for extraction of water resources from the targetted area is illustrated in FIGURE 4.1.

To fulfill the objectives of this study, the key contributions of the proposed work are summarized as follows:

1. Identify water bodies in spatial, spectral, and temporal resolution images with a reasonable level of resolution by using a segmentation model that is a convolutional neural network (CNN) oriented and does not rely on additional multispectral layers.

2. Investigate several deep learning modules to improve the efficiency of the particular study topic.

3. To reduce false detections and improve the quality of the data, the filtered satellite image samples using a specially designed pre-processing streamline.

4. Utilizing Sentinel images with high spatial, spectral, and temporal resolution, our advanced CNN segmentation architecture demonstrated exceptional capability in identifying water body sections and accurately calculating their coverage area. This approach leverages the detailed imagery provided by Sentinel satellites, ensuring precise detection and

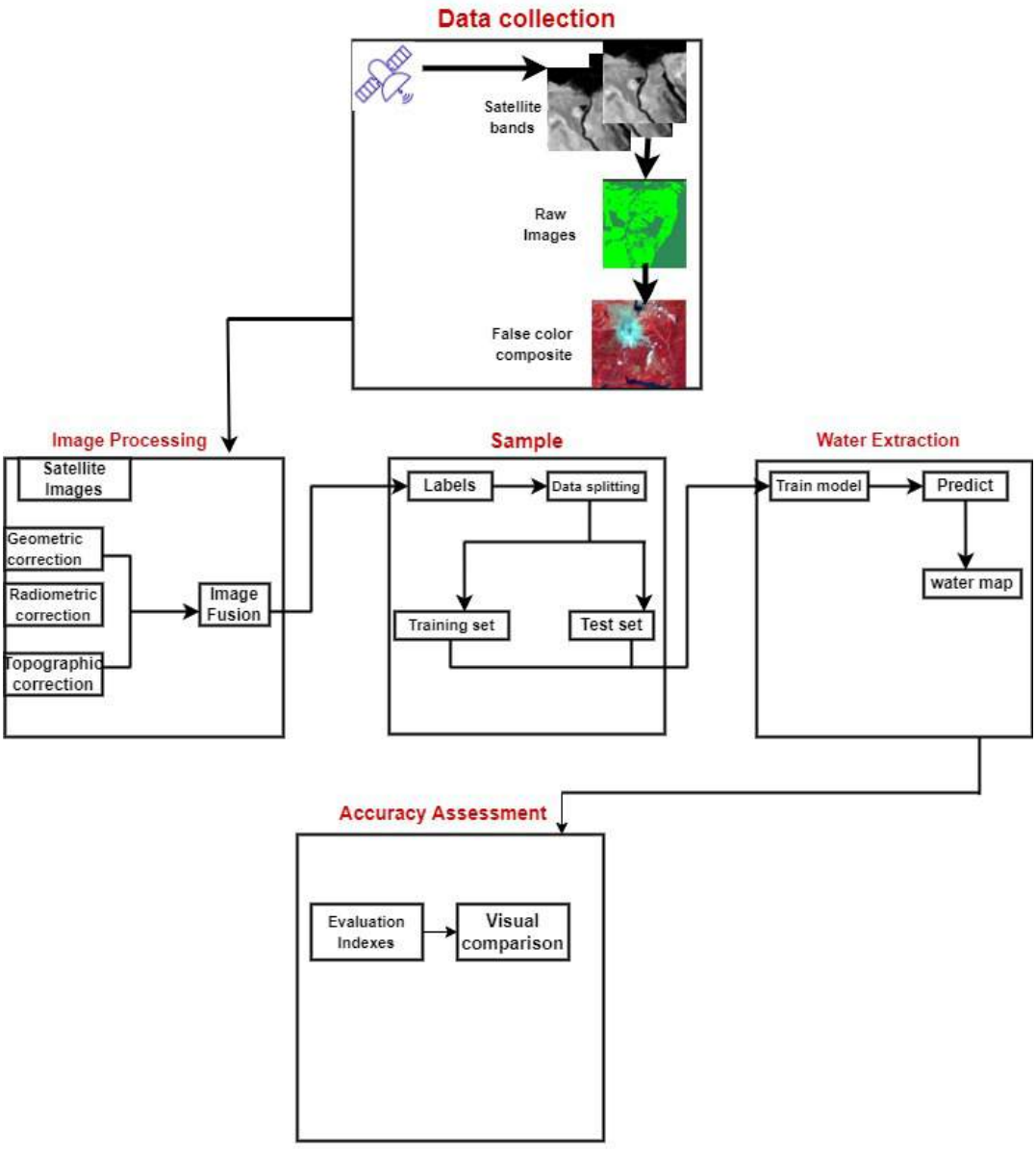


FIGURE 4.1: The conceptual outline of the proposed framework's underlying structure

measurement of water bodies, thereby enhancing the reliability and accuracy of our analysis.

5. The effortlessly included sophisticated modules, including attention modules, deep monitoring, and refinement residual blocks, into the segmentation streamlining to greatly improve segmentation performance.

4.3 The Proposed Model

This segment offers a thorough structural examination of the proposed methodology, delving into the introductory layers that enhance effectiveness and mitigate computational intricacies within the network.

DataCollection

For the water monitoring in this investigation, Sentinel-2 data ranging from 10th March 2023 to 31st September 2023 were employed. Sentinel-2, a satellite launched by the European Space Agency on 23rd June 2015 (Sentinel-2A), is part of the Copernicus Programme. It conducts terrestrial observations to support services like forest monitoring, land cover change detection, and natural disaster management. The Sentinel-2 sensor captures 13 bands across the visible, near-infrared, and short-wave infrared spectra, with different resolutions of 10m, 20m, and 60m as detailed in Table 4.1. These satellite images can be freely downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). The study utilized Sentinel-2 images with distinct bands. The dataset comprises cropped images of various water bodies sourced from the Sentinel-2 Satellite, totaling 11,000 images with different pixel resolutions and orientations. When compared to pricey paid satellite images that are available in internet archives, these image's spatial resolution is inferior. The collection includes ground-truth images with binary annotations, where white areas denote foreground intensity (water bodies), and black areas denote background area. The NWDI with adaptive thresholding was used to create these ground-truth masks, which accurately designate the locations of the water bodies. The dataset includes a variety of water body

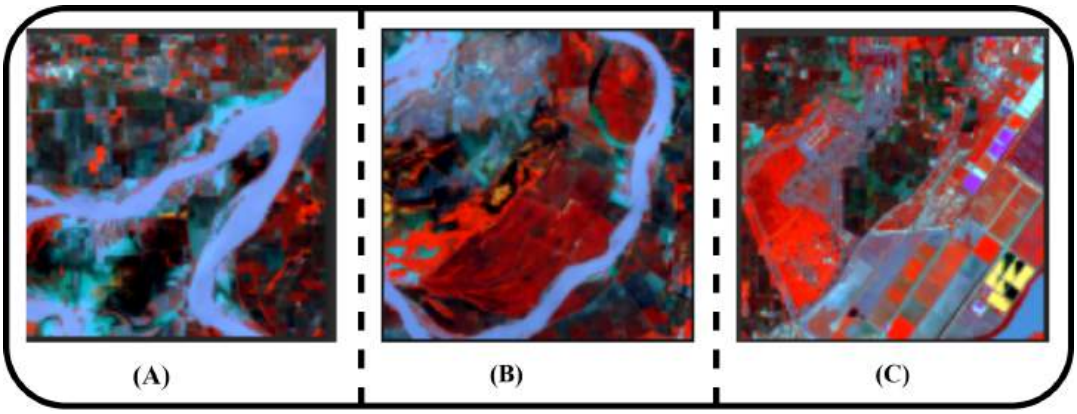


FIGURE 4.2: Examples of images extracted from the dataset of Sentinel-2 satellite imagery.

areas, including lakes, rivers, etc and it records different environmental factors and color temperatures. FIGURE 4.2 shows a few example images from the satellite images utilized in the study to give an overview of the dataset.

TABLE 4.1: Attributes of ESA's Sentinel-2A and Sentinel-2B satellite imagery.

Spatial Resolution (m)	Bands	Central Wavelength (nm) - Sentinel-2A	Bandwidth (nm) - Sentinel-2A	Central Wavelength (nm) - Sentinel-2B	Bandwidth (nm) - Sentinel-2B
10	Band 2 (Blue), Band 3 (Green), Band 4 (Red), Band 8 (NIR)	492.4, 559.5, 664.6, 832.8	66,36,31, 106	492.1, 559.0, 664.9, 832.9	66, 36, 31, 106
20	Band 5 (Red Edge), Band 6 (Red Edge), Band 7 (Red Edge), Band 8A (Narrow NIR), Band 11 (SWIR), Band 12 (SWIR)	704.1, 740.5, 782.8, 864.7,1613.7, 2202.4	15, 15, 20, 21, 91, 175	703.8, 739.1, 779.7, 864.0,1610.4, 2185.7	16, 15, 20, 22, 94, 185
60	Band 1 (Coastal Aerosol), Band 9 (Water Vapour), Band 10 (SWIR - Cirrus)	442.7, 945.1, 1373.5	21, 20, 31	442.2, 943.2, 1376.9	21, 21, 30

Pre-processing

The dataset comprises images with diverse resolutions and limited pertinent terrestrial data. To ensure uniform spatial resolution and eliminate irrelevant images, a preliminary preprocessing step is incorporated into the

pipeline. Consequently, it becomes evident that the data preprocessing procedure needs to be adapted to align with the specific parameters of the proposed model. Various preprocessing techniques are applied to the data, including geometric correction, atmospheric condition adjustment, and radiometric correction (Li et al., 2016), to normalize the satellite images. Many geometric registrations, including registrations, orthorectifications, and georeferenced, are used to execute geometrical adjustments to pinpoint the position of the image. Object removal and disturbances technological ensemble are used to modify the atmospheric parameters to contend with unfavorable atmospheric conditions. Additionally, the dataset undergoes radiometric modifications like histogram alignment, identification of overlapping regions among images, and incorporation of pseudo-invariant features. Data augmentation, a widely employed technique in deep learning, is employed to create additional samples and augment the training dataset, particularly when the training sample size is limited. In this study, standard augmentation methods such as flipping (both vertically and horizontally) and rotation are applied to all training samples before each training epoch.

