Two-Phase Methods to Segment Man-made Objects Around Reservoirs

Nayereh Hamidishad

Thesis presented to the
Institute of Mathematics and Statistics
of the University of São Paulo
for the degree of
Doctor of Philosophy in
Computer Science

Program: Computer Science

Supervisor: Prof. Dr. Roberto Marcondes Cesar Junior

During the development of this work the author received financial assistance from CAPES
and
FAPESP #2015/22308-2

São Paulo
February 2023

Acknowledgments

I would like to express my appreciation to my supervisor, Professor Roberto Marcondes Cesar Junior, for generously providing guidance, knowledge, expertise, and support. I am grateful to my defense committee, Prof. Nina Sumiko Tomita Hirata, Prof. Nelson Delfino d'Ávila Mascarenhas, Prof. Luciano da Fontoura Costa, and Prof. Celina Maki Takemura. I am also thankful to Professor Jurandir Itizo Yanagihara from PME/Poli for working in collaboration and for permission to use his laboratory. I would like to thank CAPES and FAPESP for their support during my Ph.D.

Words can not express my gratitude to my Mother. I am deeply indebted to her and also to my father and sister, who I could not terminate this journey without them. Thanks and apology from the depth of my heart to my little son, Mostafa, for spending the hours that was owned to him on working on my thesis. Finally, my profound thanks go to my spouse, Ali, for his encouragement, understanding, and consistent support.

Resumo

Hamidishad, N. **Métodos de duas fases para segmentar áreas construídas ao redor de reservatórios**. Tese- Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

Os reservatórios são infraestruturas fundamentais para a gestão dos recursos hídricos. Eles reduzem os efeitos das flutuações de fluxo de água intersazonais e interanuais e, portanto, facilitam o abastecimento de água, a geração de energia hidrelétrica e o controle de enchentes, para citar alguns exemplos. Há uma interação significativa entre o meio ambiente e os reservatórios. Por exemplo, as atividades humanas podem afetar a qualidade da água afluente do reservatório e seus processos químicos e biológicos. As construções ao redor dos reservatórios são um exemplo de tais atividades. Essa questão social pode ser detectada segmentando os objetos criados pelo homem em torno dos reservatórios nas imagens de sensoriamento remoto (RS). Os métodos tradicionais baseados em pixels, baseados em objetos (OB) e de aprendizado profundo são três abordagens de mapeamento de cobertura da terra (LCM).

Desenvolvemos uma nova abordagem baseada em técnicas de processamento de imagens e no método OB para segmentar regiões selecionadas ao redor de reservatórios. As desvantagens da abordagem OB, como a alta dependência dos resultados na escolha dos parâmetros, nos levaram a usar DL para contornar a necessidade excessiva de especificação de parâmetros e o ajuste frequentemente exigido pelos métodos OB.

Nos últimos anos, o DL atraiu considerável atenção como um método para segmentar imagens das imagens do RS e alcançou um sucesso notável. Para segmentar objetos artificiais em torno dos reservatórios utilizando um fluxo de trabalho de ponta a ponta, segmentar os reservatórios e destacar a região de interesse (ROI) em torno deles é essencial. No entanto, os reservatórios são normalmente considerados em uma classe ampla denominada corpos d'água. Além disso, os estudos de segmentação de objetos feitos pelo homem implementados em imagens RS de alta resolução urbana consideram menos frequentemente construções na zona rural. Portanto, eles frequentemente não consideram estruturas desafiadoras de cobertura do solo, como estradas não asfaltadas.

Nesta pesquisa, desenvolvemos uma nova abordagem baseada em técnicas de processamento de DL e imagens para segmentação de objetos feitos pelo homem em torno dos reservatórios. No fluxo de trabalho em duas fases proposto, o reservatório é inicialmente segmentado usando um modelo DL. Em seguida, uma etapa de pós-processamento é proposta para remover

erros da vegetação flutuante nos reservatórios. Em seguida, a RoI ao redor do reservatório (RoIaR) é destacada usando as técnicas de processamento de imagem. Finalmente, os objetos artificiais na roiar são segmentados por um modelo DL. Coletamos imagens de alta resolução do Google Earth (GE) de oito reservatórios no Brasil, localizados principalmente nas paisagens, em dois anos disponíveis para treinar os modelos de fluxo de trabalho. Além disso, validamos o fluxo de trabalho preparado com um conjunto de dados de teste não visto durante o treinamento. As pontuações F1 do estágio de segmentação semântica da fase 1, estágio de pós-processamento e estágio de segmentação semântica da fase 2 no conjunto de teste externo são 92,54%, 94,68% e 88,11%, respectivamente, que mostram alta capacidade de generalização do fluxo de trabalho preparado.

palavras-chave: imagem RS de alta resolução; segmentação de reservatórios; segmentação feita pelo homem; processamento de imagem; aprendizagem profunda.

Abstract

Hamidishad, N. **Two-Phase Methods to Segment Man-made Objects Around Reservoirs**. Thesis (Ph.D.)- Institute of Mathematics and Statistics, Universidade de São Paulo, São Paulo, 2023.

Reservoirs are fundamental infrastructures for the management of water resources. They reduce the effects of interseasonal and interannual streamflow fluctuations and hence facilitate water supply, hydroelectric power generation, and flood control, to name a few. There is a significant interaction between the environment and reservoirs. For example, reservoirs affect the quality of the water downstream of their dams, and human activities affect the quality of the reservoir's inflowing water and its chemical and biological processes.

Construction around reservoirs is a human activity that can negatively impact the reservoirs' water quality. This social issue can be detected by segmenting the man-made objects around reservoirs in the Remote Sensing (RS) images. Traditional pixel-based, Object-Based (OB), and Deep-Learning (DL) methods are three Land-Cover Mapping (LCM) approaches.

We developed a new approach based on image processing techniques and the OB method for LCM of the selected regions around reservoirs. Disadvantages of the OB approach, such as the high dependency of results on the choice of parameters, led us to use DL to circumvent excessive parameter specification and tunning that are often required by OB methods.

In recent years, DL has attracted considerable attention as a method for segmenting the RS imagery semantically and has achieved remarkable success. To segment man-made objects around the reservoirs utilizing an end-to-end workflow, segmenting reservoirs and detaching the Region of Interest (RoI) around them are essential. However, reservoirs are always considered in a broad class termed water bodies in RS semantic segmentation studies. Besides, man-made object semantic segmentation in the RoIaR is not explored in the literature. Moreover, man-made object segmentation in high-resolution images, especially countryside man-made object segmentation, is not extensively explored in the literature.

In this research, we develop a new approach based on DL and image processing techniques for man-made object segmentation around the reservoirs. In the proposed two-phase workflow, the reservoir is initially segmented using a DL model. Then, a post-processing stage is proposed to remove errors such as floating vegetation. Next, the RoI around the Reservoir

(RoIaR) is detached using the proposed image processing techniques. Finally, the man-made objects in the RoIaR are segmented by a DL model. We collected high-resolution Google Earth (GE) images of eight reservoirs in Brazil, mainly located in the countrysides, over two different available years to train the workflow models. Furthermore, we validated the prepared workflow with a test dataset not seen during training. The F1-scores of the phase-1 semantic segmentation stage, post-processing stage, and phase-2 semantic segmentation stage on the external test set are 92.54%, 94.68%, and 88.11%, respectively, which show high generalization ability of the prepared workflow.

Keywords: high-resolution RS imagery; reservoir segmentation; man-made object segmentation; image processing; object-based method; deep learning.

List of Figures

1.1 An RS image (left) and generated land cover map (right) by Geng et al. [2017]. 2
1.2 Construction around reservoirs can be considered a social problem with potentially serious consequences to society and the reservoir's water quality (a) Example of Brazilian news reporting unauthorized constructions around an important reservoir ; (b) Photo from a Brazilian news agency illustrating unauthorized constructions around a reservoir. 3
1.3 Remote sensing image showing man-made objects around Guarapiranga reservoir. 4

2.1 Overview of land cover map generation using OB approach. 7

3.1 An example of an RS image and its segmentation and classification outputs produced by the OB approach Duro et al. [2012a] 13
3.2 Flowchart of the proposed approach. 14
3.3 Visual pipeline of the proposed tow-phase OB classification approach. 15
3.4 Proposed approach for data collecting. 17
3.5 Example of a GE image and two segmentation issues. The rooftops in over-segmentation are covered by several segments. Whereas in under-segmentation, the segments over rooftops (e.g., the segment highlighted with a red border) contain other scene classes as well. 19
3.6 Classification comparison for three land cover types. 21
3.7 Example of a test set image. This image is collected from zone-1 in 2017. . . 23
3.8 Result of implementing multi-resolution segmentation on Figure 3.7. 24
3.9 The output of applying spectral difference segmentation on Figure 3.8. Applying this segmentation step reduced The number of generated segments over the reservoirs significantly. 24
3.10 produced image after classifying reservoir in Figure 3.7. 25
3.11 The output of implementing spectral difference segmentation on segments in Figure 3.9 that are unclassified in Figure 3.10. 25
3.12 Figure 3.7 final land cover map. Red: man-made, green: vegetation, black: shadow, blue: reservoir. 26
3.13 A test set image collected from Zone-1 in 2009. 27

3.14 Figure 3.13 land cover map. Red: man-made, green: vegetation, black: shadow, blue: reservoir. 27
3.15 Two test set images from different locations and their LCM. (a) and (b) are the GE image collected from Zone-2 in 2009 and Zone-3 in 2017, respectively; (c) OB LCM of (a); (d) OB LCM of (b). 28
3.16 Segmentation comparison for images of the same region in 2009 and 2017. The red and green pixels represent man-made and vegetation classes, respectively. In this figure the environmental degradation around the reservoir is depicted. 29
3.17 Average percentage of different land cover types in three studied zones. . . . 29

4.1 Overview of the proposed analysis workflow. 37
4.2 Visualization of the studied reservoirs locations. 39
4.3 Proposed data preparation and annotation pipeline. 40
4.4 The U-Net architecture. 41
4.5 The SegNet architecture. 41
4.6 An overview of PSPNet. The size of feature map channels is denoted below each box. The size of the last feature map in (a) is 1/8 of the input image size. 43
4.7 An overview of FPN. 44
4.8 An overview of LinkNet architecture. 44
4.9 Jupyter Notebook (Google Colaboratory) created to prepare the proposed method. 47
4.10 Examples of the test set patches beside their corresponding ground truths and segmentation outputs. 49
4.11 Examples of non-reservoir water bodies in the collected dataset and corresponding segmentation results. 50
4.12 Two examples of produced mosaic images, corresponding ground truths, prediction outputs, and post-processing outputs. Anta-2014 with 11687 × 14430 pixel size, and Nova-2021 with 24830 × 23193 pixel size are depicted in the first and second columns, respectively. 51
4.13 Samples of four studied reservoirs RoI patches beside their corresponding ground truths and prediction outputs. 53
4.14 Two samples of Barra phase-1 patches, besides their corresponding ground truths and semantic segmentation results. 54
4.15 Two samples of Barra RoI patches, their corresponding ground truths, and semantic segmentation results. 55

A.1 62

B.1 64

List of Tables

3.1 The geographical coordinates of the studied zones. 16
3.2 Proportions of "from vegetation to man-made" and "from shadow to man-made" from 2009 to 2017 in three studied zones. 28
3.3 Confusion matrix when there are k classes. 30
3.4 Accuracy assessment of Zone-1 calibration set images land cover maps. . . . 31
3.5 Accuracy assessment of Zone-1 test images land cover maps. 31
3.6 Accuracy assessment of Zone-2 calibration set images land cover maps. . . . 31
3.7 Accuracy assessment of Zone-2 test images land cover maps 32
3.8 Accuracy assessment of Zone-3 calibration set images land cover maps 32
3.9 Accuracy assessment of Zone-3 test images land cover maps 32

4.1 Locations of studied reservoirs. 38
4.2 Acquisition years of each reservoir dataset. Some of the older year images of Luiz and Nova belong to 2004 and 2010, respectively. 38
4.3 Performance of trained architectures for phase-1 semantic segmentation stage. 46
4.4 The performance of models with healthy learning curves for Phase-1 semantic segmentation on the validation set. 47
4.5 SegNet_d performance on the test set. 48
4.6 Prediction and refinement performance metrics for Anta-2014 an Nova-2021. 48
4.7 The highest achieved performances using trained models for Phase-2 semantic segmentation on train and test sets. 52
4.8 FPN performance in segmenting test set to man-made, and non-man-made. 52
4.9 FPN performance in segmenting countryside and urban man-made objects. C and U are the abbreviations for countryside and urban. 52
4.10 Performance of phase-1 semantic segmentation and post processing stages in segmenting Barra dataset to reservoir and non-reservoir. 53
4.11 Performance of phase-2 semantic segmentation stage in segmenting Barra RoI to man-made and non-man-made. 54
4.12 U-Net performance in segmenting train and test sets into reservoir, man-made, and else. 55

Contents

List of Figures ix

List of Tables xi

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Goal . . . 5
- 1.3 Contributions . . . 5
- 1.4 Text Organization . . . 5

2 Bibliographical review 7
- 2.1 Object-Based Image Analysis . . . 7
- 2.2 Deep Learning . . . 9
- 2.3 DL-Based RS Segmentation . . . 11

3 Land Cover Mapping Using OB Method 13
- 3.1 Introduction . . . 13
- 3.2 Proposed Approach . . . 14
  - 3.2.1 Data Formation . . . 15
  - 3.2.2 Two-phase OB Classification . . . 16
  - 3.2.3 Data Analysis: Post-classification comparison change detection . . . 21
- 3.3 Experimental Results and Accuracy Assessment . . . 22
  - 3.3.1 Implementation Details . . . 22
  - 3.3.2 Experimental Results . . . 23
  - 3.3.3 Workflow Validation . . . 30
- 3.4 Conclusion . . . 31

4 A DL-based approach for man-made object segmentation around reservoirs 35
- 4.1 Introduction . . . 35
- 4.2 Materials and Methods . . . 36
  - 4.2.1 Data Collection . . . 36
  - 4.2.2 Data Preparation and Annotation . . . 36
  - 4.2.3 Phase-1: Reservoir segmentation . . . 38

4.2.4 Phase-2: Man-made object segmentation 42
4.3 Experimental Results 45
4.3.1 Performance Evaluation Metrics 45
4.3.2 Phase-1 Experimental Results 46
4.3.3 Phase-2 Experimental Results 48
4.3.4 Workflow evaluation 52
4.3.5 Benchmark 54
4.4 Discussion 55

5 Concluding remarks **57**
5.1 Conclusion 57
5.2 Future work 58

A Paper: Object-based Method for Identifying New Constructions around Water Reservoirs: Preliminary Results **61**

B Paper: An End-to-End Two-Phase Deep Learning-Based Workflow to Segment Man-made Objects Around Reservoirs **63**

Bibliography **65**

Chapter 1

Introduction

1.1 Motivation

Earth Remote Sensing is a technology that involves remote imaging systems (e.g., aircraft, satellites, unmanned aerial vehicles) to obtain information from a target on the earth through the analysis of the acquired data. The sensors mounted on remote imaging systems measure the electromagnetic radiation reflected or emitted by the target. The characteristics of the sensor and its platform determine the type of information that may be accessed by remote sensing technologies Bhandari et al. [2012], Rozanda et al. [2015]. There are two overall categories for remote sensing systems, named passive and active. In the passive mode, the sensor uses the naturally available energy (sunlight) for image capturing, and in the active mode, the system sends microwaves toward the surface and waits for the reflected waves.

The first aerial photo was acquired in 1858 from a balloon in France by Gaspard-Felix Tournachon. In the next years, numerous improvements were made in photographic technology. The first aerial photograph taken from an aircraft belongs to 1909 from an Italian landscape. In 1972, the first satellite designed to observe earth's land areas was launched. The images had four spectral bands, each 80m pixel size. Numerous improvements followed from these beginnings, and now a day, satellites, airplanes, or UAVs are utilized for acquiring RS images with different spatial, spectral, and temporal resolutions Campbell and Wynne [2011]. For example, some satellite sensors, such as WordView-3, collect multi-spectral imagery (images with at least three spectral bands) at 1.84 m spatial resolution. WordView-1 and GeoEye-1 collect panchromatic imagery at 0.5 m, 0.46 m, and 0.41 m spatial resolutions, respectively. Panchromatic (or black-and-white) images have a single band that combines the information from different bands (e.g., green and red) instead of partitioning it into different spectral bands. Moreover, some sensors, such as AVIRIS, produce images with hundreds of bands that are called hyper-spectral images Liau [2014], Jensen [2015], Zhu et al. [2018].