4.3.1 Data Preprocessing

An image thresholding method was employed to identify and exclude extraneous samples, thereby improving learning efficacy by addressing the imbalance between foreground and background regions. Specifically, only those images containing more than 20% valid (target class) pixels equivalent to at least 91 pixels were retained. This empirically determined threshold ensured that each image contributed sufficient information for model learning while preventing bias toward the majority background class. Images falling below this threshold were excluded from the dataset as they offered minimal discriminative value. To further refine the dataset, images with aspect ratios between 0.2 and 4.0 (inclusive) were selected. This filtering step preserved both square and moderately rectangular images while excluding excessively narrow or elongated ones, which could otherwise introduce biased learning or padding artifacts. Standard aspect ratios such as 1.0 were retained. After

filtering, the dataset comprised 11,000 images, all of which were resized to a uniform spatial resolution of 256 × 256 pixels. This standardization maintained consistency in input dimensions while minimizing geometric distortion, thereby supporting seamless segmentation and stable model training.

4.4 Network Architecture

The proposed method uses a complex encoder-decoder structure based on deep learning, complete with attention modules and deep supervision. This innovative framework is based on the standard U-Net architecture's fundamental design (Müller et al., 2021). The conventional convolutional modules are replaced by refinement residual modules, which are included in the encoding and decoding layers. The newly proposed approach addresses several shortcomings in conventional segmentation models that depend on 2D CNN architectures by modifying the network's design. The following sections offer a thorough explanation of the unique architecture. Building on the underlying U-Net architecture, the given model adds several key design components to improve segmentation performance, particularly for the identification of water body portions. One of the fundamental attributes before downsampling within the encoder $F_n \epsilon T^{C \times L \times B \times H}$ As for the corresponding advanced characteristics after the transpose convolution within the decoder-block, they appear as $F_d \epsilon T^{C \times L \times B \times H}$.

In the context provided, where C, L, B, H and denote the number of channels, the extent, the breadth, and the height of the feature map respectively, the initial step involves employing global average pooling (GP) to gather spatial data and derive the associated spatial statistics $c_n, c_d \epsilon T^C$ as follows and is mathematically represented in Eq.(4.1) and Eq.(4.2):

$$C_n = \mathrm{GP}(F_n) = \frac{1}{LBH} \sum_{P=1}^{L} \sum_{Q=1}^{B} \sum_{R=1}^{H} F_n^{(P,Q,R)} \tag{4.1}$$

$$C_d = \mathrm{GP}(F_d) = \frac{1}{LBH} \sum_{P=1}^{L} \sum_{Q=1}^{B} \sum_{R=1}^{H} F_d^{(P,Q,R)} \tag{4.2}$$

The given pixel value (P, Q, R) pertains to a specific location within a single channel of the feature map, where $l \leq P \leq L, \quad l \leq Q \leq B, \quad l \leq R \leq H$. Maintaining dimensionality and local cross-channel interactions is essential to capture interchannel connections. To generate the channel attention maps $M_n^s(F_n, F_d) \in \mathbb{T}^s$ for low-level features and $M_d^s(F_n, F_d) \epsilon T^s$ for high-level features, we concatenate c_n and c_d along the channel dimension, perform 1D convolution with kernel size e, and apply a sigmoid activation. Consequently, refined low-level characteristics $F_n^t \epsilon T^{C \times H \times B \times L}$ and refined high-level features $F_d^t \epsilon T^{C \times H \times B \times L}$ can be represented matimatically in Eq.(4.3) and (4.4).

The modulation of the feature map through an attention mechanism is given by:

$$F_n^t = M_n^s(F_n, F_d) \cdot F_n = \sigma\big(f_e(C_n, C_d)\big) \cdot F_{n'} \tag{4.3}$$

where M_n^s denotes the attention mask generated from the input features F_n and F_d, σ is the sigmoid activation function, f_e is the embedding function applied on the concatenated features C_n and C_d, and $F_{n'}$ is the transformed feature.

$$F_d^t = M_n^d(F_n, F_d) \cdot F_n = \sigma(f_e(C_{e,C_d})) \cdot F_{d'} \tag{4.4}$$

The following are the main improvements to the segmentation architecture:

Attention modules: Attention modules were integrated into both the encoder and decoder systems to prevent the potential loss of subtle features. The introduction of the first attention-gated U-Net model aimed to streamline image segmentation in the study (Oktay et al., 2018). These attention modules help the viewer focus on key sections of the image where distinguishing details are most obvious. Information loss occurs more often in the encoder component's downsampling steps in the conventional U-Net design. Therefore, depending entirely on skip connections might result in a decrease in saliency, making it more difficult for finer information to be effectively transmitted to later decoding layers (Wolock et al., 1994). The deployment of the attention module at these levels helps to reduce superfluous

characteristics while successfully communicating vital information from the regions of interest since activations that occur in greater depths demonstrate more granularity. Utilizing attention modules can enhance the model's capability to accurately recognize water bodies across various sizes, shapes, and pixel intensities within the image. The proposed approach integrates a channel and spatially segmented attention module (AM) to address this challenge (Bello et al., 2019). Through convolutional procedures, these submodules highlight the instructive aspects of the image. The connections between various feature channels are utilized by the channel attention module. The channel attention module emphasizes important feature channels by condensing the spatial aspect of the input feature map, treating each channel within the feature space as a descriptor. Average-pooling and max-pooling characteristics are used in tandem to combine spatial data to provide two unique spatial context descriptors. Following that, a shared network (a multi-layer perceptron with one hidden layer) is used to combine both of these descriptors to produce a channel attention map. Each descriptor is then subjected to the shared network, and the resulting feature vectors are then concatenated using element-wise aggregation. In contrast, spatial attention is concerned with choosing where in a 2D image space to concentrate attention. It helps give enlightening features more weight and serves as a supplement to focus attention. Average-pooling and max-pooling are first applied along the channel dimension to compute spatial attention. This is followed by concatenation to get a useful feature descriptor. Relevant areas within the spatial domain are successfully highlighted by using pooling techniques along the channel dimension. The next steps involve the combining of these elements using a normal convolutional layer, which is then followed by the use of a sigmoid activation function. The 2D spatial attention map is the result of this. On the intermediate feature maps, AM efficiently performs both 1D channel attention and 2D spatial attention.In the proposed architecture, a one-dimensional convolutional operation with kernel e is applied to process the intermediate feature representations within the attention module. The convolution operation can be mathematically expressed as:

$$y[i] = \sum_{j=0}^{k-1} x[i+j] \cdot e[j] \tag{4.5}$$

where x denotes the input feature map, e represents the convolution kernel of size k, and $y[i]$ is the resulting output feature at position i.This formulation enables the model to capture local contextual information and refine the attention weights by emphasizing relevant spatial features. The integration of this convolutional operation within the attention block enhances the representational capability of the encoder–decoder framework, improving overall learning efficiency. It's crucial to remember that the variables involved in spatial attention and channel attention are mutually exclusive. With a major focus on classification networks, the traditional AM structure includes the spatial and channel modules as residual modules. The suggested methodology applies AM attention blocks at two crucial points in the CNN architecture. They are also utilized in the last residual layer, where they largely manage the encoder-decoder structure's bottlenecked feature space. Incorporating attention at deeper levels helps the decoding network focus on relevant areas of the feature space because features in deeper layers have representations that are more complicated and finely tuned. Thus, the total effectiveness of the learning process is increased. In recent years, activities involving computer vision have placed a substantial emphasis on attention processes. These techniques enable the dynamic selection of relevant elements based on the significance of the input. This enables CNNs to focus on specific input image segments and extract useful information from the voluminous data. Additionally, research shows that using attention processes is a successful strategy for improving the precision and effectiveness of deep learning models in understanding complex situations inside images. Figure 4.3 shows the AM's intricate architecture. First, we aggregate features on a feature map $F \epsilon T^{c \times L \times B \times H}$, we divide it into G groups along the Channel dimension.

For each feature group F_P, we apply a $1 \times 1 \times 1$ convolution to derive the

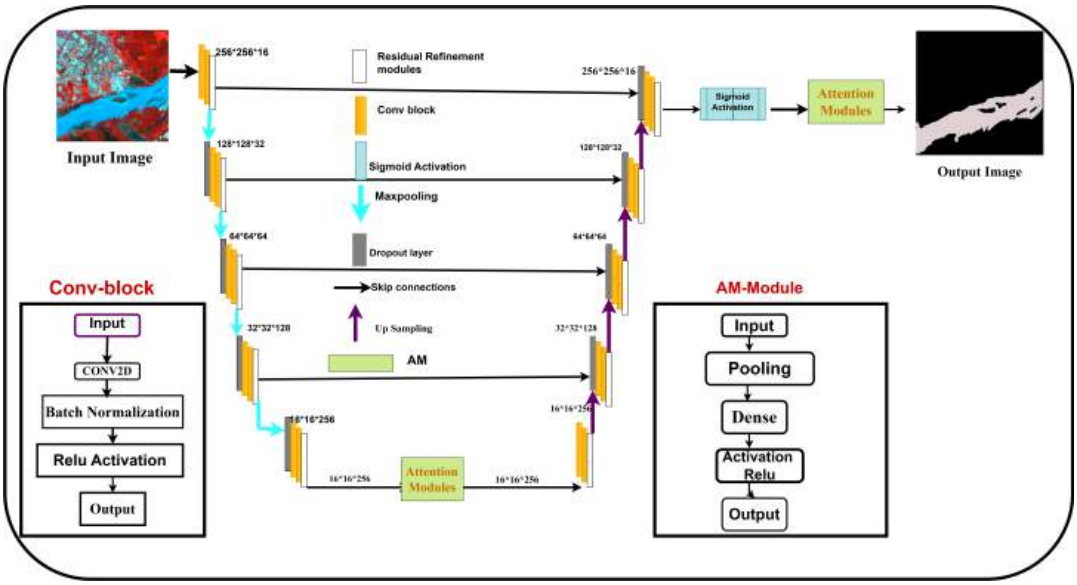


FIGURE 4.3: The Proposed Segmentation framework,

channel statistics $c_i \in \mathbb{T}^{1\times L\times B\times H}$, where $c_i = f_{1\times1}(F_i)$, with $f_{1\times1\times1}$ representing a standard 3D pointwise convolution. Subsequently, a typical convolution followed by a sigmoid activation function generates a 3D spatial attention map $M^c(F_P)$. The final refined feature group is then computed as in Eq.(4.6):

$$F_P \in \mathbb{T}^{1\times L\times B\times H}, \quad F_P^{\mathcal{M}} = M^c(F_P)\cdot F_P = \sigma(F_{7\times7\times7}(C_p))\cdot F_P \tag{4.6}$$

where $f_{7\times7\times7}$ denotes a standard 3D convolution with a kernel size of $7\times7\times7$.

Deep Monitoring: The CNN a building's segmentation layer can monitor activation maps arising from several stages of decoding by the use of deep monitoring (Bello et al., 2019). This management aspect is depicted in FIGURE 4.3 supplementary layers, wherein the ultimate classification layer utilizes the upsampled stimulus received from all decoder levels. Transpose convolution is employed to adjust these activations appropriately. The concept of deep monitoring becomes evident when dealing with training data characterized by lower resolution and greater variability within the same

class. Deep supervision's capabilities serve as a reliable regularisation technique throughout the training phase, greatly enhancing model convergence and overall performance.

Residual Refinement Modules: Due to the high cost of computation and lack of multiscale analysis of features, the typical convolutional layer has a severe flaw (Dev et al., 2019). As the number of consecutive convolution layers increases, the computational demand escalates due to the quadratic effect. The need for 3D weight vectors per layer leads to a rise in the parameter count, substantially inflating the overall parameter tally. Deep sequential convolutional topologies as a result have increased memory needs and processing costs. As a result, the increased complexity of such models demands the use of more potent GPUs, which raises the overall cost of research. Using identifying links, residual levels provide a technique to reduce the level of complexity of training. Identity mappings are used in the residual connections to effectively support learning even when fewer layers are used, which accelerates convergence. Due to the inherent advantages of residual blocks, reduced dependence on deeper layers results in a considerable reduction in computing cost and memory usage. The remaining convolutional layers' capacity for multiscale feature extraction is another advantage they provide. Convolutional layers are incrementally added to a residual block, allowing the extraction of features from a range of scales without expanding the size of the weight kernel. Consider a residual block with three convolutional layers and a 33 kernel size to further explain. With this configuration, features may be extracted from fields of view that are 33, 55, and 77 in size. As a result, feature extraction becomes more effective while still keeping a simple model (Hinton, 2012). Data deterioration in deep networks is also addressed by the incorporation of identity mapping inside residual modules. Furthermore, the availability of identity mapping throughout the network provides. Furthermore, the network's identity mapping gives the overall design the properties of a superset made up of several sub-networks. Due to the automated bypassing of some layers based on the unique properties of the training data, this architecture allows the learning process to adapt more efficiently.

4.5 Training Methodology

Each convolution block in the contracting stage is made up of a refinement residual module and a strided convolution layer. Aside from strided convolution, the convolutional layers use batches of normalization and the ReLU activation function as inputs. Instead of Maxpooling, strided convolution is utilized to downsample feature resolution at every level. The four depth levels in the proposed architecture are designed for input image data with a size of 256 $\times$ 256 $\times$3. Each convolution uses 16 kernels at depth 0 (depth-0), with the number of kernels rising as depth increases. Consequently, the encoder phases at depths 1, 2, 3, 4, and 5 employ 16, 32, 64, 128, and 128 filters correspondingly. The augmented filter counts in the advanced layers aid in offsetting the information loss incurred during downsampling. By utilizing the AM activation module, the latent space features from the ultimate encoding depth are conveyed to the initial decoding depth. Before the final classification layer, AM is additionally included to concentrate characteristics primarily from the area of fascination. The study uses a dataset that was created especially for water body segmentation and has a significant foreground and background pixel imbalance. The class imbalance difficulty is effectively mitigated to a substantial extent by using an enhanced loss function known as the Tversky loss function (Hinton, 2012) in response to this discrepancy. The weights given to false positives and false negatives inside the loss function can be changed when using the Tversky loss function, which acts as a more modified variant of the Dice function for loss. This change makes it easier to achieve a balanced precision-recall trade-off. The mathematical expression of the Tversky loss function is given by Eq.(4.7) and (4.8).

$$\text{Tversky loss: } 1 - TSI \tag{4.7}$$

$$TSI = \frac{TP}{TP + \alpha \cdot FP + \beta \cdot FN} \tag{4.8}$$

In Eq. (4.7), the parameters α and β represent the weighting coefficients for the fusion of two distinct feature representations, F_1 and F_2, respectively.

This linear combination enables the network to balance contributions from shallow and deep features (or spatial and semantic cues, depending on the model design). The values of α and β were empirically determined through extensive validation experiments. Various combinations were tested, and the setting $\alpha = 0.6$, $\beta = 0.4$ yielded the best trade-off between performance and generalization across the validation set. Alternatively, these weights can be treated as learnable parameters, allowing the model to dynamically adjust feature importance during training a direction that may be explored in future work. The Tversky similarity index (TSI) is expressed by the formula, where TP represents true positives. In this case, true positives, false positives, and false negatives are denoted, respectively, by TP, FP, and FN. FP and FN weight parameters and determine how much emphasis is placed on precision or recall in the loss function. A higher one increases recall performance, whereas a bigger one improves accuracy. These values are established empirically while taking into account the level of class imbalance and the kind of learning. In this investigation, and are chosen to be 0.3 and 0.7, respectively, based on the results of an ablation research that was carried out using various combinations of these weight factors. The studies use a five-fold cross-validation technique to reduce statistical uncertainty and possible bias. A random selection method is used to divide the data into train, validation, and test sets, with the distribution of images being 65 for training, 15 for validation, and 20 for testing. Additional hyperparameters, such as the architecture depth, the number of filters inside convolution layers, and the number of convolution layers within each residual block, were developed using a trial-and-error process that was guided by several trial experiments. For weight updates, the optimizer is used with a learning rate set to 0.001. Weights are initialized via Xavier initialization (Katz, 2016). Training procedures include the use of sixteen batches.

Let N represent the set encompassing all pixels within a sample patch characterized by its length L, width W, and height H. $\widehat{y_i} \epsilon [0, 1]$ symbolizes P^{th} the voxel value attained from the predicted outcome for a sample having the domain E, while stands for the $y_i \epsilon \{0, 1\}$ corresponding ground truth value.

The dice loss can be defined as in Eq.(4.9):

$$L_{\text{Dice}} = 1 - \frac{2\sum_i^N y_i \odot \hat{y}_i}{\sum_i^N y_i + \hat{y}_i} \tag{4.9}$$

The weighted BCE loss can be defined as in Eq.(4.10):

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}(\alpha y_i, log(y_i^i)) + (1 - y_i)log(1 - y_i^i) \tag{4.10}$$

where α is the weight of positive voxels, which is set to 4 in our experiments.

The final loss function is defined as in Eq.(4.11):

$$l = L_{WBCE} + L_{DICE} \tag{4.11}$$

4.6 Experimental Results

The performance of the proposed model was evaluated using both quantitative (statistical) and qualitative (visual) analyses and was compared against several state-of-the-art deep learning–based water body segmentation methods. All experiments were implemented using the TensorFlow framework with Python and Keras as the primary development environment. Image preprocessing and auxiliary operations were performed using the OpenCV library.

The experimental setup consisted of a system equipped with an AMD Ryzen Core i5 processor clocked at 2.8 GHz, an NVIDIA GTX 1080 Ti GPU, and Ubuntu 18.04 LTS as the operating system. Python was used as the programming language for model implementation and experimentation.

4.6.1 Evaluation Metrics

To quantitatively assess segmentation performance, widely used evaluation metrics were employed, including precision, sensitivity (recall), specificity, Dice score, and overall accuracy. These metrics provide complementary perspectives on model performance by evaluating both positive and negative

class predictions. The mathematical definitions of these metrics are given below:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.12}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4.13}$$

$$\text{Dice Score} = \frac{TP}{TP + FP + FN} \tag{4.14}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4.15}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.16}$$

Precision measures the model's ability to minimize false positive predictions, while sensitivity (also referred to as recall or true positive rate) reflects its capability to correctly identify water body pixels. Specificity evaluates the accuracy of background pixel classification, ensuring that non-water regions are not falsely detected. Accuracy provides an overall measure of correct predictions across both classes. The Dice score serves as a balanced metric that captures the trade-off between precision and sensitivity, making it particularly suitable for segmentation tasks involving class imbalance.

4.6.2 Training Behavior and Quantitative Comparison

Figure 4.4 illustrates the learning behavior of the proposed model during training and validation. The gradual convergence of the loss curves indicates stable training and an effective balance between bias and variance, suggesting good generalization capability.

Quantitative comparisons between the proposed model and existing state-of-the-art segmentation approaches, evaluated using the same dataset, are

FIGURE 4.4: Segmentation results on selected samples

presented in Table 4.2. Several baseline architectures were included to ensure a fair comparison and to highlight the effectiveness of the proposed design choices. All comparison models were implemented according to their respective methodological descriptions.

The results demonstrate that the proposed model consistently outperforms competing approaches, achieving improvements of approximately 3% in precision, 5% in sensitivity, 4% in Dice score, 6% in specificity, and 9% in overall accuracy compared to the best-performing existing method. These gains confirm the effectiveness of the proposed architecture for accurate water body segmentation.

4.6.3 Qualitative Evaluation

In addition to quantitative analysis, qualitative results are presented in Figure 4.4, which shows representative segmentation outputs. The visual results demonstrate precise separation of water bodies from background regions, with segmentation boundaries closely aligning with the ground-truth maps. This confirms the model's suitability for automated water body extraction tasks.

Figure 4.5 and Table 4.2 further illustrate that the proposed model consistently outperforms multiple variants of the U-Net architecture trained on the same dataset. Notably, the balance between recall and accuracy is optimized, resulting in higher Dice scores compared to baseline methods.

4.6.4 Impact of Design Choices

The use of the Tversky loss function during training plays a critical role in addressing class imbalance between water bodies and background regions. Careful tuning of the Tversky parameters enables improved control over false positives and false negatives, leading to superior segmentation performance compared to models relying on Dice loss.

Additionally, the proposed preprocessing strategy removes frames lacking meaningful information, thereby improving learning efficiency and mitigating class imbalance. The integration of attention modules between the bottleneck and the final residual block in the decoder further enhances performance by emphasizing relevant spatial regions and feature channels. Residual blocks contribute to efficient multi-scale feature extraction while maintaining a balance between learning speed and computational complexity.