There are different platforms for downloading RS images. Bing[1], NASA World Wind[2], and Google Earth[3] (GE) are free widely used programs among them Qi et al. [2016], Zheng et al. [2016], Abdelrahim and Mansour [2017], Huang et al. [2018]. The increasing availability of data sources helps the development of research for new RS techniques and applications.

RS technology is important due to its spatially-explicit representation of the earth's surface, frequent temporal coverage, and relatively low observation costs. The remote sensing data has many application areas such as environmental assessment and monitoring, change detection, exploitation of renewable and non-renewable natural resources, mapping, and military surveillance and reconnaissance Schowengerdt [2006]. One of the most important applications of RS images is land cover/use mapping. An example of an RS image and corresponding land cover map is illustrated in Figure 1.1.

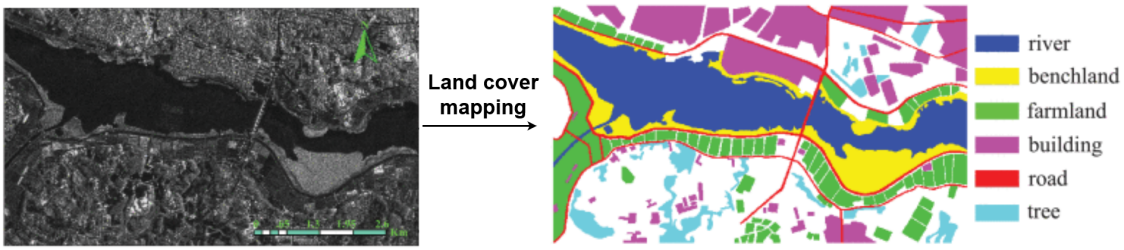

Figure 1.1: *An RS image (left) and generated land cover map (right) by Geng et al. [2017].*

A relevant application of RS segmentation techniques is monitoring the construction around water reservoirs (man-made lakes). Reservoirs reduce the effects of interseasonal and interannual streamflow fluctuations and hence facilitate water supply, hydroelectric power generation, and flood control, to name a few Gao et al. [2012]. There is a significant interaction between the environment and reservoirs as essential water resource management tools. For example, reservoirs affect the quality of the water downstream of their dams, and human activities affect the quality of the reservoir's inflowing water and the chemical and biological processes in it Votruba and Broža [1989]. The construction around reservoirs has a serious negative impact on the quality of reservoirs' water and can be considered as a social problem as illustrated in Figure 1.2 [4]. Monitoring the existence of such constructions can be modeled by segmenting RS images (see Figure 1.3).

The traditional pixel-based (called pixel-based) methods (e.g., SVM) rely on the spectral properties of individual pixels, and each pixel is independently classified. However, a single pixel does not capture the characteristics of targeted objects well when the spatial resolution of satellite images improves. It causes a reduction in the accuracy of classification using pixel-based methods Hu et al. [2013a]. Over the last twenty years, the RS community has

[1]https://www.bing.com/maps/
[2]https://worldwind.arc.nasa.gov/
[3]https://www.google.com/earth/desktop/
[4]https://g1.globo.com/sp/sao-paulo/noticia/2021/07/20/governo-estadual-e-prefeitura-de-sp-fazem-operacao-para-remover-construcoes-irregulares-as-margens-da-represa-billings.ghtml and https://agenciabrasil.ebc.com.br/foto/2018-02/construcoes-margens-da-represa-billings-em-sao-paulo-1581289720-0

(a)

(b)

Figure 1.2: *Construction around reservoirs can be considered a social problem with potentially serious consequences to society and the reservoir's water quality (a) Example of Brazilian news reporting unauthorized constructions around an important reservoir ; (b) Photo from a Brazilian news agency illustrating unauthorized constructions around a reservoir.*

undertaken considerable efforts to promote the use of OB technology for LCM Blaschke et al. [2014a], Qu et al. [2021]. The OB classification approach is less sensitive to the spectral variance within the objects despite the pixel-based methods. Since it first segments images and then uses both segments' features and spatial relations between them to classify them into different land covers.

One of the most commonly used and popular algorithms in the OB approach to segment the images is multi-resolution segmentation Hossain and Chen [2019]. However, this method requires users to determine a set of parameters. Furthermore, multi-resolution segmentation

Figure 1.3: *Remote sensing image showing man-made objects around Guarapiranga reservoir.*

parameters can differ between different RS images and even between different objects in an image. Many studies believe that to achieve high accurate segments over land cover classes of different sizes, multi-scale segmentation (i.e., defining one set of segmentation parameters for each class in the scene) is necessary Drăguţ et al. [2014], Johnson and Jozdani [2018].

Whereas the OB approach almost always outperforms the pixel-based approach Silveira et al. [2019], Wu et al. [2023], the DL approach is a more sophisticated approach that has made significant strides in recent years. It enables high-level feature extraction to be carried out automatically while displaying promising results in various domains, including image semantic segmentation. Recently, convolutional neural networks (CNNs) have been among the most advanced algorithms for the semantic segmentation of RS images. Their superior performance compared to traditional methods has been proved Ghanbari et al. [2021], Wurm et al. [2021], Malerba et al. [2021]. The encoder-decoder and pyramid pooling-based networks can be counted as two state-of-the-art and widely used categories of CNNs.

To the best of our knowledge, there is no study on the semantic segmentation of reservoirs in high-resolution RS images using DL models. This class is always considered in a broad class termed water bodies. Furthermore, the RoIaR man-made object segmentation has not been explored. Moreover, man-made object segmentation in urban high-resolution RS images, especially countryside man-made object segmentation, has not been extensively explored in the literature.

It is worth noting that although elevation data can improve the detection process, they are not currently viewed as a cost-effective solution to map RS images Vakalopoulou et al. [2015]. On the other hand, spatial resolution is more critical than spectral resolution in urban

LCM Neupane et al. [2021]. Therefore, our experiments are on images collected from the GE platform, which is a widely used database Hou et al. [2019], Zhang et al. [2020]. This platform gives us access to free high-resolution RS images from target reservoirs at various times. GE covers more than 25 percent of the Earth's land surface and three-quarters of the global population Jensen [2015], Jacobson et al. [2015]. Furthermore, the appearance of GE images is improved using color balancing, warping, and mosaic processing Almeer [2012]. Therefore, it can be used for studying many other reservoirs.

1.2 Goal

This thesis aims to develop new methods for segmenting man-made object around reservoirs in high resolution RS images utilizing highest performance approaches.

1.3 Contributions

The main contributions of this thesis are:

- Development and evaluation of an OB approach for land cover mapping RS images. The results are described in Hamidishad and Marcondes Cesar Jr [2019].
- Proposition of a multi-phase segmentation for increasing the accuracy of detected segments in the segmentation step of the OB approach
- Development of an end-to-end workflow for man-made object segmentation around reservoirs. The proposed approach and results are described in Hamidishad and Marcondes Cesar Jr [2023].
- Preparation and evaluation of two DL models for segmenting reservoirs and man-made objects from urban and countryside areas in RS images with different spatial resolutions, textures, and temporal conditions.
- Proposing a post-processing approach to increase the accuracy of generated reservoir maps.

1.4 Text Organization

The organization of this thesis is as follows. In Chapter 2, we discuss the bibliography. The proposed OB and multi-phase segmentation approaches, utilized data, and corresponding results are presented in Chapter 3. The proposed two-phase DL-based workflow, the data collection, and experimental results are explored in Chapter 4. Finally, we discuss the results and findings of the study and propose some future works in Chapter 5.

Chapter 2

Bibliographical review

Generally, mapping RS images is a complicate process in which many factors influence the quality of final products. This chapter presents a bibliography review of two widely used and popular approaches for RS imagery mapping.

2.1 Object-Based Image Analysis

In the past, most digital image classification methods were based on processing the entire scene pixel by pixel, commonly called pixel-based classification Blaschke [2001], Myint et al. [2011]. Nonetheless, by improving the spatial resolution of RS images, the image analysis paradigm moved from pixel-based to OB. This method can increase the classification accuracy compared to pixel-based image classifications that ignore spatial coherence among groups of pixels Yu et al. [2006], Wang et al. [2007], Zhou et al. [2015], Qu et al. [2021].

the OB approach consists of two steps; see Figure 2.1. The segmentation step allows decomposing the scene into many relatively homogeneous image segments (referred to patches, objects, polygons, or super-pixels) using an image segmentation process. Segments in this context are groups of pixels that share some common properties (e.g. a connected component that belongs to a water body or a roof-top). Then, various statistical characteristics of the created homogeneous image segments are used by different classification methods for classifying these segments Blaschke et al. [2014b].

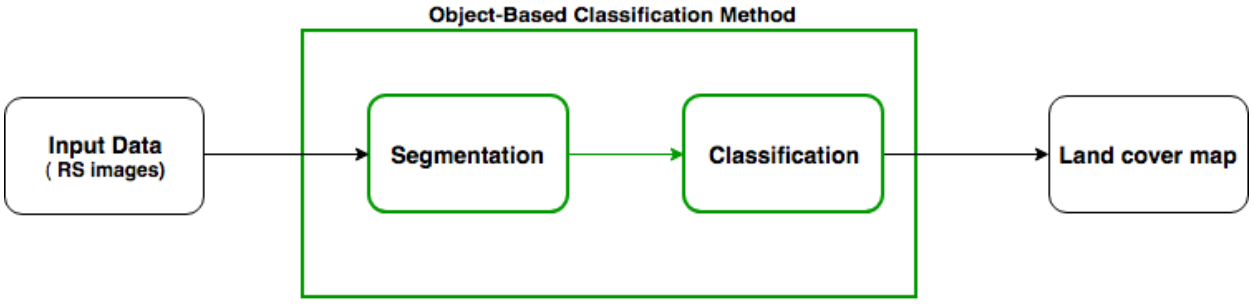

Figure 2.1: *Overview of land cover map generation using OB approach.*

An important image segmentation aspect regards the possible varying scale of the objects of interest. The multi-resolution algorithm has been developed to address this issue. This method, proposed by Baatz [2000], is one of the most commonly used algorithms for segmenting RS images Guirado et al. [2017], Liu et al. [2017], Zhang et al. [2018a], Johnson and Jozdani [2018] and was first implemented in eCognition software.

Assessing the quality of segmentation and identifying appropriate segmentation parameters can be performed using a supervised approach, unsupervised approach, or visually. The main goal in a supervised method is applying several combinations of parameters and then quantifying dissimilarity between reference segments (digitized manually) and generated segments. For example, Tong et al. [2012] presents a supervised approach to determine optimal image segmentation parameters for the multi-resolution algorithm in eCognition software. A complete description of this optimizer can be found in Zhang et al. [2010].

Despite the effectiveness of supervised methods, they are time-consuming and dependent on the applied range of parameter settings Witharana and Civco [2014]. Unsupervised methods do not rely on expert knowledge. These methods quantify the inter-segment homogeneity and intra-segment heterogeneity at several scales Jozdani et al. [2018]. One of the most well-known unsupervised methods is proposed by Drăguţ et al. [2010], which is improved and expanded by Drăguţ et al. [2014].

The work Grybas et al. [2017] tests several unsupervised segmentation optimization procedures qualitatively and demonstrates that optimal parameters for a particular scene do not necessarily equate to optimal parameters for the chosen classification scheme. Furthermore, the authors of Belgiu and Drăguţ [2014] compare supervised and unsupervised approaches for selecting multi-resolution parameters. They show that the unsupervised method produces very different image segments from the supervised method. Nonetheless, surprisingly, the classification accuracy was still very similar. Although the over-segmentation (where the created segments are smaller than the optimal segments) is always preferred to the under-segmentation (where the generated segments are larger than the optimal segments), they show that as long as under-segmentation is acceptable, high classification accuracy is achievable.

As mentioned before, the quality of generated segments can be assessed visually. This method have been utilized in several studies Gao et al. [2011], Duro et al. [2012b], Hu et al. [2013a], Liu et al. [2017], Johnson and Jozdani [2018], Zaabar et al. [2022].

The second step in the OB approach is the classification of generated segments. This step can be implemented using unsupervised, supervised, semi-supervised, weakly-supervised, or rule-based methods. Unsupervised methods, also called clustering methods, group segments based on their similarity of feature values. When these clusters are obtained, the reference data is used to classify them. When there are more clusters than the land cover classes, these methods group clusters with no reference data with the closest classified clusters.

One way to ensure that all groups of the land covers are taken into account by the

classification algorithm is using the knowledge in assigning each segment to a given class during the learning process. This method is called the supervised method. In supervised methods, we directly construct the decision function that separates segments and associates them with their corresponding classes. Because of this reason, supervised methods are more efficient than unsupervised methods. However, they require more reference data Baghdadi and Zribi [2016]. The semi-supervised approach is implemented when the dataset is partially annotated, and the weakly-supervised approach is implemented when the images are incompletely annotated.

The rule-based method is based on user-defined rules. A rule describes intervals of a feature or features that determine whether the constructed segments in the segmentation step belong to a particular class. Since the analyst must create the rules, the major challenge in this method is developing rule sets to map relevant features. This method is implemented in different studies with different names, such as rule-based Ziaei et al. [2014], Jensen [2015], Ma et al. [2017], visual interpretation Liu et al. [2017], Membership function classifier Myint et al. [2011], and knowledge-based Chen et al. [2018a].

The application of the OB approach in mapping RS images is studied in many works. For example, the Landsat-8 images of the Yangtze River Delta in China are mapped by Qu et al. [2021] using OB and pixel-based approaches. Simple non-iterative clustering is used to segment images, and the random forest is applied to classify the segments. In their experiments, the OB method outperformed the pixel-based method, where the random forest is used as the classifier. An object-based convolutional neural network approach is proposed by Lv et al. [2021] for LCM images collected using the GE platform from an area in California. In the proposed approach, every image is segmented into several zones (residential, natural, and industrial zones), and each zone is then segmented using the multi-resolution method. A binary tree sampling is proposed to generate appropriate convolutional windows for the segments. These windows are semantically segmented using an already trained CNN model, and then all pixels in each segment are labeled with the most frequent label in that segment. The authors of Rittenhouse et al. [2022] applied the simple non-iterative clustering method to produce the segments based on the similarity of spectral reflectance, spectral texture, vegetation height, and texture of vegetation height of each pixel. They proposed a ruleset for classifying the segments.

2.2 Deep Learning

DL is a machine learning field that explores Neural Networks to learn from data by successive layers. The idea is that such (possible large number of) layers may extract meaningful input representations. The term "deep" in deep learning (DL) refers to these subsequent layers of representation. Convolutional networks LeCun et al. [1989] are among the most popular DL networks that have quickly found their place within computer vision

tasks, including RS LCM.

CNNs are mainly composed of three operation layers termed convolution, pooling, and nonlinear activation function. Krizhevsky et al. [2012] proposed a pioneer large CNN, AlexNet, to classify the ImageNet dataset that achieved the best results of the year. In a successful process, a CNN model called VGGNet was introduced by Simonyan and Zisserman [2014]. VGGNet architectures consist of five convolutional blocks, each followed by a max-pooling layer, and three Fully-Connected (FC) blocks. The VGG-16 (containing 16 convolutional layers) won the ImageNet challenge in 2013 and is this category's most frequently used model.

Although increasing the network's depth in some cases is proved to be crucial Szegedy et al. [2015], it causes the vanishing/exploding gradients. Residual learning frameworks fixed this problem. The shortcut connection in these networks, which passes the input tensor of each block to its output, is the key insight of this approach. After VGG-16, two proposed residual networks by He et al. [2016] called ResNet-50 and ResNet-101 are the most utilized CNNs Ghanbari et al. [2021].

The results of the proposed architectures were encouraging. Nevertheless, the high-resolution reduction in the input of models caused by several pooling layers was a severe drawback for the pixel-wise classification tasks. Therefore, the encoder-decoder paradigm has been introduced in the semantic segmentation field. The fully convolutional network proposed by Long et al. [2015] achieved the best results in the PASCAL VOC visual recognition competition. The decoder-encoder-based networks consist of an encoder path and a decoder path. The encoder path consists of convolutional layers to extract the feature maps. Next, these features are transformed/up-sampled to dense label maps in the decoder path.

A state-of-the-art architecture of this category is named SegNet and was proposed firstly by Badrinarayanan et al. [2015]. Each decoder in this architecture up-samples the output of the corresponding encoder using the already stored pooled indices. Despite the SegNet that was primarily motivated by scene understanding, the U-Net architecture proposed by Ronneberger et al. [2015] was first applied to biomedical segmentation. However, it was later successful in other fields as well Emek and Demir [2020], Kashani et al. [2019]. The proposed concatenation between cropped feature map of each encoder level and the corresponding decoder improves the segmentation using the U-Net architecture. Although it is reported that U-Net needs a smaller dataset to achieve satisfying results, it is at a higher memory cost than SegNet. U-Net, SegNet, and FPN Lin et al. [2017a] are the encoder-decoder-based architectures that have demonstrated solid performances and are frequently utilized in RS semantic segmentation Neupane et al. [2021], Bischke et al. [2019].

The spatial pyramid pooling-based networks are also CNN-based state-of-the-art networks. These networks contain pyramid pooling modules to collect multi-level global information of the input. PSPNet proposed by Zhao et al. [2017] is a widely used architecture of this category Zhang et al. [2020], Yuan et al. [2021] that won the ImageNet scene parsing challenge

in 2016.

CNNs are employed for a variety of tasks. Such as image classification Krizhevsky et al. [2012], which in each image is labeled by a class; object detection Galvez et al. [2018] that involves localizing certain objects in the images with bounding boxes accurately and efficiently; speech processing that converts human speech into spoken words Deng et al. [2013], semantic segmentation Pham [2021], Etc. The widely used application of CNNs in the RS community is semantic segmenting RS images into different land covers or/and land uses. A bibliography of this application is presented in the next section.

2.3 DL-Based RS Segmentation

CNNs are implemented in various RS semantic segmentation tasks and have achieved successful results. For example, the authors of Abdollahi et al. [2020] implemented a composition of SegNet and U-Net models to segment buildings in high-resolution aerial images. Water detection in high-resolution RGB images collected from Australia is performed using a CNN model by Malerba et al. [2021]. CNNs are even used to assess the quality of OpenStreetMap data Xie et al. [2019]. In this paper, OpenStreetMap building footprints are assessed using the results extracted from the high-resolution RS images using a U-Net-based model.

Semantic segmentation of water bodies is studied in several works. For example, Song et al. [2020] has fused panchromatic and RGB images and used a DL-based workflow for water body recognition. A DL encoder-decoder framework is proposed by Li et al. [2021] to extract water bodies from 4-band RS images with resolutions greater than one meter. The authors of Chen et al. [2018b] combine an enhanced super-pixel method with DL to extract urban water bodies from multi-spectral bands with low spatial resolutions (>4m). The RapidEye 5m resolution images are used by Zhang et al. [2018b] to compare pixel-based methods with DL methods in segmenting gorges reservoir areas to water bodies and other land covers. The capacity of NDWI and NDSWI indices in mapping water surfaces in 4-band (near-infrared, Red, Green, and Blue) high-resolution RS images using DL and ML methods is studied by Aryal et al. [2021]. The utilized data belongs to National Oceanic and Atmospheric Administration (NOAA), which covers areas inside the USA. The authors of Van Soesbergen et al. [2022] proposed a pipeline where a DL model in the first stage segments the water bodies in moderate spatial resolution RS images. Next, bounding boxes of individual water bodies are classified into two classes, dam reservoir and natural water, by a classifier.

DL is also popular among studies on semantic segmentation of man-made objects in RS images. For example, a DL-based approach is proposed by Makantasis et al. [2015] to classify ROSIS hyper-spectral images as man-made and non. The man-made class in this work consists of asphalt, metal sheets, bricks, bitumen, and tiles. Before feeding to the

network, the data is pre-processed by Randomized Principal Component Analysis for input dimension reduction. The authors of Papadomanolaki et al. [2019] have proposed an OB-DL framework to semantic segment two publicly available ISPRS datasets. These datasets are annotated to the impervious surface, building, low vegetation, tree, car, and clutter. Two extension versions of U-Net are proposed by Abdollahi et al. [2021] to segment buildings and roads in RGB RS images with 0.5m resolution. In the proposed workflow, each class is trained in a distinct network because of the type of available ground truths. Residential land, industrial land, traffic land, woodland, and unused land are five defined classes by Yu et al. [2022] for collected RGB images with 0.5 m resolution. They proposed a workflow to segment images in which the images are fed to two networks in parallel. Next, their output feature maps are fused to produce the final map. The built infrastructures in two sites on the North Slope of Alaska are mapped by Manos et al. [2022] by applying a DL model on 4-band commercial satellite images with resolutions from 0.5 up to 0.87. The utilized model in this work is the U-Net with ResNet50 as the backbone.

Chapter 3

Land Cover Mapping Using OB Method

The content described in this chapter is based on the publication Hamidishad and Marcondes Cesar Jr [2019].

3.1 Introduction

The OB land cover mapping consists of two steps, segmentation and classification (see Figure 3.1). Segments extracted in the segmentation step are regions generated by one or more homogeneity criteria. Thus, segments have additional spectral information compared to single pixels, such as mean values per band, median values, minimum and maximum values, and variance, to name but a few. The more significant advantage of classifying segments instead of pixels is the additional spatial information of segments Hay and Castilla [2008], Van der Werff and Van der Meer [2008]. However, The additional spatial features (such as distances, neighborhood, Etc) are the main reasons for the increase in using this method Blaschke [2010]. An example of this approach on an RS image is illustrated in Figure 3.1.

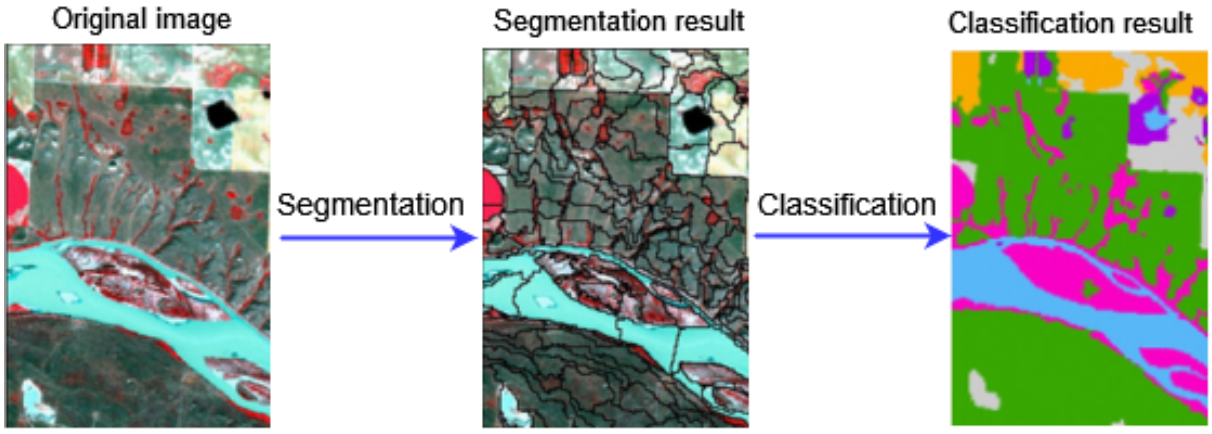

Figure 3.1: *An example of an RS image and its segmentation and classification outputs produced by the OB approach Duro et al. [2012a]*

Multi-resolution segmentation and rule-based classification are among the most commonly used algorithms for segmenting and classifying RS images. However, the multi-resolution

segmentation method requires users to determine a set of parameters and is significantly slower than some other segmentation techniques. Furthermore, visual assessment is the most common method employed for assessing these parameters' values, making the process very empirical and heuristic Zhang et al. [2008]. Moreover, the multi-resolution segmentation parameters can differ between different RS images and even between different objects in an image. Therefore, determining and assessing these parameters are challenging and time-consuming.

Our research developed a new OB approach based on two phases. Phase 1 aims at detecting the reservoir objects in the images. Then, phase 2 classifies the non-reservoir image segments into vegetation, man-made object, and shadows. This chapter is organized as follows: the proposed method is described in section 3.2. In section 3.3, implementation details, experimental results, and results validation are depicted. Finally, the chapter is concluded in section 3.4.

3.2 Proposed Approach

The framework of the proposed procedure is illustrated in Figure 3.2. The framework is composed of three parts: data formation, the proposed two-phase OB classification, and a data analysis application. The core idea is the two-phase OB classification that is visually illustrated in Figure 3.3.

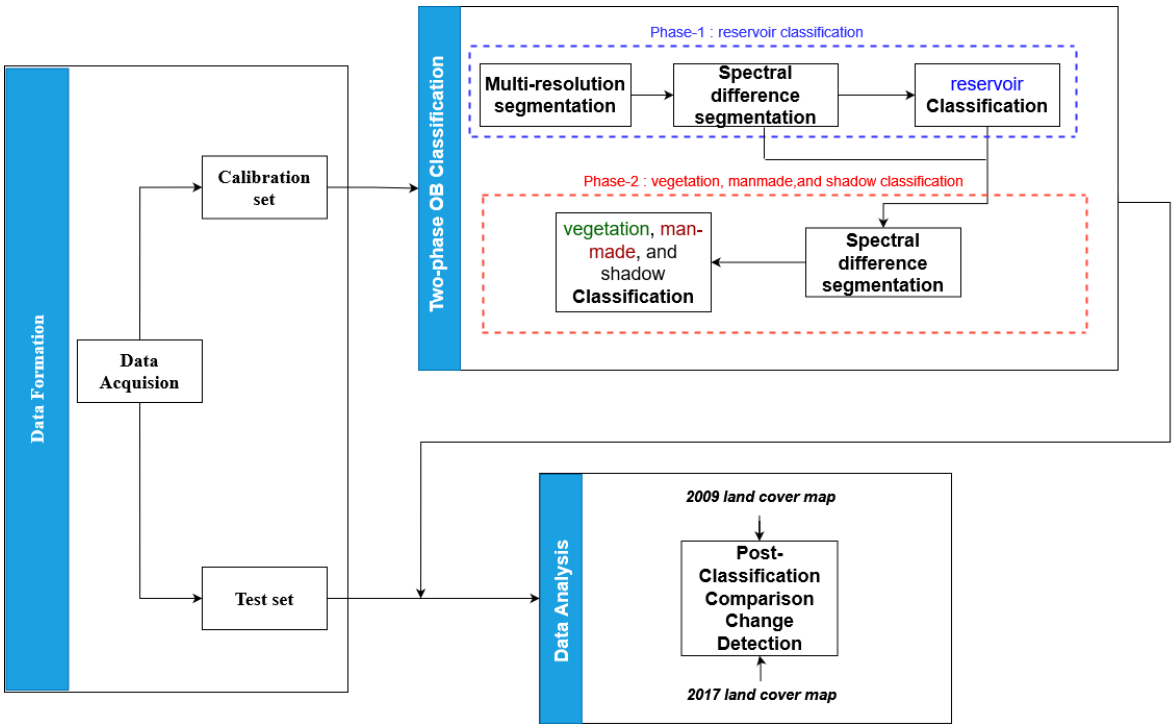

Figure 3.2: *Flowchart of the proposed approach.*

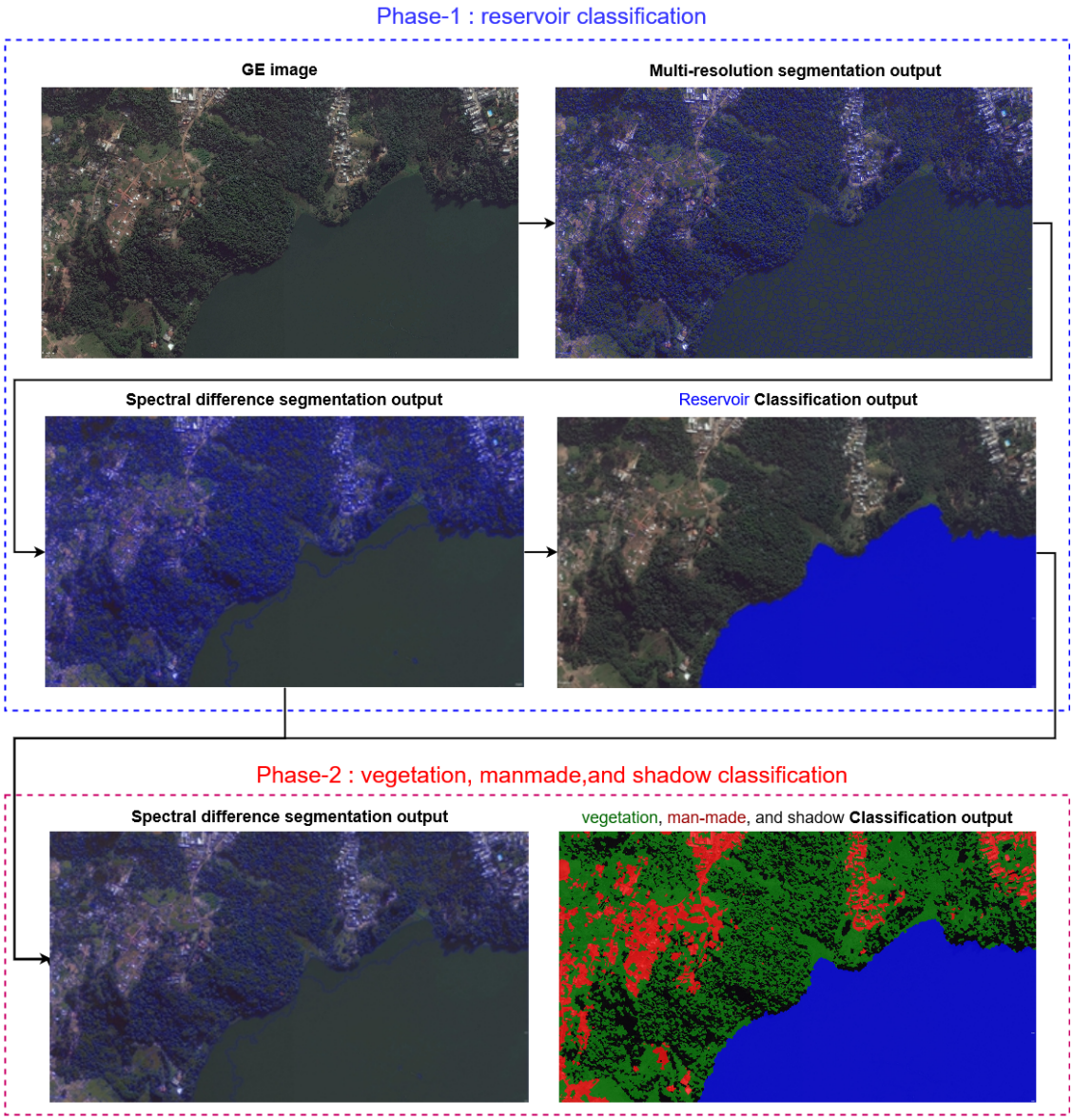

Figure 3.3: *Visual pipeline of the proposed tow-phase OB classification approach.*

3.2.1 Data Formation

The Google Earth Pro© is the data source adopted in our experiments, although the proposed method may also be applied to other sources. The database of this software is mainly provided by DigitalGlobe[1], Inc., and contains high-resolution images of many different areas on the earth Jensen [2015]. GE covers over 25% of the Earth's land surface and three-quarters of the global population Jensen [2015], Jacobson et al. [2015]. To improve the appearance of images when are displayed in the GE, the spectral information of images with more than three bands is reduced to only three bands (red, green, and blue) Visser et al. [2014], Guo et al. [2016], Potere [2008]. Furthermore, the appearance of GE images is improved using color balancing, warping, and mosaic processing Almeer [2012]. Containing an open database of historical RS images with flexibility in selecting the spatial resolution

[1]https://www.digitalglobe.com/

are other advantages of this software. The spatial resolution of GE imagery is flexible since it depends on the resolution of original images and the zoom level Liu et al. [2017].