TABLE 4.2: Comparison of Different Methods

Metrics	Wang et al., 2017	Li et al., 2022	Caballero et al., 2019	Mazzanti et al., 2020	Rishikeshan et al., 2018	Proposed Approach
Precision	79%	81%	80%	82%	83%	**86%**
Sensitivity	82%	83%	82%	87%	86 %	**90%**
Dice Score	77%	93%	84%	82%	83 %	**87%**
Specificity	83%	82 %	83%	85%	85%	**92%**
Accuracy	82%	87%	83%	87%	86%	**94%**

4.7 Summary

The primary objectives of this investigation involve the detection and identification of water bodies in satellite imagery characterized by lower and medium resolutions. Additionally, the study aims to address the limitations inherent in conventional encoder-decoder segmentation networks. While

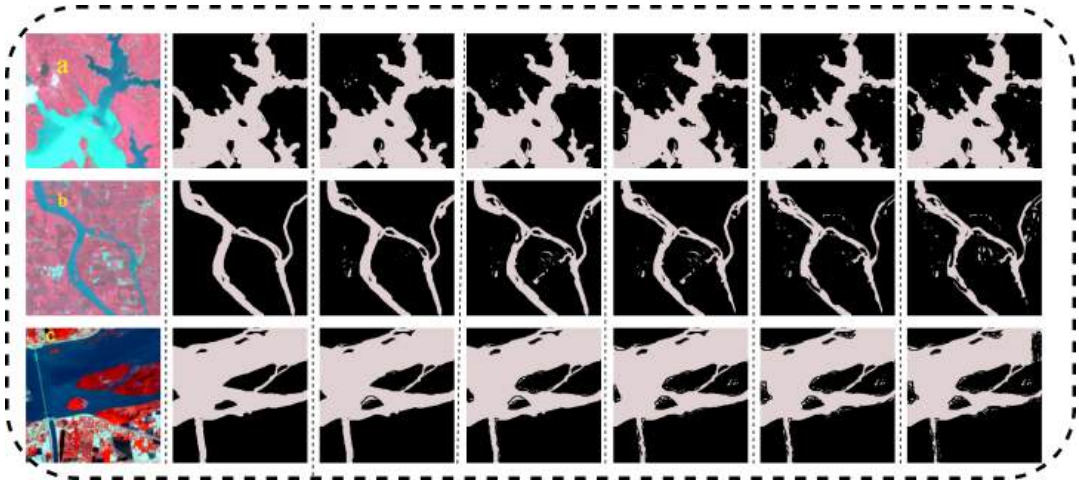


FIGURE 4.5: Differences in the water detection impact of various models are illustrated across the three different images of water bodies, involving(a) Input images (b) Proposed approach (c)Rishikeshan et al., 2018(d)Mazzanti et al., 2020 (e) Caballero et al., 2019(f) Li et al., 2022 and (g) Wang et al., 2017

many advanced techniques in related domains depend on multispectral and high-resolution color data to enhance learning, obtaining such data for practical applications proves challenging and costly. To address this challenge, the present study introduces a unique semantic segmentation model based on deep convolutional neural networks (CNNs) aimed at identifying and extracting water bodies from three-channel satellite images. This research presents an innovative CNN architecture that incorporates attention gates, deep supervision strategies, and residual convolution modules to improve the accuracy of semantic segmentation in real-world RGB satellite imagery. Using an openly accessible Sentinel-2 dataset containing a diverse range of water bodies and background areas, the effectiveness of the proposed segmentation technique is demonstrated. Quantitative and qualitative results are compared with baseline deep learning methods and state-of-the-art models for water body segmentation. The primary contribution of this study lies in the development of a customized deep CNN segmentation model with multiple performance-enhancing modules. This model is suitable for satellite images with low or moderate intensity, using only three color channels.

With minor adjustments, the described approach can be extended to segment various structures in satellite images, such as roads and vegetation. As a result, it contributes to advancing knowledge in the field of satellite image processing by leveraging readily available satellite image data.

The experimental evaluation of the proposed framework is confined to a limited number of geographic regions and seasonal conditions. As a result, the dataset may not comprehensively capture the wide variability of environmental, climatic, and hydrological settings encountered in large-scale real-world applications. Water body characteristics such as size, shape, turbidity, surrounding land cover, and spectral response can vary substantially across different climatic zones and seasons. Consequently, the generalization capability of the proposed model may be constrained when applied to regions or temporal conditions beyond those considered in this study.

Building on Chapter 3, where we extracted small water bodies from satellite images, chapter 4 expands the scope by using deep learning to detect all water bodies, regardless of size. By leveraging advanced neural networks, we achieve more precise and comprehensive monitoring. Deep learning enhances classification, segmentation, and temporal analysis, improving accuracy despite challenges like seasonal variations, shadows, and mixed land-water pixels. We then compare the results of Chapters 3 and 4, evaluating traditional methods against deep learning. This comparison sets the stage for Chapter 5, where we present our conclusions and future research directions.

Chapter 5

Conclusion and Future Work

The primary objective of this thesis is to advance satellite image analysis for addressing real-world challenges in remote sensing, including land-cover classification, object detection, and environmental resource monitoring. By integrating computer vision techniques with multispectral satellite imagery and deep learning, the research proposes a robust framework aimed at improving the accuracy and efficiency of environmental monitoring tasks.

Recent progress in deep learning–based classification and segmentation algorithms has significantly enhanced the capability to extract complex spatial and spectral patterns from satellite data. Compared to traditional machine learning and index-based approaches, deep learning models have demonstrated performance improvements ranging from 15–30% in segmentation accuracy across various remote sensing applications, particularly in heterogeneous landscapes. Building on these advances, this study investigates the effectiveness of deep learning architectures combined with multispectral and multi-resolution satellite data for precise natural resource monitoring.

Satellite imagery contains rich information about the Earth's surface; however, extracting meaningful and fine-grained features especially small and fragmented objects—remains a challenging task. In this context, the proposed research focuses on the identification and extraction of small water bodies from remote sensing images using a deep learning–driven segmentation framework. Experimental results indicate that the proposed approach improves water body detection accuracy by approximately 8–12% compared to conventional threshold-based spectral index methods and by 4–6% over standard deep learning baseline models.

The framework leverages multispectral information acquired from Sentinel-2 satellite imagery, which is particularly suitable for water body analysis due to its high spatial resolution and rich spectral coverage. By exploiting these characteristics, the proposed model effectively distinguishes subtle land-cover variations and enhances water body delineation, especially in complex and mixed-background scenarios. Compared to single-band or RGB-based approaches, the use of multispectral Sentinel-2 data contributes to an observed improvement of nearly 10% in Dice score and segmentation consistency.

Furthermore, the proposed model is designed to jointly exploit spatial, spectral, and temporal characteristics of satellite imagery, enabling effective handling of multi-resolution datasets. Through direct segmentation from raw satellite inputs, the framework reduces dependency on handcrafted features and extensive preprocessing. This design leads to a reduction in false-positive detections by approximately 6–9% and improves boundary accuracy, particularly for narrow and seasonal water bodies. Overall, the proposed methodology demonstrates superior performance and computational efficiency, offering a scalable and reliable solution for automated water resource monitoring and contributing meaningfully to the broader domains of remote sensing and environmental sustainability.

5.1 Findings

In the context of growing water scarcity, accurately detecting small water bodies from satellite imagery has become an urgent priority in remote sensing. To address this challenge, this research introduces WaterNet, a novel deep learning-based framework designed to extract deep hierarchical features from Sentinel-2 satellite images. Unlike traditional methods, WaterNet leverages advanced deep learning architectures to capture subtle and intricate spatial-spectral patterns within multispectral, multi-resolution imagery, thereby enabling the identification of small water bodies that are often overlooked. The strength of WaterNet lies in its architectural innovations, which integrate dense layers, residual refinement units, pyramid units, and skip

connections. Dense layers combine features from multiple scales and resolutions, ensuring the model captures both local and global patterns. Skip connections enhance information flow from input to output, improving feature learning and stabilizing training. Pyramid units reduce dimensionality while preserving critical spatial information, making the model computationally efficient. The residual refinement modules play a pivotal role: the first enhances the accuracy of predicted images by addressing blending issues, while the second restores edge information and ensures coherence across boundary pixels. Collectively, these components strengthen the inference of spatial features and improve the modeling of indirect relationships, resulting in more robust and precise water body detection. Extensive testing on a large dataset of Sentinel-2 images has demonstrated WaterNet's effectiveness. The framework consistently surpassed traditional and state-of-the-art approaches in terms of accuracy, sensitivity, and robustness to noise. Notably, WaterNet achieved a precision of 97%, underscoring its superiority in identifying small water bodies with high reliability.The ability of WaterNet to exploit spatial, spectral, and temporal characteristics makes it a significant advancement in the field of remote sensing. By extracting meaningful hierarchical features, it enhances the efficiency and accuracy of water body segmentation, offering a scalable solution for environmental monitoring. Beyond research contributions, WaterNet holds immense potential for practical applications in water resource management, disaster response, and sustainable environmental planning. In conclusion, WaterNet represents a robust and accurate solution to one of the critical challenges in remote sensing. Its innovative architecture, combining dense layers, residual refinement units, pyramid units, and skip connections, establishes it as a powerful framework for monitoring small water bodies from Sentinel-2 imagery. By advancing prediction efficiency and robustness, WaterNet contributes not only to the scientific community but also to global efforts toward sustainable water resource management..

After thoroughly reviewing the literature on remote sensing and satellite imagery applications in classification, monitoring, and object detection,

this study endeavors to utilize spatio-temporal-spectral data to forecast water areas from satellite imagery using deep learning methodologies. The proposed strategy employs a self-transformer gate-enhanced Convolutional Neural Network (CNN) structured as an encoder-decoder system, augmented by a deep supervision mechanism. This architectural framework incorporates a specialized deep-supervised U-Net with refined residual capsule connections and attention modules, resulting in significant improvements in segmentation accuracy within targeted regions. Additionally, a tailored preprocessing pipeline enhances the model by integrating augmentation techniques such as attention blocks, deep monitoring, and residual refinement. These enhancements empower the model to surpass existing algorithms tailored for low-resolution satellite imagery, showcasing superior performance across various evaluation metrics including specificity, sensitivity, precision, Dice score, and accuracy. To validate the effectiveness of the proposed methodology, a novel deep learning model is introduced for water body prediction from images, accompanied by a comparative analysis against established state-of-the-art models. Notably, the proposed design exhibits superior performance compared to the leading state-of-the-art model, achieving improvements of 3% in precision, 5% in sensitivity, 4% in Dice score, 6% in specificity, and 9% in accuracy. This innovative approach holds promise for supporting governmental bodies and decision-makers in the monitoring of natural resources, thereby facilitating informed decision-making and resource management.