In this research, we studied three zones around the Guarapiranga reservoir in São Paulo, Brazil. Their geographic coordinates are presented in Table 3.1.

Zone-1	Zone-2	Zone-3
23°43′58.01” S, 46°45′53.17” W	23°41′23.08” S, 46°44′10.14” W	23°46′43.74” S, 46°47′43.79” W

Table 3.1: *The geographical coordinates of the studied zones.*

We collected images from each study region over two years, 2009 and 2017. These six images are considered the test set. In order to form the calibration set, we collected an image from a sub-region of each test set image. Figure 3.4 illustrates the data collection method. The size of the calibration and test set images are 1116 x 632 and 4836 x 2739 pixels, respectively, and their spatial resolution is 0.3 meters.

3.2.2 Two-phase OB Classification

This stage aims to map RS images into four land covers; reservoir, man-made, vegetation, and shadow. The man-made class consists of roads (asphalted and not-asphalted), buildings, pools, impervious surfaces, and unfinished constructions. The implemented steps in this stage are explained in the following subsections.

3.2.2.1 Multi-resolution segmentation

This section is based on the multi-resolution concept adopted in Benz et al. [2004]. This approach is adopted in the eCognition software to analyze the images using the proposed OB method. In that work, Benz et al. [2004], the resolution is associated with objects of different possible sizes or scales that appear in RS images. For instance, trees are possible objects of small scale, rooftops are objects of intermediate scale, and reservoirs are large-scale objects. Multi-resolution, in this context, refers to hierarchical processing to analyze such possible scales Benz et al. [2004].

In the multi-resolution segmentation method, each pixel is essentially considered as a segment. Next, each segment is merged with its neighbors based on relative homogeneity (or heterogeneity) criteria. In every step, each segment is merged with the adjacent segment with minimum increase in heterogeneity, similar to region-growing strategies. If the heterogeneity increase be more than a given scale criterion, then no merge will take place. The heterogeneity criterion and the scale constitute the most critical parameters of the multi-resolution segmentation method.

The heterogeneity measures how heterogeneous a segment is based on color heterogeneity and shape heterogeneity. The color and shape heterogeneities are based on the standard deviation of the spectral colors and the deviation of a compact (or smooth) shape, respectively.

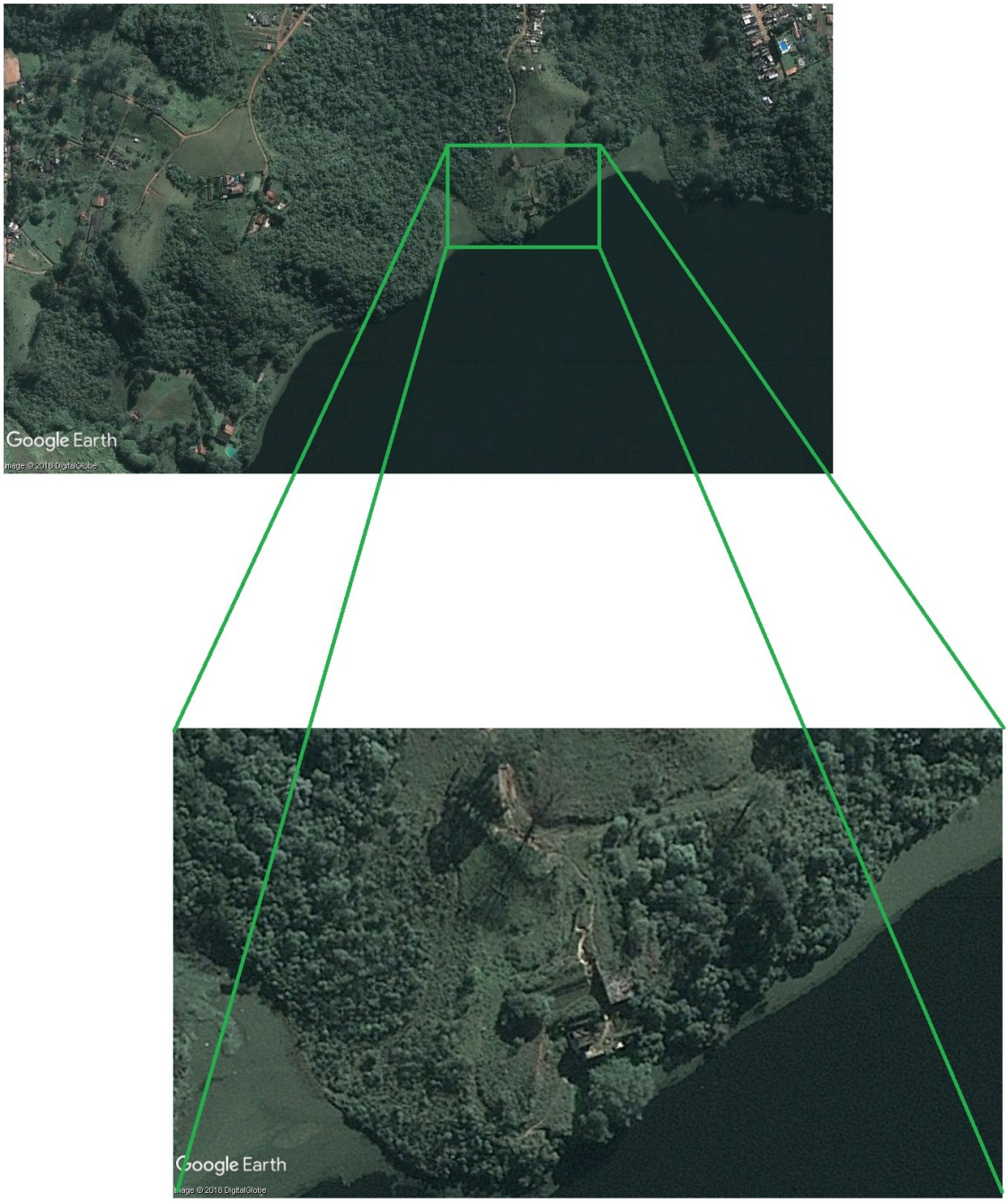

Figure 3.4: *Proposed approach for data collecting.*

The heterogeneity criteria can be customized by weighting these two criteria. The shape ratio determines to what degree shape influences the segmentation compared to color, and the compactness ratio gives it a relative weighting against smoothness. The scale parameter is related to the average size of the detected objects.

Consider Δf as heterogeneity change of a possible merge defined as:

$$W_{shape}\Delta h_{shape} + W_{color}\Delta h_{color} \tag{3.1}$$

where $0 < W_{shape}, W_{color} < 1$ and $W_{shape} + W_{color} = 1$. The color heterogeneity is computed as follow:

$$\Delta h_{color} = \Sigma_{i=1}^{N} W_i (n_{merge}\sigma_i^{merge} - (n_{seg1}\sigma_i^{seg1} + n_{seg2}\sigma_i^{seg2})) \tag{3.2}$$

where N is number of channels of the image, W_i is the weight of ith channel, n_{seg1} is number of pixels in segment 1, n_{seg2} is number of pixels in segment 2, n_{merge} is $n_{seg1} + n_{seg2}$, σ_i^{merge} is the standard deviation of channel ith within merged segment, and σ_i^{seg1} and σ_i^{seg2} are the standard deviations of ith channel of the segments 1 and 2, respectively.

The shape heterogeneity is defined as:

$$\Delta h_{shape} = W_{compact}\Delta h_{compact} + (1 - W_{compact})\Delta h_{smooth} \tag{3.3}$$

where $\Delta h_{compact}$ and Δh_{smooth} are :

$$\Delta h_{compact} = n_{merge}\frac{l_{merge}}{\sqrt{n_{merge}}} - (\frac{l_{seg1}}{\sqrt{n_{seg1}}} + \frac{l_{seg2}}{\sqrt{n_{seg2}}}) \tag{3.4}$$

$$\Delta h_{smooth} = n_{merge}\frac{l_{merge}}{b_{merge}} - (\frac{l_{seg1}}{b_{seg1}} + \frac{l_{seg2}}{b_{seg2}}) \tag{3.5}$$

l is the border length and b is the perimeter of the segment bounding box Benz et al. [2004].

Two issues arise when assessing segmentation results: Over-segmentation and under-segmentation. It has been shown that as long as under-segmentation is acceptable, a high classification accuracy is achievable Belgiu and Drăguţ [2014], Grybas et al. [2017]. Therefore, in this phase of segmentation, we followed the over-segmentation idea. Examples of over-segmentation and under-segmentation are illustrated in Figure 3.5.

We segmented calibration set images with different scales, shape, and compactness factors and compared segmentation results by visual inspection. When the scale, shape, and compactness parameters are 35, 0.7, and 0.5, respectively, segments become internally homogeneous, and all pixels within a segment belong to one class.

3.2.2.2 Spectral difference segmentation

Spectral difference segmentation merges segments produced by another segmentation method. In this method, neighbor segments whose differences between their layer mean intensities (e.g., mean of Red band in the segments) are below the value given by the maximum spectral difference are merged Tri [2018]. Consider seg_1 and seg_2 as two neighbor segments. Let $(R_{seg_1}, G_{seg_1}, B_{seg_1})$ and $(R_{seg_2}, G_{seg_2}, B_{seg_2})$ be the mean spectral values for R, G, and B bands of these two segments. If we consider w_R, w_G, and w_B as the weights for R, G, and B bands, respectively, the spectral difference between these two segments can be computed as:

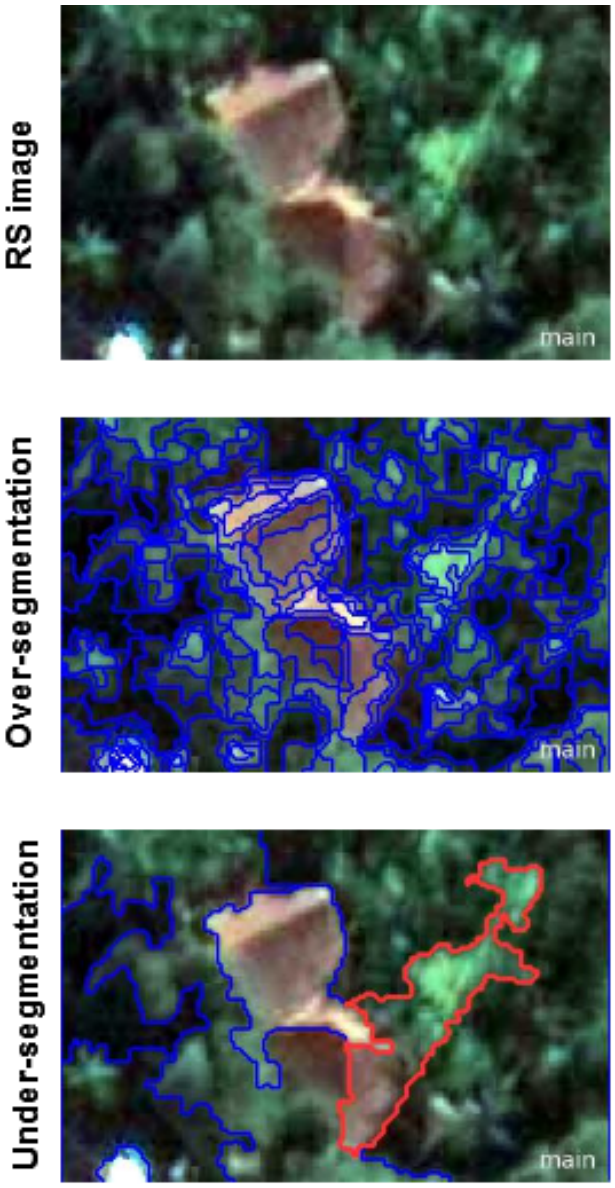

Figure 3.5: *Example of a GE image and two segmentation issues. The rooftops in over-segmentation are covered by several segments. Whereas in under-segmentation, the segments over rooftops (e.g., the segment highlighted with a red border) contain other scene classes as well.*

$$\frac{w_R\left|R_{seg_1} - R_{seg_2}\right| + w_G\left|G_{seg_1} - G_{seg_2}\right| + w_B\left|B_{seg_1} - B_{seg_2}\right|}{w_R + w_G + w_B} \tag{3.6}$$

The spectral differences between generated segments over the reservoirs tend to be small. On the other hand, classifying them using their spectral information leads to poor results because of their spectral similarity with shadow objects and some man-made objects. Therefore, the spectral difference segmentation method is implemented to merge neighbor segments when the difference between their mean intensities is very low (smaller than 1 in our experiments). Employing this method reduces the number of segments, especially over reservoirs, and generates more accurate segments over the scene reservoir objects. Utilizing these generated segments, we could classify reservoir objects without their spectral

information.

The Spectral difference segmentation is applied in two stages of our pipeline, as seen in Figure 3.2. The second application of this method merges segments with low spectral differences that belong to the same classes and, consequently, constructs more accurate segments over unclassified objects.

3.2.2.3 Classification

The second step in the OB method is classifying segments obtained from the segmentation step. In our two-phase approach, classification is applied twice: in the first phase, for reservoir classification, and in the second phase, to classify the other classes of interest.

In this study, the rule-based classification method is implemented. As mentioned, the spectral difference segmentation generated more accurate segments over the reservoir objects in the GE images. It made it possible to classify the reservoir objects using two spatial information of generated segments: "the segment area" and "the relative border of the segment to reservoir". Consider l is the border length of a segment and l_t is the common border of this segment with segments classified as the reservoir, then "the relative border of the segment to the reservoir" will be l/l_r. Therefore, for example, "the relative border of the segment to reservoir" for a segment surrounded by reservoir objects equals one.

The classification step in the second phase aims at classifying segments into vegetation, man-made object, and shadow. As the number of target classes increases and different classes have similar spectral properties, applying the rule-based method is more challenging and time-consuming. Such difficulties illustrate a drawback of the OB approach, which requires heuristics and an empirical setup of features and parameters. Even so, this method may lead us to the best results empirically Ziaei et al. [2014]. Therefore, we implemented two steps to speed up the process of extracting features and suitable thresholds to be used for classifying the segments:

Spectral similarity analysis A new step is implemented to analyze spectral similarities between segments of the same classes in different GE images. This step applies the nearest neighbor classifier to the calibration set images. The classifier is trained utilizing some proposed features in the literature Pande-Chhetri et al. [2017], Hu et al. [2013b] (e.g., mean values per band, standard deviation, brightness, area) and samples of segments of 2009 calibration images. Not only were the model outputs for 2017 images too poor but also, the 2009 LCMs were not satisfying. Regarding our experiments, we decided to classify the GE images of each year separately to achieve better results.

Feature analysis In this step, firstly, the segments from all classes from each year of the calibration set images are selected. Next, using these samples, the histograms of defined features are constructed for every two classes. Finally, the features with small histogram overlap between classes are selected to be used in the classification. The features adopted in our experiments are brightness, standard deviation, number of pixels/length, quantile,

the minimum distance between a man-made object with reservoir objects, and $R - G$ and $(R - B)^2 - (R - G)^2$, where R, G, and B represent the mean value of Red, Green, and Blue bands of segments, respectively.

The necessary rules and related applied thresholds for classifying and proper segmentation parameters have been adopted using the calibration set images. These are then applied to the GE images in the test set. The calibration set images consist of significantly smaller numbers of segments. It causes the process of adopting the rules, features, and parameters' values to be more straightforward.

3.2.3 Data Analysis: Post-classification comparison change detection

In order to illustrate a possible application of the proposed approach, we apply it to compare man-made object land cover in different years. In Figure 3.6, the experimental setup is illustrated.