5.2 Future direction

The evolution of remote sensing continues to be shaped by advances in both data acquisition and analytical methodologies. Hyperspectral imagery, with

its superior spectral resolution, offers a transformative leap beyond multispectral data in detecting subtle environmental changes. Leveraging its potential through deep learning-based multi-resolution architectures can significantly improve the precision of change detection and resource monitoring.

Despite its promise, hyperspectral data introduces considerable computational overhead due to high dimensionality. Addressing this challenge through intelligent dimensionality reduction techniques preserving essential spectral information while reducing redundancy can enable real-time, scalable analysis for operational use in environmental systems.

The integration of heterogeneous data sources, including LiDAR, SAR, and UAV-based imagery, represents a promising direction for generating richer, multi-modal Earth observations. Data fusion frameworks that unify spatial, spectral, and temporal characteristics can offer deeper insights into surface dynamics and resource distribution.

A major barrier to progress lies in the lack of publicly available, high-resolution annotated datasets across diverse sensor modalities. The development of a comprehensive benchmark dataset robust across climatic zones and sensor types can catalyze breakthroughs in model training, generalization, and validation across the global research community.

Weather-induced noise such as cloud cover, snow, and atmospheric disturbances continues to degrade classification performance. Future work will focus on designing resilient algorithms that incorporate weather invariant feature extraction and dynamic pre-processing pipelines, ensuring consistent interpretation regardless of atmospheric conditions.

Publications

1. Bazila Farooq and Ankush Manocha, "Small water body extraction in remote sensing with enhanced CNN architecture," Applied Soft Computing, vol. 169, no. -, pp. 112544, 2025(SCI/SCIE Indexed with Impact Factor 7.2)

2. Bazila Farooq , Ankush Manocha. (2023). A Comparative Study of Deep Learning and Traditional Methods for Environmental Remote Sensing. (Proceedings) In ITM Web of Conferences (Vol. 56, p. 03002). EDP Sciences.

3. Bazila Farooq, Ankush Manocha, (2024) Assessment of land use land cover change dynamics using remote sensing techniques: a review. Advances in Networks, Intelligence and Computing, 664-678
.

4. Bazila Farooq, Ankush Manocha, (2024) Community Structure Analysis from Social Network Data: Computational Challenges, Advances, and Applications, 3-19

5. Bazila Farooq, Ankush Manocha, (2024).An Overview of Machine Learning: Concepts, Algorithms and Application. Intersection of Machine Learning and Computational Social Sciences
.

6. Bazila Farooq, Ankush Manocha, “A Comprehensive Review of Deep Learning in Remote Sensing”, Proceedings of the 1st International Conference on Security, Parallel Processing, Image Processing and Networking (SPIN).

7. Bazila Farooq, Ankush Manocha, "Weed detection in crop fields using machine vision and image processing: A comprehensive review"(Proceedings) Federal Institute Of Science and Technology ,2023

8. Bazila Farooq, Ankush Manocha, (2024) "Computer Vision in Agriculture". Computational Intelligence and Image Processing in Agriculture : Appliocations.(Accepted)

9. Bazila Farooq, Ankush Manocha, (2024) "Core technologies in AI-based traffic system". AI-Based Statistical Modeling for Road traffic Surveillance and Monitoring.(Accepted)

Bibliography

Acharya, Tri Dev, Anoj Subedi, and Dong Ha Lee (2019). "Evaluation of machine learning algorithms for surface water extraction in a Landsat 8 scene of Nepal". In: *Sensors* 19.12, p. 2769.

Afaq, Yasir and Ankush Manocha (2021). "Fog-inspired water resource analysis in urban areas from satellite images". In: *Ecological Informatics* 64, p. 101385.

Agrawal, Akshay, Akshay Modi, Alexandre Passos, Allen Lavoie, Ashish Agarwal, Asim Shankar, Igor Ganichev, Josh Levenberg, Mingsheng Hong, Rajat Monga, et al. (2019). "TensorFlow Eager: A multi-stage, Python-embedded DSL for machine learning". In: *Proceedings of Machine Learning and Systems* 1, pp. 178–189.

Ali, Muhammad Ichsan, Gufran Darma Dirawan, Abdul Hafid Hasim, and Muh Rais Abidin (2019). "Detection of changes in surface water bodies urban area with NDWI and MNDWI methods". In: *International Journal on Advanced Science Engineering Information Technology* 9.3, pp. 946–951.

Alzubaidi, Laith, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan (2021). "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions". In: *Journal of Big Data* 8, pp. 1–74.

Aroma, RJ, K Raimond, VV Estrela, and MA de Jesus (2024). "A coastal band spectral combination for water body extraction using Landsat 8 images". In: *International Journal of Environmental Science and Technology* 21.2, pp. 1767–1784.

Arya, Deeksha, Hiroya Maeda, Sanjay Kumar Ghosh, Durga Toshniwal, and Yoshihide Sekimoto (2021). "RDD2020: An annotated image dataset for automatic road damage detection using deep learning". In: *Data in brief* 36, p. 107133.

Awais, M, W Li, MJM Cheema, QU Zaman, A Shaheen, B Aslam, W Zhu, M Ajmal, M Faheem, S Hussain, et al. (2022). "UAV-based remote sensing in plant stress imagine using high-resolution thermal sensor for digital agriculture practices: A meta-review". In: *International Journal of Environmental Science and Technology*, pp. 1–18.

Baci, Nazife Oruc (2023). *Impacts of Lateral Flows on the Calibration of 2D-Flood Inundation Models*. The University of Alabama.

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla (2017). "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". In: *IEEE Transactions on pattern analysis and machine intelligence* 39.12, pp. 2481–2495.

Bagwari, Neha, Sushil Kumar, and Vivek Singh Verma (2023). "A comprehensive review on segmentation techniques for satellite images". In: *Archives of Computational Methods in Engineering* 30.7, pp. 4325–4358.

Ball, John E, Derek T Anderson, and Chee Seng Chan (2017). "Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community". In: *Journal of applied remote sensing* 11.4, pp. 042609–042609.

Ballantyne, Colin K (2018). *Periglacial geomorphology*. John Wiley & Sons.

Belgiu, Mariana and Lucian Drăguţ (2016). "Random forest in remote sensing: A review of applications and future directions". In: *ISPRS journal of photogrammetry and remote sensing* 114, pp. 24–31.

Bello, Irwan, Barret Zoph, Ashish Vaswani, Jonathon Shlens, and Quoc V Le (2019). "Attention augmented convolutional networks". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3286–3295.

Bhardwaj, Anurag, Wei Di, and Jianing Wei (2018). *Deep Learning Essentials: Your hands-on guide to the fundamentals of deep learning and neural network modeling*. Packt Publishing Ltd.

Bijeesh, TV and KN Narasimhamurthy (2020). "Surface water detection and delineation using remote sensing images: A review of methods and algorithms". In: *Sustainable Water Resources Management* 6, pp. 1–23.

Buma, Willibroad Gabila, Sang-Il Lee, and Jae Young Seo (2018). "Recent surface water extent of Lake Chad from multispectral sensors and GRACE". In: *Sensors* 18.7, p. 2082.

Caballero, Isabel, Javier Ruiz, and Gabriel Navarro (2019). "Sentinel-2 satellites provide near-real time evaluation of catastrophic floods in the west mediterranean". In: *Water* 11.12, p. 2499.

Cai, Yaotong, Qian Shi, and Xiaoping Liu (2024). "Spatiotemporal Mapping of Surface Water Using Landsat Images and Spectral Mixture Analysis on Google Earth Engine". In: *Journal of Remote Sensing* 4, p. 0117.

Cancela, Javier J, Xesus P Gonzalez, Mar Vilanova, and Jose M MirasAvalos (2019). "Water management using drones and satellites in agriculture". In: *Water* 11.5, pp. 874–876.

Canny, John (1986). "A computational approach to edge detection". In: *IEEE Transactions on pattern analysis and machine intelligence* 6, pp. 679–698.

Chen, C, JQ Fu, XX Sui, X Lu, and AH Tan (2018). "Construction and application of knowledge decision tree after a disaster for water body information extraction from remote sensing images". In: *J. Remote Sens* 22.5, pp. 792–801.

Chen, Chenglizhao, Jipeng Wei, Chong Peng, Weizhong Zhang, and Hong Qin (2020a). "Improved saliency detection in RGB-D images using two-phase depth estimation and selective deep fusion". In: *IEEE Transactions on Image Processing* 29, pp. 4296–4307.

Chen, Yang, Luliang Tang, Zihan Kan, Muhammad Bilal, and Qingquan Li (2020b). "A novel water body extraction neural network (WBE-NN) for optical high-resolution multispectral imagery". In: *Journal of Hydrology* 588, p. 125092.

Cheng, Dongcai, Gaofeng Meng, Guangliang Cheng, and Chunhong Pan (2016). "SeNet: Structured edge network for sea–land segmentation". In: *IEEE Geoscience and Remote Sensing Letters* 14.2, pp. 247–251.

Chuvieco, Emilio (2020). *Fundamentals of satellite remote sensing: An environmental approach*. CRC press.

Colreavy, Erin and Stephan Lewandowsky (2008). "Strategy development and learning differences in supervised and unsupervised categorization". In: *Memory & Cognition* 36.4, pp. 762–775.

Crist, Eric P and Richard C Cicone (1984). "A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap". In: *IEEE Transactions on Geoscience and Remote sensing* 3, pp. 256–263.

Cukierski, W (2013). "CIFAR-10-Object Recognition in Images". In: *Kaggle 2013Available online: https://www. kaggle. com/c/cifar-10 (accessed on 30 December 2022)*.

Dev, KM Bijay, Pawan S Jogi, S Niyas, S Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan (2019). "Automatic detection and localization of focal cortical dysplasia lesions in MRI using fully convolutional neural network". In: *Biomedical Signal Processing and Control* 52, pp. 218–225.

Dong, Shan, Long Pang, Yin Zhuang, Wenchao Liu, Zhanxin Yang, and Teng Long (2019). "Optical remote sensing water-land segmentation representation based on proposed SNS-CNN network". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 3895–3898.