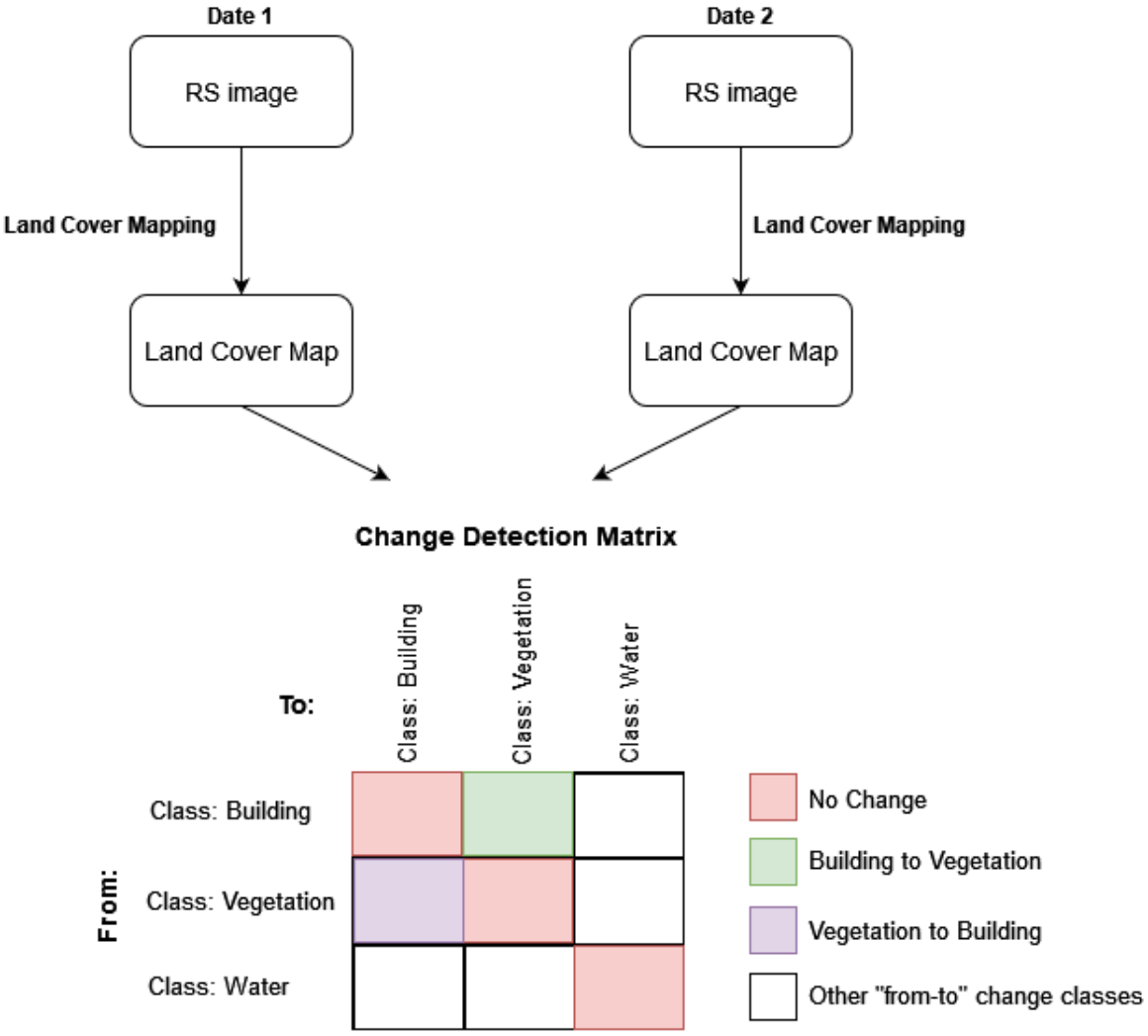

Figure 3.6: *Classification comparison for three land cover types.*

In this stage, a change detection matrix for each studied zone using the results of the mapping test set images are constructed. Then, the important "from-to" changes are

computed.

3.3 Experimental Results and Accuracy Assessment

Expansion of dormitory neighborhoods around Guarapiranga reservoir has threatened this reservoir since 1991 Osava [2008]. Since this reservoir is one of the most important water supply sources of São Paulo city and its preservation is critical for urban management, we decided to analyze the constructions around it to illustrate the proposed method.

3.3.1 Implementation Details

In this research, RS images are obtained using Google Earth Pro software. The results of segmentation and classification are obtained using eCognition software, and results of accuracy assessment and change detection are obtained using Python. This section aims to give an overall view of the implementation of the method described in section 3.2.

This research provides RS images via Google Earth Pro 7. 3. 1. Downloading this user-friend software is free. Four steps are necessary to obtain GE images from a geographical area of interest: first, finding the area of interest on the map manually or by setting its geographical coordinates; second, setting the desired eye altitude; third, selecting the date of the GE image using "historical imagery" tool; and fourth, saving the GE image in available resolutions using "save image" tool.

In addition, segmentation and classification are implemented in eCognition Developer 9.3.1 software. This software prepares different segmentation and classification methods to be utilized. After loading the image, appending a new rule in the "process tree" window to access the list of all possible algorithms is necessary for segmenting an RS image. By selecting the multi-resolution segmentation method, a new window opens that contains the parameters of this method. After changing the value of parameters to the desired values and executing the algorithm, the result of segmentation on the current image is visible. Users can set values from 0 to 1 for the shape and compactness factor to determine objects at a certain scale level. The larger scale parameter results in larger image segments and vice versa. The spectral difference segmentation method can be implemented similarly.

Furthermore, to classify segments using the rule-based method, the first step is listing the names of intended classes in the "class hierarchy" window. The second step is defining all features (e.g., mean values per band, variance, distances, neighborhood, Etc.) in the "Features" tab of "Image object Information" window. The third step is making the set of intended rules in the "process tree" window to classify images using them. To create each of these rules, following these steps are necessary: 1)appending a new rule in the "process tree" window, 2) selecting "assign class" in the list of algorithms, 3) making the intended rule in the newly opened window.

As explained in Chapter 3, implementing the nearest neighbor classification method

is suggested for the spectral similarity analysis step. A possible implementation of this algorithm with details may be found at Miller et al. [2013]. Also, in the feature extraction step, constructing histograms of defined features for every two classes is proposed. For this, the first step is selecting the segments of every defined class from the current map as samples, and the second step is opening "Sample Editor" using Classification > Samples > Sample Editor from the main menu. This window shows constructed histograms using selected samples.

3.3.2 Experimental Results

Figures 3.7-3.12 demonstrate the collected GE image from Zone-1 in 2017, besides the results of implementing the proposed two-phase OB classification stage on this image. As is illustrated in Figure 3.9, after implementing the spectral difference segmentation method in Figure 3.8, a segment is generated that covers almost the whole reservoir. The result of applying spectral difference segmentation for the second time is illustrated in Figure 3.11. This Figure shows the high reduction in the number of segments and the creation of segments closer to the scene objects.

Figure 3.7: *Example of a test set image. This image is collected from zone-1 in 2017.*

Figure 3.8: *Result of implementing multi-resolution segmentation on Figure 3.7.*

Figure 3.9: *The output of applying spectral difference segmentation on Figure 3.8. Applying this segmentation step reduced The number of generated segments over the reservoirs significantly.*

Figure 3.10: *produced image after classifying reservoir in Figure 3.7.*

Figure 3.11: *The output of implementing spectral difference segmentation on segments in Figure 3.9 that are unclassified in Figure 3.10.*

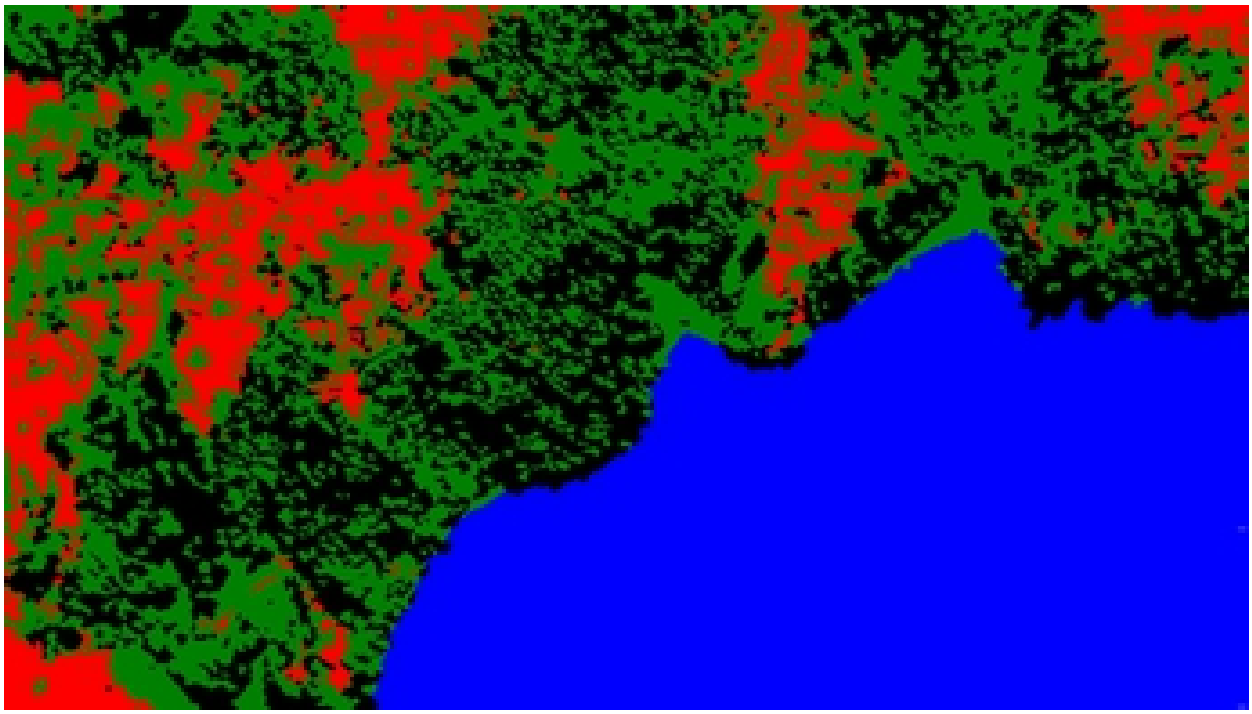

Figure 3.12: *Figure 3.7 final land cover map. Red: man-made, green: vegetation, black: shadow, blue: reservoir.*

Figures 3.13 and 3.14 illustrate the collected GE image from Zone-1 in 2009 and its derived land cover map, respectively. Examples of Zone-2 and Zone-3 test set images beside their LCM produced utilizing the proposed pipeline are illustrated in Figure 3.15. The percentages of "from vegetation to man-made" and "from shadow to man-made" change classes in new man-made areas in 2017 test images are presented in Table 3.2. These percentages show a potential environmental impact around the Guarapiranga reservoir. Since most of the GE images' shadow areas correspond to vegetation, a high proportion of the changes indicated as "from shadow to man-made" belongs to the "from vegetation to man-made" class. Furthermore, a comparison of the proportions of man-made class in the land cover maps from different years shows that the man-made area in the three studied zones increased from approximately 40% to 400% from 2009 to 2017.

Figure 3.13: *A test set image collected from Zone-1 in 2009.*

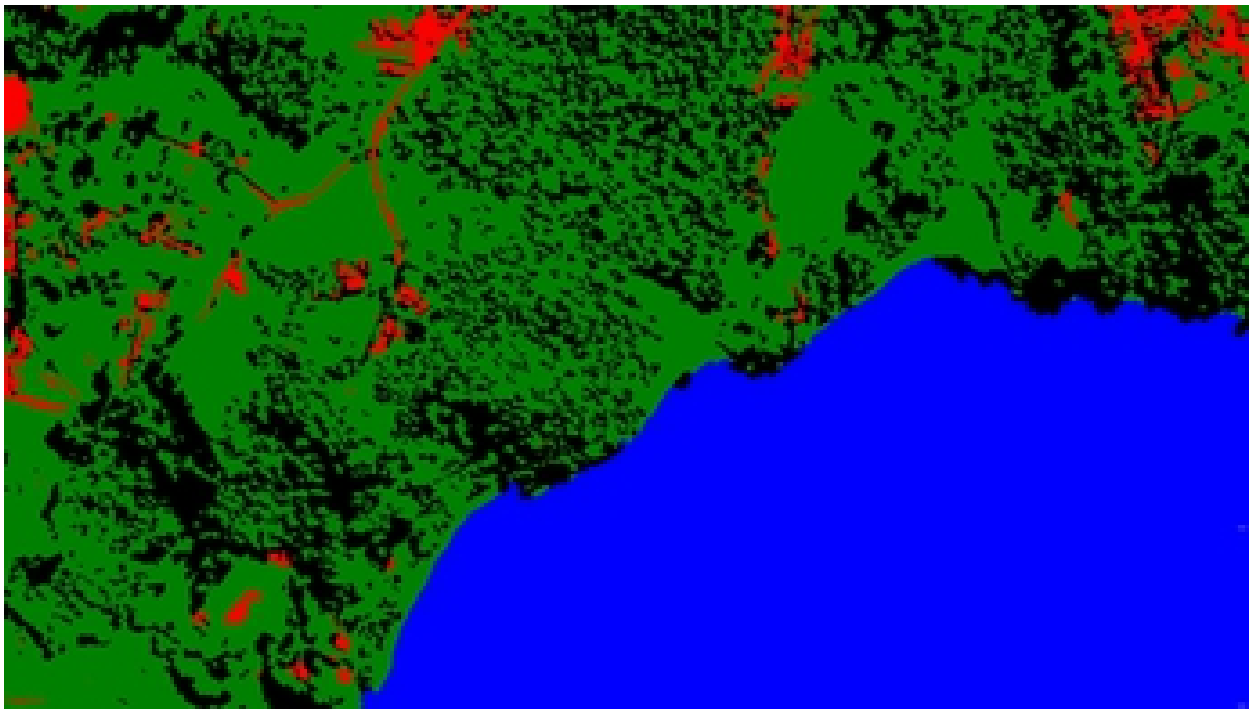

Figure 3.14: *Figure 3.13 land cover map. Red: man-made, green: vegetation, black: shadow, blue: reservoir.*

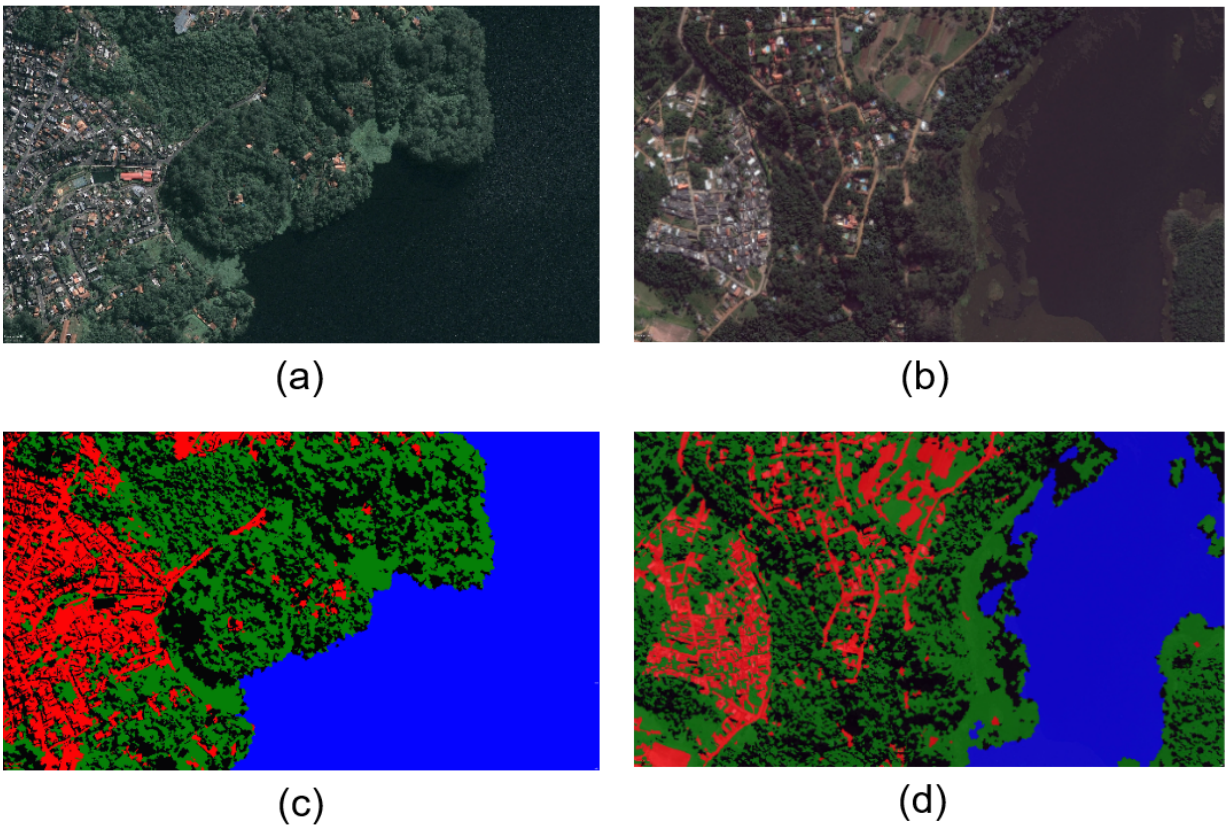

Figure 3.15: *Two test set images from different locations and their LCM. (a) and (b) are the GE image collected from Zone-2 in 2009 and Zone-3 in 2017, respectively; (c) OB LCM of (a); (d) OB LCM of (b).*

Table 3.2: *Proportions of "from vegetation to man-made" and "from shadow to man-made" from 2009 to 2017 in three studied zones.*

	from vegetation to man-made (%)	from shadow to man-made (%)
Zone-1	87.22	12.78
Zone-2	28.83	71.17
Zone-3	68.39	31.61

Figure 3.16 summarizes the segmentation results for images of the same region in 2009 and 2017. Another format to display land cover changes is the histogram. Figure 3.17 displays and compares the average proportions of every land cover type in the studied zones. We can easily conclude from the results that the man-made area is rapidly expanding, from 8% to 14%.

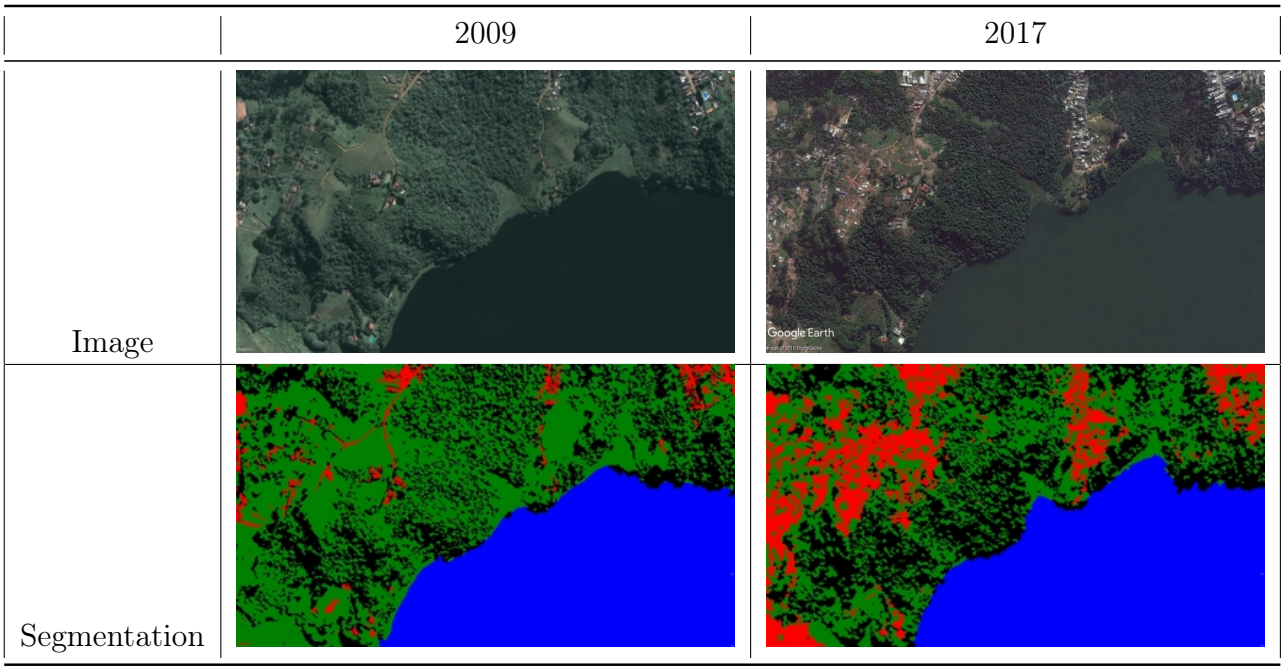

Figure 3.16: *Segmentation comparison for images of the same region in 2009 and 2017. The red and green pixels represent man-made and vegetation classes, respectively. In this figure the environmental degradation around the reservoir is depicted.*

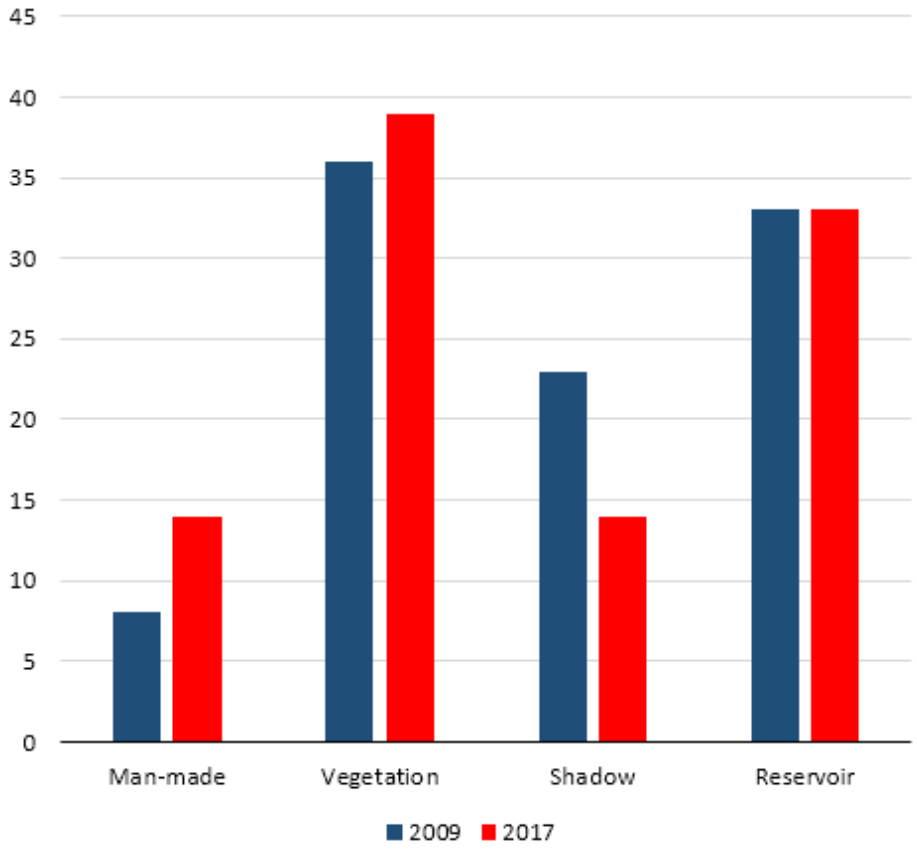

Figure 3.17: *Average percentage of different land cover types in three studied zones.*

3.3.3 Workflow Validation

Each collected GE image is individually assessed using a number of independent validation points generated by a stratified random sampling scheme. Stratified random sampling is an unbiased sample selection method that ensures adequate sampling in each class. Since it selects a minimum number of samples from each class Congalton and Green [2008]. A minimum of 50 samples for each class is suggested by Lillesand et al. [2014]. Therefore, due to the lack of survey points in the study areas, 250 and 500 independent points from calibration and test set images are randomly selected, respectively. From these number of samples, 50 and 100 points are distributed to each class in calibration and test set images, respectively.

The producer's accuracy (equation 3.7), user's accuracy (equation 3.8), overall accuracy (equation 3.9), and Kappa (equation 3.10) are utilized performance evaluation metrics that are calculated based on the confusion matrix (Table 3.3). The producer's and the user's accuracy indicate the probability of a reference pixel being correctly classified and the probability that a pixel classified on the map represents that category on the land, respectively. We also computed Kappa to determine whether the presented results in the confusion matrix are significantly better than a random result or not Jensen [2015], Congalton and Green [2008].

Table 3.3: *Confusion matrix when there are k classes.*

		Ground truth class				
		1	2	3		k
Map class	1	x_{11}	x_{12}	x_{13}	....	x_{1k}
	2	x_{21}	x_{22}	x_{23}	....	x_{2k}
	3	x_{31}	x_{32}	x_{33}	....	x_{3k}
	.	.	.	.	....	.
	k	x_{k1}	x_{k2}	x_{k3}	....	x_{kk}

$$Producer's\ accuracy\ for\ class\ j = \frac{x_{jj}}{x_{+j}} \tag{3.7}$$

$$User's\ accuracy\ for\ class\ j = \frac{x_{jj}}{x_{j+}} \tag{3.8}$$

$$Overall\ accuracy = \frac{\Sigma_{j=1}^{k} x_{jj}}{\Sigma_{i=1}^{k}\Sigma_{j=1}^{k} x_{ij}} \tag{3.9}$$

$$Kappa = \frac{N\Sigma_{i=1}^{k} x_{ii} - \Sigma_{i=1}^{k}(x_{i+}\ x_{+i})}{N^2 - \Sigma_{i=1}^{k}(x_{i+}\ x_{+i})} \qquad N = \Sigma_{i=1}^{k}\Sigma_{j=1}^{k} x_{ij} \tag{3.10}$$

Tables 3.4-3.9 indicate the producer's, user's, overall accuracy, and Kappa of generated land

cover maps. In these tables, "Pro" and "Use" are abbreviations of the producer's and user's accuracy, respectively. Presented accuracy assessment results are computed using one run.

Table 3.4: *Accuracy assessment of Zone-1 calibration set images land cover maps.*

	2009		2017	
	Use (%)	Pro (%)	Use (%)	Pro (%)
man-made	90.20	97.87	84.62	96.49
Vegetation	88.00	84.62	81.08	81.08
Shadow	89.86	96.87	95.08	87.88
Reservoir	100	90.16	96.55	91.80
overall accuracy (%)	91.60		88.76	
Kappa (%)	88.71		84.98	

Table 3.5: *Accuracy assessment of Zone-1 test images land cover maps.*

	2009		2017	
	Use (%)	Pro (%)	Use (%)	Pro (%)
man-made	92.08	93.94	88.70	87.18
Vegetation	84.11	92.03	81.75	81.10
Shadow	96.58	91.87	88.01	92.50
Reservoir	100	93.57	100	97.97
overall accuracy (%)	92.80		90.04	
Kappa (%)	90.35		86.41	

Table 3.6: *Accuracy assessment of Zone-2 calibration set images land cover maps.*

	2009		2017	
	Use (%)	Pro (%)	Use (%)	Pro (%)
man-made	92.31	96.00	87.50	97.25
Vegetation	83.82	81.43	81.20	94.34
Shadow	80.03	84.48	90.77	80.73
Reservoir	96.77	90.91	97.00	84.38
overall accuracy (%)	87.70		88.31	
Kappa (%)	83.55		84.41	

3.4 Conclusion

We proposed a new approach for OB LCM RGB RS images in this study. In the proposed approach, a multi-phase segmentation is adopted. Avoiding testing different values

Table 3.7: *Accuracy assessment of Zone-2 test images land cover maps*

	2009		2017	
	Use (%)	Pro (%)	Use (%)	Pro (%)
man-made	92.86	100	92.62	95.76
Vegetation	91.2	91.93	91.49	91.49
Shadow	91.2	87.69	94.06	90.48
Reservoir	94.93	92.26	98.54	98.54
overall accuracy (%)	92.60		94.21	
Kappa (%)	90.11		92.24	

Table 3.8: *Accuracy assessment of Zone-3 calibration set images land cover maps*

	2009		2017	
	Use (%)	Pro (%)	Use (%)	Pro (%)
man-made	94.12	92.31	92.98	96.36
Vegetation	88.00	89.19	82.67	92.54
Shadow	92.31	87.27	96.43	84.37
Reservoir	95.65	100	100	96.92
overall accuracy (%)	92.31		92.43	
Kappa (%)	89.66		89.89	

Table 3.9: *Accuracy assessment of Zone-3 test images land cover maps*

	2009		2017	
	Use (%)	Pro (%)	Use (%)	Pro (%)
man-made	88.99	96.04	93.16	97.32
Vegetation	95.74	85.44	87.22	87.88
Shadow	89.31	94.35	92.59	91.24
Reservoir	98.45	100	100	96.9
overall accuracy (%)	93.33		93.14	
Kappa (%)	91.07		90.84	

for adjusting the parameters of a time-consuming high-performance segmentation method called multi-resolution, obtaining segments with shapes close to the shapes of scene objects, and gaining the spatial features of high accurate generated segments for classifying reservoirs, are advantages of the proposed method. This is analogous to some selective attention-based methods that may be found in the literature.

Spectral similarity analysis and feature extraction steps are proposed to be implemented before defining the rules. These two steps indicate the suitable parameters and thresholds to be used for defining rules. Collecting two datasets as calibration and test sets are proposed, where the calibration set image sizes are almost one-sixteenth of the test set images and have

significantly less number of scene objects compared to the test set images. Utilizing these methods can significantly decrease the necessary time for achieving the desired results when we visually assess the parameters of the multi-resolution segmentation method and classify images with the rule-based method. We selected three zones around the Guarapiranga reservoir in São Paulo, Brazil, to collect our remote sensing images.

Chapter 4

A DL-based approach for man-made object segmentation around reservoirs

This chapter content is based on the publication Hamidishad and Marcondes Cesar Jr [2023].

4.1 Introduction

Despite the efficacy and popularity of the OB image analysis approach, this method has some disadvantages. For instance, most of its applications rely on pricey commercial software; and it heavily relies on expert experience in establishing the most suitable parameters. Therefore, its robustness is undermined, and its outputs may differ highly from case to case.

In recent years, DL has attracted considerable attention as a method for segmenting the RS imagery semantically and has achieved remarkable success. In this chapter, we develop a new approach based on DL and image processing techniques for man-made object semantic segmentation around the reservoirs. In order to segment man-made objects around the reservoirs in an end-to-end procedure, segmenting reservoirs and identifying the region of interest (RoI) around them are essential. In the proposed two-phase workflow, the reservoir is initially segmented using a DL model. A post-processing stage is proposed to remove errors such as floating vegetation. Next, the RoI around the reservoir (RoIaR) is identified using the proposed image processing techniques. Finally, the man-made objects in the RoIaR are segmented by a DL model. In order to illustrate the proposed approach, our task of interest is segmenting man-made objects around some of the most important reservoirs in Brazil. Therefore, we trained the proposed workflow using collected Google Earth (GE) images of eight reservoirs in Brazil over two different years.

The organization of this chapter is as follows: Section 4.2 describes the studied reservoirs, collected data characteristics, applied data pre-processing pipeline, the proposed workflow for segmenting man-made objects around the reservoirs, and corresponding utilized methods. Next, the performance of each workflow stage, besides results visualization and workflow

evaluation, is explored in section 4.3. Finally, the results and findings of the study are discussed in section 4.4.

4.2 Materials and Methods

Our task of interest is segmenting man-made objects around reservoirs. The proposed approach (see Figure 4.1) is based on three main steps: 1- Reservoir map generation; 2- RoIaR detection; 3- Man-made object segmentation in the RoIaR. The data is initially collected and pre-processed to be prepared in a suitable manner. Then, patches of constructed images are fed to phase-1 for reservoir segmentation. The reservoir map is passed to phase-2, where the RoIaR is detected. This RoIaR is used as a mask where man-made objects are finally segmented. The workflow is detailed in the following subsections.

4.2.1 Data Collection

Our experiments are performed on RGB remote sensing images collected from eight reservoirs in Brazil using the Google Earth Pro© software. GE images represent an integration of multiple satellite data sources, mainly DigitalGlobe's QuickBird commercial satellite and EarthSat that mostly is from Landsat-7 Qian et al. [2020]. Aiming to improve the images' appearance, the spectral information of images with more than three bands is reduced to RGB Visser et al. [2014]. Furthermore, the GE images' appearance is improved using color balancing, warping, and mosaic processing Almeer [2012]. Besides being an open dataset of RS images, including historical images and flexibility in selecting images of different resolutions are additional advantages of this platform. The GE coordinate system is the World Geodetic System WGS84 standard.