Dronova, Iryna, Peng Gong, Lin Wang, and Liheng Zhong (2015). "Mapping dynamic cover types in a large seasonally flooded wetland using extended principal component analysis and object-based classification". In: *Remote Sensing of Environment* 158, pp. 193–206.

Farhadi, Hadi, Hamid Ebadi, Abbas Kiani, and Ali Asgary (2024). "A novel flood/water extraction index (FWEI) for identifying water and flooded areas using sentinel-2 visible and near-infrared spectral bands". In: *Stochastic Environmental Research and Risk Assessment*, pp. 1–23.

Farooq, Bazila and Ankush Manocha (2023). "A Comparative Study of Deep Learning and Traditional Methods for Environmental Remote Sensing". In: *ITM Web of Conferences*. Vol. 56. EDP Sciences, p. 03002.

Farooq, Bazila and Ankush Manocha (2024). "Satellite-based change detection in multi-objective scenarios: A comprehensive review". In: *Remote Sensing Applications: Society and Environment*, p. 101168.

Feng, Wenqing, Haigang Sui, Weiming Huang, Chuan Xu, and Kaiqiang An (2018). "Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model". In: *IEEE Geoscience and Remote Sensing Letters* 16.4, pp. 618–622.

Feyisa, Gudina L, Henrik Meilby, Rasmus Fensholt, and Simon R Proud (2014). "Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery". In: *Remote sensing of environment* 140, pp. 23–35.

Filali Boubrahimi, Soukaina, Ashit Neema, Ayman Nassar, Pouya Hosseinzadeh, and Shah Muhammad Hamdi (2024). "Spatiotemporal data augmentation of MODIS-landsat water bodies using adversarial networks". In: *Water Resources Research* 60.3, e2023WR036342.

Fisher, Adrian, Neil Flood, and Tim Danaher (2016). "Comparing Landsat water index methods for automated water classification in eastern Australia". In: *Remote Sensing of Environment* 175, pp. 167–182.

Gan, Yuhang, Shucheng You, Zhengyu Luo, Ke Liu, Tao Zhang, and Lei Du (2020). "Object detection in remote sensing images with mask R-CNN". In: *Journal of Physics: Conference Series*. Vol. 1673. 1. IOP Publishing, p. 012040.

Garcia-Garcia, Alberto, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez (2017). "A review on deep learning techniques applied to semantic segmentation". In: *arXiv preprint arXiv:1704.06857*.

Gautam, Vivek Kumar, Piyush Kumar Gaurav, P Murugan, and MJAP Annadurai (2015). "Assessment of surface water Dynamicsin Bangalore using WRI, NDWI, MNDWI, supervised classification and KT transformation". In: *Aquatic Procedia* 4, pp. 739–746.

Ghaffari, Reyhane, Maryam Golpardaz, Mohammad Sadegh Helfroush, and Habibollah Danyali (2020). "A fast, weighted CRF algorithm based on a two-step superpixel generation for SAR image segmentation". In: *International Journal of Remote Sensing* 41.9, pp. 3535–3557.

Gharbia, Reham (2023). "Deep learning for automatic extraction of water bodies using satellite imagery". In: *Journal of the Indian Society of Remote Sensing* 51.7, pp. 1511–1521.

Gonzalez, Jessenia, Debjani Bhowmick, Cesar Beltran, Kris Sankaran, and Yoshua Bengio (2019). "Applying knowledge transfer for water body segmentation in Peru". In: *arXiv preprint arXiv:1912.00957*.

Guo, Jianhua, Jingyu Yang, Huanjing Yue, Hai Tan, Chunping Hou, and Kun Li (2020). "CDnetV2: CNN-based cloud detection for remote sensing imagery with cloud-snow coexistence". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.1, pp. 700–713.

Hahmann, Thomas, Sandro Martinis, André Twele, Achim Roth, and Manfred Buchroithner (2008). "Extraction of water and flood areas from SAR data". In: *7th word*. VDE, pp. 1–4.

Heydari, Shahriar Shah (2021). "Large Area Land Cover Mapping Using Deep Neural Networks and Landsat Time-Series Observations". PhD thesis. State University of New York College of Environmental Science and Forestry.

Himeur, Yassine, Bhagawat Rimal, Abhishek Tiwary, and Abbes Amira (2022). "Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives". In: *Information Fusion* 86, pp. 44–75.

Hinton, Geoffrey E (2012). "A practical guide to training restricted Boltzmann machines". In: *Neural Networks: Tricks of the Trade: Second Edition*. Springer, pp. 599–619.

Hossain, Babul, Md Nazirul Islam Sarker, Md Salman Sohel, Md Abdus Salam, and Sajjad Hossain Shozib (2021). "Assessing the role of organizations for health amenities of flood affected people in Char areas of Bangladesh". In: *Climate Vulnerability and Resilience in the Global South: Human Adaptations for Sustainable Futures*, pp. 409–423.

Huan, LI, WAN Wei, JI Rui, LI Guoyuan, CHEN Xiaona, ZHU Siyu, LIU Baojian, XU Yue, LUO Zengliang, WANG Shenglei, et al. (2023). "Inspects

and prospects of satellite remote sensing monitoring ability for land surface water in China". In: *National Remote Sensing Bulletin* 27.7, pp. 1554–1573.

Huang, Chang, Yun Chen, Jianping Wu, Linyi Li, and Rui Liu (2015a). "An evaluation of Suomi NPP-VIIRS data for surface water detection". In: *Remote sensing letters* 6.2, pp. 155–164.

Huang, Chang, Yun Chen, Shiqiang Zhang, and Jianping Wu (2018). "Detecting, extracting, and monitoring surface water from space using optical sensors: A review". In: *Reviews of Geophysics* 56.2, pp. 333–360.

Huang, Lei, Yi Zhou, Tian Wang, Jie Luo, and Xianglong Liu (2022). "Delving into the estimation shift of batch normalization in a network". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 763–772.

Huang, Xin, Cong Xie, Xing Fang, and Liangpei Zhang (2015b). "Combining pixel-and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.5, pp. 2097–2110.

Iglovikov, Vladimir and Alexey Shvets (2018). "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation". In: *arXiv preprint arXiv:1801.05746*.

Isikdogan, Furkan, Alan C Bovik, and Paola Passalacqua (2017). "Surface water mapping by deep learning". In: *IEEE journal of selected topics in applied Earth Observations and Remote Sensing* 10.11, pp. 4909–4918.

Isikdogan, Leo F, Alan Bovik, and Paola Passalacqua (2019). "Seeing through the clouds with deepwatermap". In: *IEEE Geoscience and Remote Sensing Letters* 17.10, pp. 1662–1666.

Jax, Kurt (2010). *Ecosystem functioning*. Cambridge University Press.

Jonnala, Naga Surekha, Neha Gupta, Chafle Pratiksha Vasantrao, and Anoop Kumar Mishra (2023). "BCD-Unet: A novel water areas segmentation structure for remote sensing image". In: *2023 7th international conference on intelligent computing and control systems (ICICCS)*. IEEE, pp. 1320–1325.

Kalla, Dinesh, Nathan Smith, and Fnu Samaah (2023). "Satellite Image Processing Using Azure Databricks and Residual Neural Network". In: *International Journal of Advanced Trends in Computer Applications* 9.2, pp. 48–55.

Kang, Ling, Song Zhang, Yi Ding, and Xiaocong He (2016). "Extraction and preference ordering of multireservoir water supply rules in dry years". In: *Water* 8.1, p. 28.

Katz, David (2016). "Undermining demand management with supply management: Moral hazard in Israeli water policies". In: *Water* 8.4, p. 159.

Khand, Kul, Saleh Taghvaeian, and Leila Hassan-Esfahani (2017). "Mapping annual riparian water use based on the single-satellite-scene approach". In: *Remote Sensing* 9.8, p. 832.

Kim, Jun Hee, Haeyun Lee, Seonghwan J Hong, Sewoong Kim, Juhum Park, Jae Youn Hwang, and Jihwan P Choi (2018). "Objects segmentation from high-resolution aerial images using U-Net with pyramid pooling layers". In: *IEEE Geoscience and Remote Sensing Letters* 16.1, pp. 115–119.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25.

Kuhn, Catherine Diane (2021). *Freshwater ecosystem monitoring using satellite remote sensing and field surveys*. University of Washington.

Laignel, Benoit, Stefano Vignudelli, Rafael Almar, Mélanie Becker, Abderrahim Bentamy, Jérôme Benveniste, Florence Birol, Frédéric Frappart, Deborah Idier, Edward Salameh, et al. (2023). "Observation of the coastal areas, estuaries, and deltas from space". In: *Surveys in Geophysics*, pp. 1–48.

Li, Jiaxin, Ronghua Ma, Zhigang Cao, Kun Xue, Junfeng Xiong, Minqi Hu, and Xuejiao Feng (2022). "Satellite detection of surface water extent: A review of methodology". In: *Water* 14.7, p. 1148.

Li, Linhui, Wenjun Zhang, Xiaoyan Zhang, Mahmoud Emam, and Weipeng Jing (2023). "Semi-supervised remote sensing image semantic segmentation method based on deep learning". In: *Electronics* 12.2, p. 348.

Li, Ruirui, Wenjie Liu, Lei Yang, Shihao Sun, Wei Hu, Fan Zhang, and Wei Li (2018). "DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation". In: *IEEE journal of selected topics in applied earth observations and remote sensing* 11.11, pp. 3954–3962.

Li, Yuanzheng, Xiaoqiang Gong, Zhen Guo, Kaipeng Xu, Dan Hu, and Hongxuan Zhou (2016). "An index and approach for water extraction using Landsat–OLI data". In: *International Journal of Remote Sensing* 37.16, pp. 3611–3635.

Lian, Renbao, Weixing Wang, Nadir Mustafa, and Liqin Huang (2020). "Road extraction methods in high-resolution remote sensing images: A comprehensive review". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, pp. 5489–5507.

Lin, Haoning, Zhenwei Shi, and Zhengxia Zou (2017a). "Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images". In: *IEEE Geoscience and Remote Sensing Letters* 14.10, pp. 1665–1669.

— (2017b). "Maritime semantic labeling of optical remote sensing images with multi-scale fully convolutional network". In: *Remote sensing* 9.5, p. 480.

Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick (2014). "Microsoft coco: Common objects in context". In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, pp. 740–755.