The eight studied reservoirs are Anta, Billings (the largest reservoir in São Paulo, Brazil), Dona Francisca, Guarapiranga, Jaguara, Luiz Barreto, Nova Avanhandav (Nova), and Salto Osório. Their locations are visualized in Figure 4.2. Their geographic coordinates are listed in Table 4.1. For each reservoir, images over two different years are collected (Table 4.2). The size of collected images is 2683 $\times$ 4800 pixels. They are captured in different view altitudes and have consequently different resolutions (from approximately one meter up to two meters). A total of 206 images are collected, covering about 3000 square kilometers.

4.2.2 Data Preparation and Annotation

Data preparation involves two aspects: pre-processing for mosaic image formation and data annotation. The data preparation scheme is illustrated in Figure 4.3 using Guarapiranga reservoir samples. The data preparation aims to prepare data for training the phase-1 and phase-2 semantic segmentation models in Figure 4.1.

As is shown in Figure 4.3, the input images are initially mosaicked to eliminate overlapping

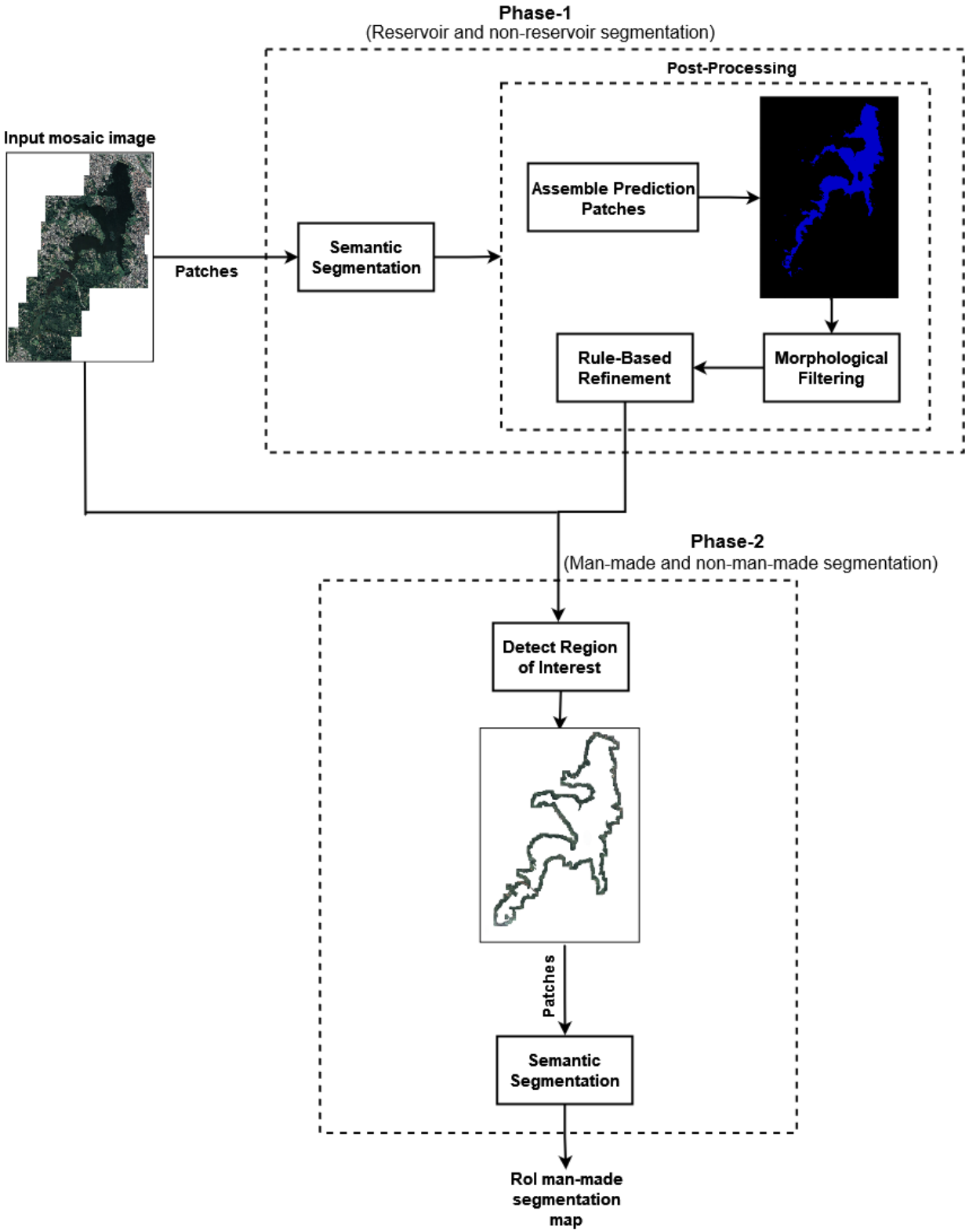

Figure 4.1: *Overview of the proposed analysis workflow.*

areas in collected GE images. Constructing the mosaic images is also essential for implementing the next steps. The mosaicked images are annotated into two classes, reservoir, and non-reservoir. To annotate images the Adobe Photoshop tools are used.

Next, to simplify the contour around the reservoirs, a polygonal approximation is initially

Table 4.1: *Locations of studied reservoirs.*

Reservoir	State	Coordinates
Anta	Minas Gerais and Rio de Janeiro	22°02′33.20″ S, 43°01′16.85″ W
Billings	São Paulo	23°48′50.62″ S, 46°32′19.39″ W
Dona Francisca	Rio Grande do Sul	29°26′34.18″ S, 53°16′09.09″ W
Guarapiranga	São Paulo	23°43′16.93″ S, 46°44′22.23″ W
Jaguara	Minas Gerais and São Paulo	20°05′01.85″ S, 47°24′10.44″ W
Luiz Barreto	São Paulo	20°14′18.50″ S, 47°11′01.95″ W
Nova	São Paulo	21°10′34.54″ S, 50°07′34.03″ W
Salto Osório	Paraná	25°33′28.60″ S, 52°57′07.61″ W

Table 4.2: *Acquisition years of each reservoir dataset. Some of the older year images of Luiz and Nova belong to 2004 and 2010, respectively.*

	Acquisition Years	
Reservoir	Older	Earlier
Anta	2014	2020
Billings	2009	2019
Dona Francisca	2011	2017
Guarapiranga	2009	2019
Jaguara	2010	2020
Luiz Barreto	2010	2020
Nova	2011	2021
Salto Osório	2005	2019

carried out Douglas and Peucker [1973], Cubes [1987], Costa and Marcondes Cesar Jr [2000]. This allows controlling the coarseness by the polygonal approximation parameter. Then, a rectangular box connecting each pair of consecutive polygon corners is defined. These boxes are enlarged to cover an at least distance from the border of the reservoir. The RoIaR is defined as the union of these boxes (see Figure 4.3) and is used to mask the mosaic image.

The masked RoIaR image is annotated to man-made and non-man-made objects:

- **Man-made objects**: road (asphalted and not-asphalted), rooftop, bridge, pool, urban and countryside constructions, impervious surface.
- **Non-man-made objects**: Vegetation, water body, bare land, plantations, Etc.

4.2.3 Phase-1: Reservoir segmentation

This step explores a deep neural network that segments input RGB patches to reservoir and non-reservoir. Encoder-decoder-based models have been trained and compared for this step. Below, we briefly describe the two models assessed in this study: U-Net and SegNet. Based on our evaluation, the SegNet-based model has been selected as the best in our experiments.

Figure 4.2: *Visualization of the studied reservoirs locations.*

The U-Net architecture introduced by Ronneberger et al. [2015] is based on a downsampling-upsampling procedure that concatenates feature maps between each encoder and corresponding decoder by skip connections (see Figure 4.4). In each step in the encoder path, two 3x3 convolutions followed by a ReLU and a 2x2 max-pooling with stride two are repeated. Furthermore, the number of feature channels in each downsampling step is doubled. After each upsampling in the decoder path, a 2x2 convolution that halves the number of feature channels is applied. These features are concatenated with the cropped feature of the corresponding encoder step, and then two 3x3 convolution-ReLU blocks are implemented.

Due to the unpadded convolutions utilized in the U-Net, the output size of the model is smaller than the input. Therefore, we avoided unpadded convolutions to keep the size of each output equal to the corresponding input and set the dropout to 0.3 to overcome the overfitting (named U-Net_p). On the other hand, a common strategy in DL research for training the CNNs properly and avoiding training from scratch is utilizing a pre-trained

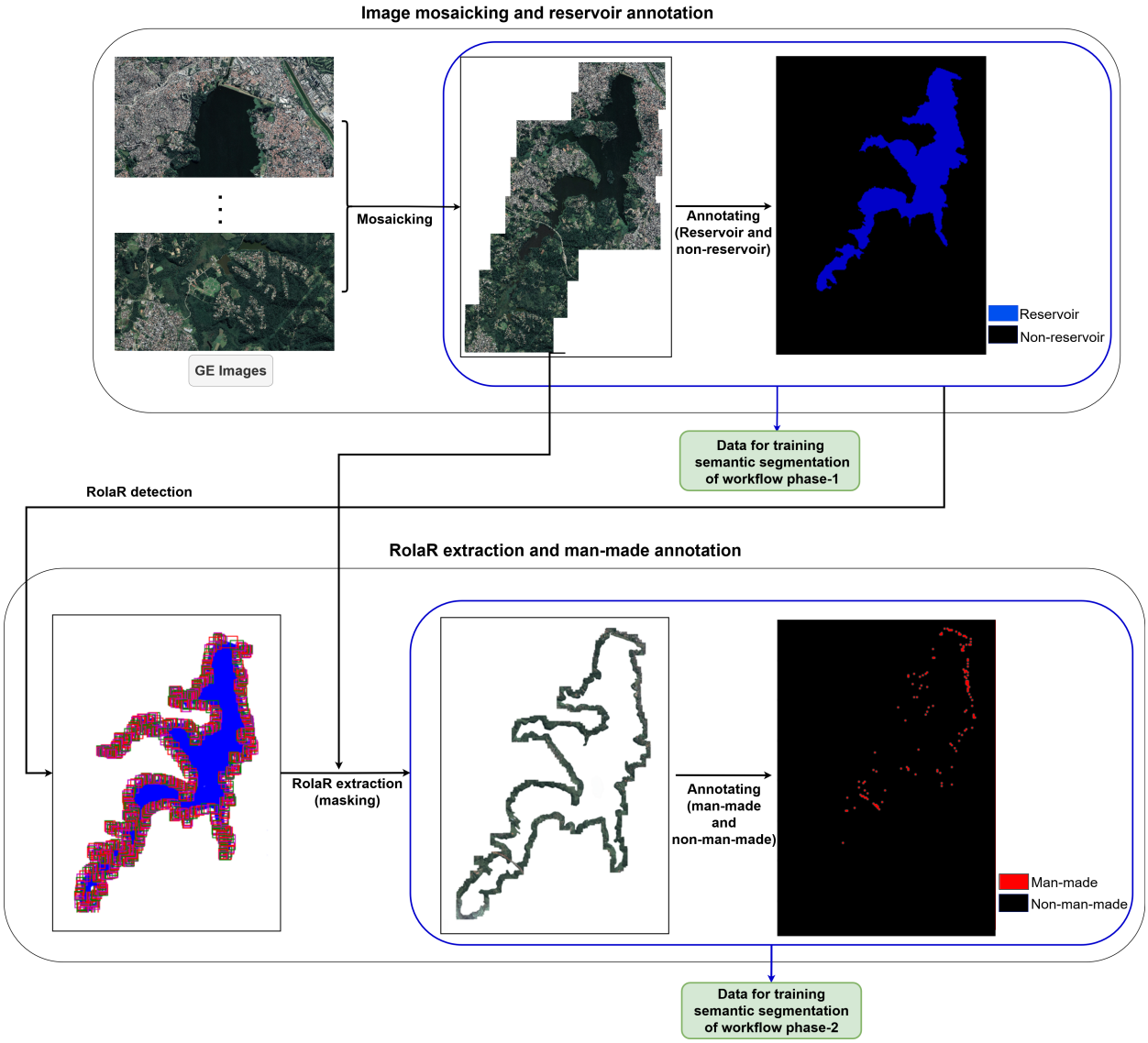

Figure 4.3: *Proposed data preparation and annotation pipeline.*

CNN as the initializer or as the fixed feature extractor, called transfer learning. Therefore, we trained a U-Net model whose encoder path was replaced by VGG-16 (named U-Net_v) and initialized with weights trained on the ImageNet dataset. However, the model overfitted extremely the train set. The other trained U-Net-based model (named U-Net_s), has fewer features. In this model, there is only one convolution block in each layer that is also batch normalized Ioffe and Szegedy [2015].

The SegNet architecture was first introduced by Badrinarayanan et al. [2015]. Similar to the U-Net, SegNet includes an encoder and a decoder part with the advantage that the need for learning to up-sample is eliminated. Since each decoder uses pooling indices computed in the max-pooling step of the corresponding encoder. After each convolution layer in the encoder path, a ReLU not-linearity is used, whereas, in the decoder, no ReLU not-linearity is presented. Furthermore, the number of channels per layer is constant (see figure 4.5). In the employed architecture (named SegNet_d), despite the original form, the number of feature

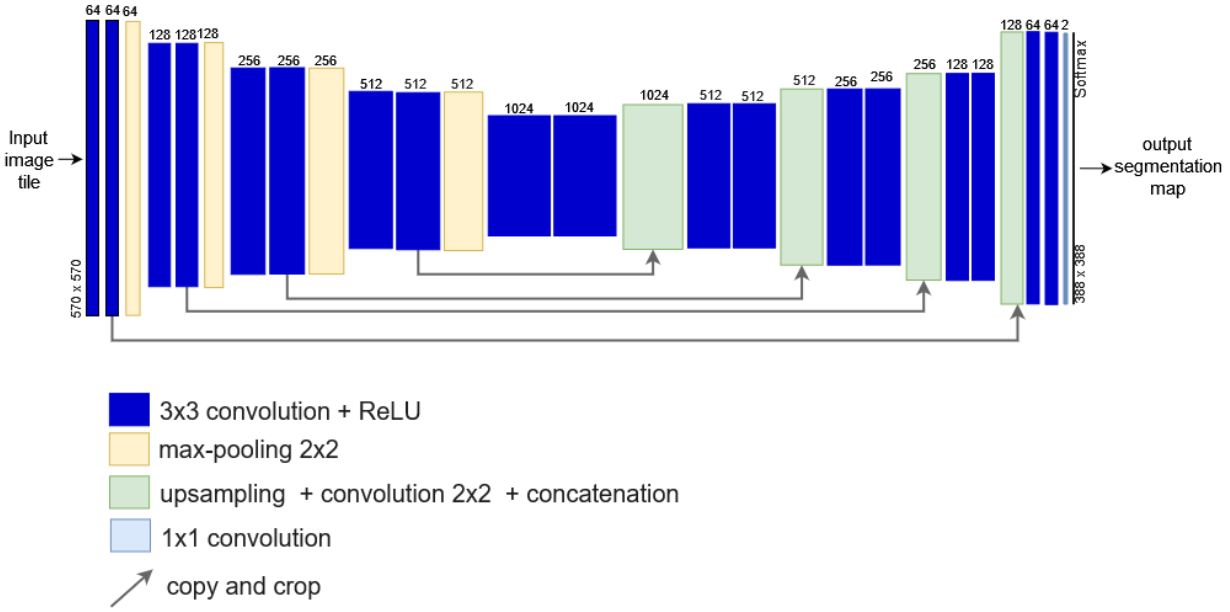

Figure 4.4: *The U-Net architecture.*

channels is doubled at each down-sampling step. Moreover, batch normalization is applied after each convolution layer.