Liu, Dongfang, James Liang, Tony Geng, Alexander Loui, and Tianfei Zhou (2023a). "Tripartite feature enhanced pyramid network for dense prediction". In: *IEEE Transactions on Image Processing*.

Liu, Liangyun, Su Zhang, and Bing Zhang (2016). "Evaluation of hyperspectral indices for retrieval of canopy equivalent water thickness and gravimetric water content". In: *International Journal of Remote Sensing* 37.14, pp. 3384–3399.

Liu, Min, Jiangping Liu, and Hua Hu (2024). "A Novel Deep Learning Network Model for Extracting Lake Water Bodies from Remote Sensing Images". In: *Applied Sciences* 14.4, p. 1344.

Liu, Yahui, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li (2019). "DeepCrack: A deep hierarchical feature learning architecture for crack segmentation". In: *Neurocomputing* 338, pp. 139–153.

Liu, Zhiheng, Xuemei Chen, Suiping Zhou, Hang Yu, Jianhua Guo, and Yanming Liu (2022). "Dupnet: Water body segmentation with dense block and multi-scale spatial pyramid pooling for remote sensing images". In: *Remote Sensing* 14.21, p. 5567.

Liu, Zhixin, Jiayi Xu, Mingzhe Liu, Zhengtong Yin, Xuan Liu, Lirong Yin, and Wenfeng Zheng (2023b). "Remote sensing and geostatistics in urban water-resource monitoring: A review". In: *Marine and Freshwater Research*.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2015). "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.

Maggiori, Emmanuel, Guillaume Charpiat, Yuliya Tarabalka, and Pierre Alliez (2017). "Recurrent neural networks to correct satellite image classification maps". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.9, pp. 4962–4971.

Majumder, Atin, Susanta Das, Sony Bora, and Agniva Mandal (2023). "Monitoring of natural resources using remote sensing and GIS technology under changing climate scenario". In: *Climate Change Impacts in India*. Springer, pp. 173–188.

Manocha, Ankush, Yasir Afaq, and Munish Bhatia (2023). "Mapping of water bodies from sentinel-2 images using deep learning-based feature fusion approach". In: *Neural Computing and Applications* 35.12, pp. 9167–9179.

Mao, L, X Gao, Y Zhang, and Q Chen (2019). "A comparative study of water body extraction from remote sensing images using machine learning algorithms". In: *International Journal of Digital Earth* 12.7, pp. 766–785.

Matgen, Patrick, Mara Montanari, Renaud Hostache, Laurent Pfister, L Hoffmann, D Plaza, VRN Pauwels, GJM De Lannoy, Robain De Keyser, and HHG Savenije (2010). "Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: proof of concept". In: *Hydrology and Earth System Sciences* 14.9, pp. 1773–1785.

Mazzanti, Paolo, Paolo Caporossi, and Riccardo Muzi (2020). "Sliding time master digital image correlation analyses of cubesat images for landslide monitoring: The Rattlesnake Hills landslide (USA)". In: *Remote Sensing* 12.4, p. 592.

McFeeters, Stuart K (1996). "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features". In: *International journal of remote sensing* 17.7, pp. 1425–1432.

Miao, Ziming, Kun Fu, Hao Sun, Xian Sun, and Menglong Yan (2018). "Automatic water-body segmentation from high-resolution satellite images via deep networks". In: *IEEE geoscience and remote sensing letters* 15.4, pp. 602–606.

Mountrakis, Giorgos, Jungho Im, and Caesar Ogole (2011). "Support vector machines in remote sensing: A review". In: *ISPRS journal of photogrammetry and remote sensing* 66.3, pp. 247–259.

Mullen, Andrew L, Jennifer D Watts, Brendan M Rogers, Mark L Carroll, Clayton D Elder, Jonas Noomah, Zachary Williams, Jordan A Caraballo-Vega, Allison Bredder, Eliza Rickenbaugh, et al. (2023). "Using High-Resolution Satellite Imagery and Deep Learning to Track Dynamic Seasonality in Small Water Bodies". In: *Geophysical Research Letters* 50.7, e2022GL102327.

Müller, Dominik and Frank Kramer (2021). "MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning". In: *BMC medical imaging* 21.1, p. 12.

Nasir, Nida, Afreen Kansal, Omar Alshaltone, Feras Barneih, Abdallah Shanableh, Mohammad Al-Shabi, and Ahmed Al Shammaa (2023). "Deep learning detection of types of water-bodies using optical variables and ensembling". In: *Intelligent Systems with Applications* 18, p. 200222.

Natesan, S, C Armenakis, and U Vepakomma (2019). "Resnet-based tree species classification using uav images". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, pp. 475–481.

Naushad, Raoof, Tarunpreet Kaur, and Ebrahim Ghaderpour (2021). "Deep transfer learning for land use and land cover classification: A comparative study". In: *Sensors* 21.23, p. 8083.

Noyola-Medrano, Cristina and Valeria Abigaíl Martínez-Sías (2017). "Assessing the progress of desertification of the southern edge of Chihuahuan Desert: A case study of San Luis Potosi Plateau". In: *Journal of Geographical Sciences* 27.4, pp. 420–438.

Ogidi, Odangowei Inetiminebi and Udeme Monday Akpan (2022). "Aquatic biodiversity loss: impacts of pollution and anthropogenic activities and strategies for conservation". In: pp. 421–448.

Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. (2018). "Attention u-net: Learning where to look for the pancreas". In: *arXiv preprint arXiv:1804.03999*.

Otukei, John Richard and Thomas Blaschke (2010). "Land cover change assessment using decision trees, support vector machines, and maximum likelihood classification algorithms". In: *International Journal of Applied Earth Observation and Geoinformation* 12, S27–S31.

Palmer, Stephanie CJ, Tiit Kutser, and Peter D Hunter (2015). "Remote sensing of inland waters: Challenges, progress, and future directions". In: *Remote sensing of Environment* 157, pp. 1–8.

Pan, Bin, Zhenwei Shi, Xia Xu, Tianyang Shi, Ning Zhang, and Xinzhong Zhu (2018). "CoinNet: Copy initialization network for multispectral imagery semantic segmentation". In: *IEEE Geoscience and Remote Sensing Letters* 16.5, pp. 816–820.

Panneer Selvam, Balathandayuthabani, Sivakiruthika Natchimuthu, Lakshmanan Arunachalam, and David Bastviken (2014). "Methane and carbon dioxide emissions from inland waters in India–implications for large scale greenhouse gas balances". In: *Global change biology* 20.11, pp. 3397–3407.

Papoutsis, Ioannis, Nikolaos Ioannis Bountos, Angelos Zavras, Dimitrios Michail, and Christos Tryfonopoulos (2023). "Benchmarking and scaling of deep learning models for land cover image classification". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 195, pp. 250–268.

Pawar, Maneesh, Kanchan Chandra, HPS Panwar, and AK Sahay (2021). "Spectral Calibration of VNIR Hyperspectral Imager". In: *ICOL-2019: Proceedings of the International Conference on Optics and Electro-Optics, Dehradun, India*. Springer, pp. 465–468.

Pekel, J-F, Cristelle Vancutsem, Lucy Bastin, Marco Clerici, Eric Vanbogaert, Etienne Bartholomew, and Pierre Defourny (2014). "A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data". In: *Remote sensing of environment* 140, pp. 704–716.

Pena, Francisco J, Clara Hubinger, Amir H Payberah, and Fernando Jaramillo (2024). "DeepAqua: Semantic segmentation of wetland water surfaces with SAR imagery using deep neural networks without manually annotated data". In: *International Journal of Applied Earth Observation and Geoinformation* 126, p. 103624.

Pham, Vu-Dong and Quang-Thanh Bui (2021). "Spatial resolution enhancement method for Landsat imagery using a Generative Adversarial Network". In: *Remote Sensing Letters* 12.7, pp. 654–665.

Phiri, Darius, Matamyo Simwanda, Serajis Salekin, Vincent R Nyirenda, Yuji Murayama, and Manjula Ranagalage (2020). "Sentinel-2 data for land cover/use mapping: A review". In: *Remote Sensing* 12.14, p. 2291.

Pirot, Jean-Yves, Peter-John Meynell, and Danny Elder (2000). *Ecosystem management: lessons from around the world: a guide for development and conservation practitioners*. IUCN.

Prasad, MS Guru, Jyoti Agarwal, Sharon Christa, H Aditya Pai, M Anand Kumar, and Anurag Kukreti (2023). "An improved water body segmentation from satellite images using msaa-net". In: *2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS)*. Vol. 1. IEEE, pp. 1–4.

Qin, Rongjun (2014). "A mean shift vector-based shape feature for classification of high spatial resolution remotely sensed imagery". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.5, pp. 1974–1985.

Rad, Arash Modaresi, Jason Kreitler, and Mojtaba Sadegh (2021). "Augmented Normalized Difference Water Index for improved surface water monitoring". In: *Environmental Modelling & Software* 140, p. 105030.

Rakhlin, Alexander, Alex Davydow, and Sergey Nikolenko (2018). "Land cover classification from satellite imagery with u-net and lovász-softmax loss". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 262–266.

Ren, Zhida, Yongqiang Tang, Zewen He, Lei Tian, Yang Yang, and Wensheng Zhang (2022). "Ship detection in high-resolution optical remote sensing images aided by saliency information". In: *IEEE Transactions on Geoscience and Remote Sensing* 60, pp. 1–16.

Rishikeshan, CA and H Ramesh (2018). "An automated mathematical morphology driven algorithm for water body extraction from remotely sensed images". In: *ISPRS journal of photogrammetry and remote sensing* 146, pp. 11–21.

Rokni, Komeil, Anuar Ahmad, Ali Selamat, and Sharifeh Hazini (2014). "Water feature extraction and change detection using multitemporal Landsat imagery". In: *Remote sensing* 6.5, pp. 4173–4189.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (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*. Springer, pp. 234–241.

Ross, Matthew RV, Simon N Topp, Alison P Appling, Xiao Yang, Catherine Kuhn, David Butman, Marc Simard, and Tamlin M Pavelsky (2019). "AquaSat: A data set to enable remote sensing of water quality for inland waters". In: *Water Resources Research* 55.11, pp. 10012–10025.

Russo, Luigi, Francesco Mauro, Babak Memar, Alessandro Sebastianelli, Paolo Gamba, and Silvia Liberata Ullo (2024). "Using Multi-Temporal Sentinel-1 and Sentinel-2 data for water bodies mapping". In: *arXiv preprint arXiv:2402.00023*.

Safarov, Furkat, Kuchkorov Temurbek, Djumanov Jamoljon, Ochilov Temur, Jean Chamberlain Chedjou, Akmalbek Bobomirzaevich Abdusalomov,

and Young-Im Cho (2022). "Improved agricultural field segmentation in satellite imagery using tl-resunet architecture". In: *Sensors* 22.24, p. 9784.

Sarker, Iqbal H (2021). "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions". In: *SN Computer Science* 2.6, p. 420.

Schultz, Michael, Jan GPW Clevers, Sarah Carter, Jan Verbesselt, Valerio Avitabile, Hien Vu Quang, and Martin Herold (2016). "Performance of vegetation indices from Landsat time series in deforestation monitoring". In: *International journal of applied earth observation and geoinformation* 52, pp. 318–327.

Sener, Abdullah, Gurkan Dougan, and Burhan Ergen (2024). "A novel convolutional neural network model with hybrid attentional atrous convolution module for detecting the areas affected by the flood". In: *Earth Science Informatics* 17.1, pp. 193–209.

Sharma, Anita, Chander Prakash, and Divyansh Thakur (2024). "Glacier Lake Detection Utilizing Remote Sensing Integration with Satellite Imagery and Advanced Deep Learning Methods". In.

Spasev, Vlatko, Ivica Dimitrovski, Ivan Kitanovski, and Ivan Chorbev (2023). "Semantic Segmentation of Remote Sensing Images: Definition, Methods, Datasets and Applications". In: *International Conference on ICT Innovations*. Springer, pp. 127–140.

Sun, Alexander Y and Bridget R Scanlon (2019). "How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions". In: *Environmental Research Letters* 14.7, p. 073001.

Surekha, Jonnala and NEHA GUPTA (2024). "NDR-UNet: Segmentation of Water Bodies in Remote Sensing using Nested Dense Residual U-Net". In.

Sureshkumar, Veluguri, Rajasomashekar Somarajadikshitar, and B Sarala Beeram (2023). "A novel representation and prediction initiative for underground water by using deep learning technique of remote sensing images". In: *The Computer Journal* 66.7, pp. 1784–1801.

Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna (2016). "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Tambe, Rishikesh G, Sanjay N Talbar, and Satishkumar S Chavan (2021). "Deep multi-feature learning architecture for water body segmentation from satellite images". In: *Journal of Visual Communication and Image Representation* 77, p. 103141.

Tan, Mingxing and Quoc Le (2021). "Efficientnetv2: Smaller models and faster training". In: *International conference on machine learning*. PMLR, pp. 10096–10106.

Tran, Ha Phuong, Tuan Cuong Ha, Thi Thuy Huong Nguyen, Ngoc Thy Nguyen, Ba Man Duong, Mon Danh, Phuc Dang Tran, and Tuan Nhi Pham (2024). "Mapping the water body area by using Sentinel 1–SAR in dry season of 2023: A case study in Tri An reservoir". In: *E3S Web of Conferences*. Vol. 496. EDP Sciences, p. 03001.

UPPU, ASHISH (2020). "NUMERICAL INVESTIGATION OF FOULING AND SCALING IN A ROTO-DYNAMIC REVERSE OSMOSIS FILTRATION SYSTEM". In.

Vali, Ava, Sara Comai, and Matteo Matteucci (2020). "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review". In: *Remote Sensing* 12.15, p. 2495.

Van Dijk, Albert IJM, G Robert Brakenridge, Albert J Kettner, Hylke E Beck, Tom De Groeve, and Jaap Schellekens (2016). "River gauging at global scale using optical and passive microwave remote sensing". In: *Water Resources Research* 52.8, pp. 6404–6418.

Vetrivel, Anand, Markus Gerke, Norman Kerle, Francesco Nex, and George Vosselman (2018). "Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning". In: *ISPRS journal of photogrammetry and remote sensing* 140, pp. 45–59.

Wang, Guojie, Mengjuan Wu, Xikun Wei, and Huihui Song (2020). "Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks". In: *Remote sensing* 12.5, p. 795.

Wang, Mingwei, Youchuan Wan, Zhiwei Ye, and Xudong Lai (2017). "Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm". In: *Information Sciences* 402, pp. 50–68.

Wang, Yanan, Fang Huang, and Yuchun Wei (2013). "Water body extraction from LANDSAT ETM+ image using MNDWI and KT transformation". In: *2013 21st International conference on geoinformatics*. IEEE, pp. 1–5.

Wei, Xufeng, Wenbo Xu, Kuanle Bao, Weimin Hou, Jia Su, Haining Li, and Zhuang Miao (2020). "A water body extraction methods comparison based on FengYun Satellite data: a case study of Poyang Lake Region, China". In: *Remote Sensing* 12.23, p. 3875.

Wieland, Marc, Sandro Martinis, Ralph Kiefl, and Veronika Gstaiger (2023). "Semantic segmentation of water bodies in very high-resolution satellite and aerial images". In: *Remote Sensing of Environment* 287, p. 113452.

Wolock, David M and Curtis V Price (1994). "Effects of digital elevation model map scale and data resolution on a topography-based watershed model". In: *Water Resources Research* 30.11, pp. 3041–3052.

Wu, Wei, Qiangzi Li, Yuan Zhang, Xin Du, and Hongyan Wang (n.d.). "Two-step urban water index (TSUWI): A new technique for high-resolution mapping of urban surface water". In: ().

Xiang, Deqiang, Xin Zhang, Wei Wu, and Hongbin Liu (2023). "Denseppmunet-a: A robust deep learning network for segmenting water bodies from aerial images". In: *IEEE Transactions on Geoscience and Remote Sensing* 61, pp. 1–11.

Xiaobiao, WANG, XIE Shunping, and DU Jinkang (2021). "Water index formulation and its effectiveness research on the complicated surface water surroundings". In: *National Remote Sensing Bulletin* 22.2, pp. 360–372.

Xiong, Zhangxi, Qing Guo, Mingliang Liu, and An Li (2021). "Pan-sharpening based on panchromatic image spectral learning using WorldView-2". In: *IEEE Geoscience and Remote Sensing Letters* 19, pp. 1–5.

Xu, Hanqiu (2006). "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery". In: *International journal of remote sensing* 27.14, pp. 3025–3033.

Xu, Yongyang, Zhong Xie, Yaxing Feng, and Zhanlong Chen (2018). "Road extraction from high-resolution remote sensing imagery using deep learning". In: *Remote Sensing* 10.9, p. 1461.

Yang, Fan, Jianhua Guo, Hai Tan, and Jingxue Wang (2017). "Automated extraction of urban water bodies from ZY-3 multi-spectral imagery". In: *Water* 9.2, p. 144.

Yang, Xiaofei, Xutao Li, Yunming Ye, Raymond YK Lau, Xiaofeng Zhang, and Xiaohui Huang (2019). "Road detection and centerline extraction via deep recurrent convolutional neural network U-Net". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9, pp. 7209–7220.

Yang, Yixin, Shangzhen Song, Delian Liu, Jonathan Cheung-Wai Chan, Jinze Li, and Jianqi Zhang (2020). "Hyperspectral anomaly detection through sparse representation with tensor decomposition-based dictionary construction and adaptive weighting". In: *IEEE Access* 8, pp. 72121–72137.

Yao, Fangfang, Chao Wang, Di Dong, Jiancheng Luo, Zhanfeng Shen, and Kehan Yang (2015). "High-resolution mapping of urban surface water using ZY-3 multi-spectral imagery". In: *Remote Sensing* 7.9, pp. 12336–12355.

Yengoh, Genesis T, David Dent, Lennart Olsson, Anna E Tengberg, and Compton J Tucker III (2015). *Use of the Normalized Difference Vegetation Index (NDVI) to assess Land degradation at multiple scales: current status, future trends, and practical considerations*. Springer.

Yeung, Henry Wing Fung, Meng Zhou, Yuk Ying Chung, Grant Moule, Wayne Thompson, Wanli Ouyang, Weidong Cai, and Mohammed Bennamoun (2022). "Deep-learning-based solution for data deficient satellite image segmentation". In: *Expert Systems with Applications* 191, p. 116210.

Yu, Long, Zhiyin Wang, Shengwei Tian, Feiyue Ye, Jianli Ding, and Jun Kong (2017). "Convolutional neural networks for water body extraction from

Landsat imagery". In: *International Journal of Computational Intelligence and Applications* 16.01, p. 1750001.

Yuan, Kunhao, Xu Zhuang, Gerald Schaefer, Jianxin Feng, Lin Guan, and Hui Fang (2021). "Deep-learning-based multispectral satellite image segmentation for water body detection". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp. 7422–7434.

Zhang, Jing (2024). "Water Body Information Extraction from Remote Sensing Images based on PSPNet". In: *International Journal of Computer Science and Information Technology* 2.1, pp. 319–325.

Zhang, Liangpei, Lefei Zhang, and Bo Du (2016). "Deep learning for remote sensing data: A technical tutorial on the state of the art". In: *IEEE Geoscience and Remote sensing magazine* 4.2, pp. 22–40.

Zhang, Lili, Yushi Lu, Rui Shi, Yipin Shen, and Yehong Shao (2024). "Towards interpretability lightweight semantic segmentation model for waterbody extraction in large-scale high resolution remote sensing images". In: *International Journal of Remote Sensing* 45.8, pp. 2721–2738.

Zhang, Xinyu, Jinjiang Li, and Zhen Hua (2022). "MRSE-Net: multiscale residuals and SE-attention network for water body segmentation from satellite images". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15, pp. 5049–5064.

Zheng, Shuai, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr (2015). "Conditional random fields as recurrent neural networks". In: *Proceedings of the IEEE International Conference on computer vision*, pp. 1529–1537.

Zhou, Yan, Jinwei Dong, Xiangming Xiao, Tong Xiao, Zhiqi Yang, Guosong Zhao, Zhenhua Zou, and Yuanwei Qin (2017). "Open surface water mapping algorithms: A comparison of water-related spectral indices and sensors". In: *Water* 9.4, p. 256.

Zhu, Xiao Xiang, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer (2017). "Deep learning in remote sensing: A comprehensive review and list of resources". In: *IEEE geoscience and remote sensing magazine* 5.4, pp. 8–36.