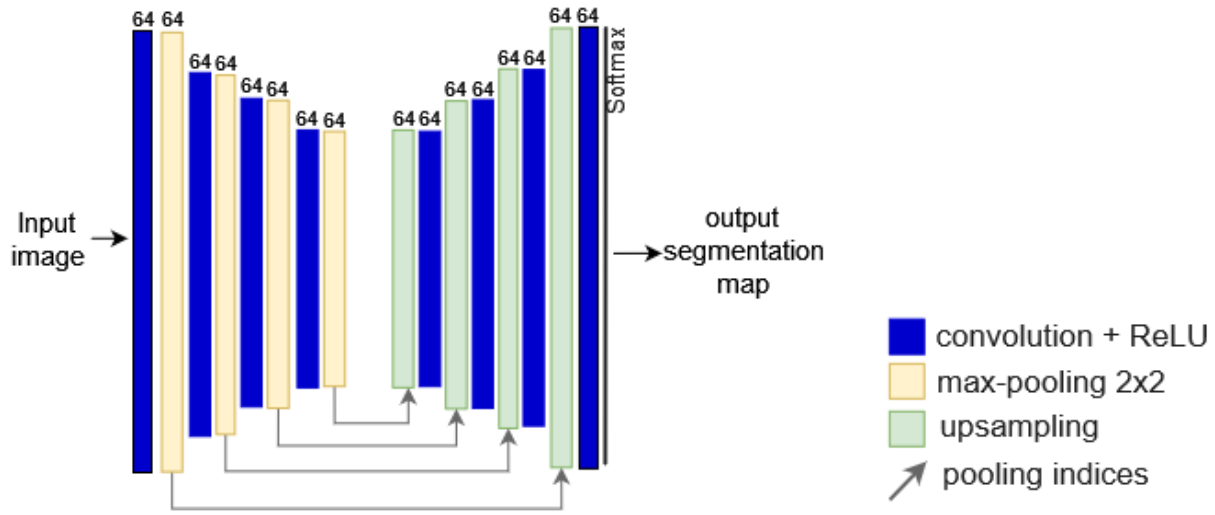

Figure 4.5: *The SegNet architecture.*

Data splitting Since the collected images correspond to different reservoirs geographically spread in Brazil, they have different visual properties. They may be obtained in different seasons, atmospheric and geological conditions, Etc. A possible approach to address such variability is to adopt the domain adaptation techniques. Since this is out of the scope of this thesis, we explored a data splitting approach to ensure variability in the train, test, and validation sets. Therefore, samples from each 16 mosaic images of reservoirs are used in these sets in the following proportions: 60% for the train set, 20% for the validation set, and 20% for the test set.

Post-processing Feeding models by mosaic images instead of patches is impossible because of the available GPU memory limits. Whereas, in many cases, patches do not contain important information about objects, such as their shapes, sizes, and locations in the images. However, this information is essential for detecting some water bodies from reservoirs. On the other hand, spectral similarities between objects of different classes also cause errors. Therefore, we proposed a post-processing stage to fix these errors.

In this stage, the segmented patches are initially assembled to form the reservoir map. Then, the morphological opening is applied to remove small false positive objects. Next, morphological closing is applied to remove small false negatives objects inside the reservoir objects.

Applying morphological transformations with a large kernel causes changes in the shapes of objects predicted as the reservoir. Accordingly, in order to remove errors inside reservoirs (such as floating vegetation) and large false positive objects (such as large water bodies around reservoirs) without removing reservoir objects that are separated because of constructed bridges over the reservoirs, the following rules are proposed:

- If a non-reservoir object is surrounded by a reservoir object, it is classified as the reservoir.
- If the size of a reservoir object is smaller than one-tenth of the size of the largest reservoir object, or the minimum distance between these two objects is greater than 300 meters, then it is classified as non-reservoir.

4.2.4 Phase-2: Man-made object segmentation

Once the reservoir is segmented, the next step is to detect and extract the RoIaR. Two possible approaches for RoIaR detection have been considered: polygonal approximation-based and mathematical morphology-based. The polygonal approximation approach has been described in Section 4.2.2, which is adopted for dataset annotation. Although this approach is useful for sparse data annotation (because we may control the polygonal approximation parameters), it produces patches of varying sizes that may not be suitable for analyzing man-made objects' evolution, for instance.

Therefore, a mathematical morphological approach is also explored. Let I denote the segmented reservoir image and s a structuring element. The dilated reservoir image is defined as $I_d = I \oplus s$, where $\oplus$ is the morphological dilation. The RoIaR R is defined as $R = I_d - I$, where $-$ denotes set difference.

Following the data annotation procedure illustrated in Figure 4.3, the detected RoIaR is applied as a mask to the original data for RoIaR extraction. The extracted RoIaR is then segmented into man-made and non-man-made objects.

Two widely used network architectures for RS semantic segmentation are the pyramid networks and encoder-decoder networks Mou and Zhu [2018]. In phase-2, the following

networks have been assessed: U-Net, Pyramid scene parsing network, Feature Pyramid network, and LinkNet, which are detailed in the following.

The Pyramid scene parsing network (PSPNet) has been introduced by Zhao et al. [2017] and won the ImageNet Scene Parsing Challenge 2016. It is a pyramid pooling module that enables the network to capture the context of the whole image. In this module, the feature map is pooled at different sizes and passed through a convolution layer. Next, these features are upsampled and concatenated with the original feature map and passed through a convolution layer to produce the final prediction (see Figure 4.6). We implemented PSPNet with different backbones in this study. Furthermore, besides PSPNet that downsamples input image to 1/8, the 1/4 downsampling is also trained.

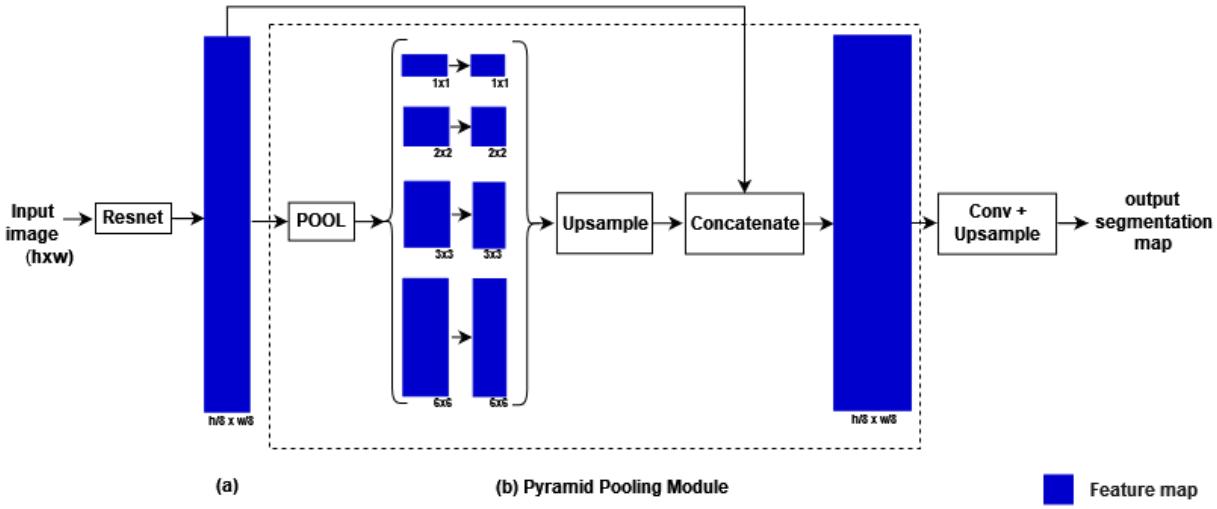

Figure 4.6: *An overview of PSPNet. The size of feature map channels is denoted below each box. The size of the last feature map in (a) is 1/8 of the input image size.*

The Feature Pyramid Network (FPN) was initially proposed by Lin et al. [2017a] for object detection. The general scheme of FPN is illustrated in Figure 4.7. The construction of this architecture involves a bottom-up path, a top-down path, and lateral connections. The scaling step in the bottom-up path (and consequently in the top-down path) is two. Each lateral link combines feature maps from the bottom-up and top-down pathways with the same spatial size. Finally, the feature maps in the top-down stages are upsampled to be the same size as the input image. These feature maps are combined and used to produce the prediction map. The ResNet is used as the backbone, whereas in this study, other backbones have also been experimented.

The LinkNet architecture proposed by Chaurasia and Culurciello [2017] is a semantic segmentation method that is constructed from an encoder and a decoder path (see Figure 4.8). Each residual block in the encoder path consists of two consequent convolution blocks. The input of each residual block is bypassed to its output. The decoder blocks consist of three convolution layers, and the middle is a full convolution. The advantage of the proposed architecture is passing the input of each encoder block to the output of the corresponding decoder block.

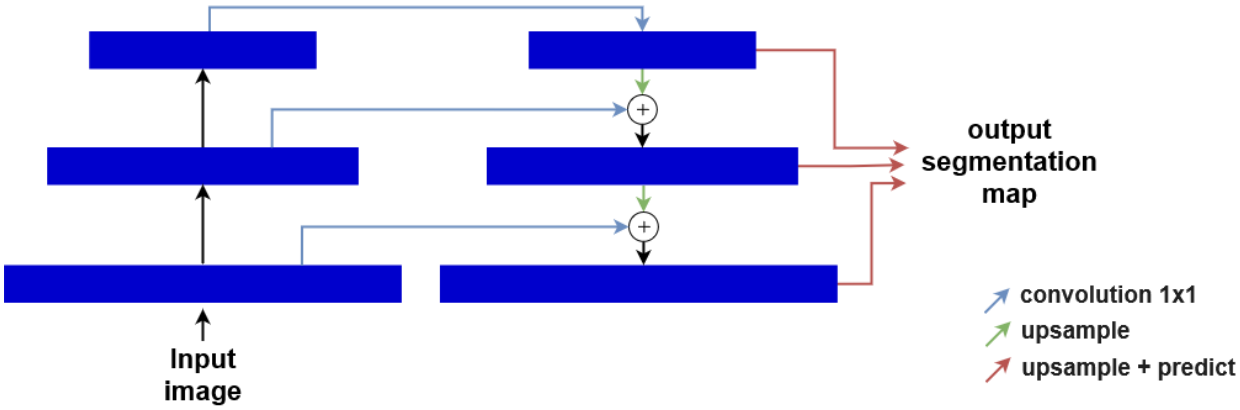

Figure 4.7: *An overview of FPN.*

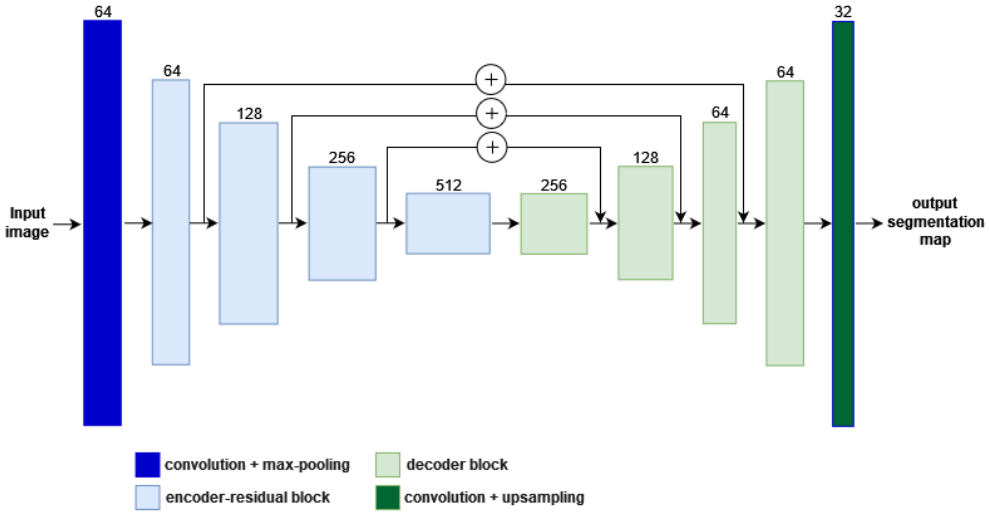

Figure 4.8: *An overview of LinkNet architecture.*

Data splitting Splitting data into the train and test sets is reported to work well when the dataset size is modest. On the other hand, the train and test sets must represent possible distributions of the addressed problem. Therefore, 70% for the train and 30% for the test set are selected randomly from each RoIaR.

Loss function The Focal loss proposed by Lin et al. [2017b] and the Dice loss that is based on Dice coefficient Sorensen [1948], Dice [1945] are utilized as the loss function for training the networks. Focal loss down-weights easy examples and hence helps the model to learn complex examples better. It is reported by Jadon [2020] that Focal loss works best when the data is highly imbalanced. To see how it works, first, consider the binary cross entropy loss (CE):

$$CE(p, y) = \begin{cases} -\log p, & \text{if } y = 1 \\ -log(p-1), & \text{otherwise} \end{cases} \tag{4.1}$$

where p is predicted probability for class with label y=1. Now lets define a new notation p_t:

$$p_t = \begin{cases} p & if y = 1 \\ p - 1, & \text{otherwise} \end{cases} \tag{4.2}$$

Using this notation we can rewrite Equation 4.1 as $CE(p_t) = -log(p_t)$. To balance the importance of positive/negative examples, we can consider α_t as the weight for class 1 and $1 - \alpha_t$ for class 0, then α-balanced CE will be written as:

$$CE(p_t) = -\alpha_t log(p_t) \tag{4.3}$$

Finally, to down-weight easy examples, they add factor $(1 - p_t)^\gamma$ to CE where $\gamma > 0$ is a tunable parameter. Based on the experiment, $\gamma = 2$ works best and is used in this study too.

The dice Loss is based on the dice coefficient (DC); see Equation 4.4. In the case of binary classification, A is the set of all positive examples, and B is the set of predicted positive examples.

$$DC = \frac{2\,|A \bigcap B|}{|A| + |B|} \tag{4.4}$$

Then, DC can be expressed as the following form:

$$DC = 2 \cdot \frac{TP}{2TP + FP + FN} \tag{4.5}$$

where TP, FP, and FN are true positive, false positive, and false negetive, respectively. The dice loss (DL) takes the following form:

$$DL = 1 - 2 \cdot \frac{\sum_{i=1}^{N} p_i r_i}{\sum_{i=1}^{N} r_i^2 + p_i^2} \tag{4.6}$$

where p_i is the predicted probability for pixel i-th and r_i is the ground truth of the corresponding pixel. The imbalance between the foreground and background can be efficiently reduced using Dice Loss. However, it disregards the imbalance in data difficulty.

4.3 Experimental Results

This section describes the experimental results of the proposed workflow. Phases 1 and 2 have been evaluated, and the results are discussed below.

4.3.1 Performance Evaluation Metrics

Three common statistics, precision (Equation 4.7), recall (Equation 4.8), and F1-score (Equation 4.9) which is a harmonic mean of precision and recall, are adopted as well as the confusion matrix of segmentation maps.

$$Precision = \frac{TP}{TP + FP} \tag{4.7}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.8}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + recall} \tag{4.9}$$

4.3.2 Phase-1 Experimental Results

The proposed methods have been implemented in Python using the Google Collaborator (Figure 4.9). The trained architectures for this phase are modified versions of U-Net and SegNet. All trained models apply the Binary Cross Entropy as the loss function. The learning rate in the Adam optimizer (proposed by Kingma and Ba [2014]) is set to 0.001, which is reduced by a factor of 0.2 after every five epochs with no reduction in validation loss down to 10^{-7}. Although the number of epochs is set to 100, training is stopped after 20 epochs with no reduction in the validation loss. Patches with 416 x 608 pixel sizes are fed into the networks, and train, validation, and test sets contain 6017, 2009, and 1998 patches, respectively. Vertical and horizontal flips are two types of augmentation that each one is applied randomly on 50% of train patches. The F1-score of trained models in segmenting the train and validation sets are presented in Table 4.3. As is illustrated in this Table, the U-Net_v overfits the train set.

Table 4.3: *Performance of trained architectures for phase-1 semantic segmentation stage.*

	F1-score	
Model	**Train set**	**Validation set**
U-Net_p	96.19	95.46
U-Net_v	92.11	68.86
U-Net_s	96.80	96.17
SegNet_d	97.22	96.56

The performances of models with healthy learning curves in segmenting the validation set are illustrated in Table 4.4. As illustrated in the Table, SegNet_d outperforms the U-Net-based models. The performance of SegNet_d in segmenting the test set is illustrated in Table 4.5. Some patches of studied reservoirs with different spectral properties besides their ground truths and SegNet_d, U-Net_s, and U-Net_p prediction outputs are shown in Figure 4.10.

Besides errors that occur because of spectral similarities between reservoirs and some other objects (such as shadows), there are small water bodies, rivers, Etc., in the images that are segmented, partially or entirely, as the reservoir by the models. This issue is unavoidable because of feeding patches to the models instead of the mosaic images. Therefore, post-

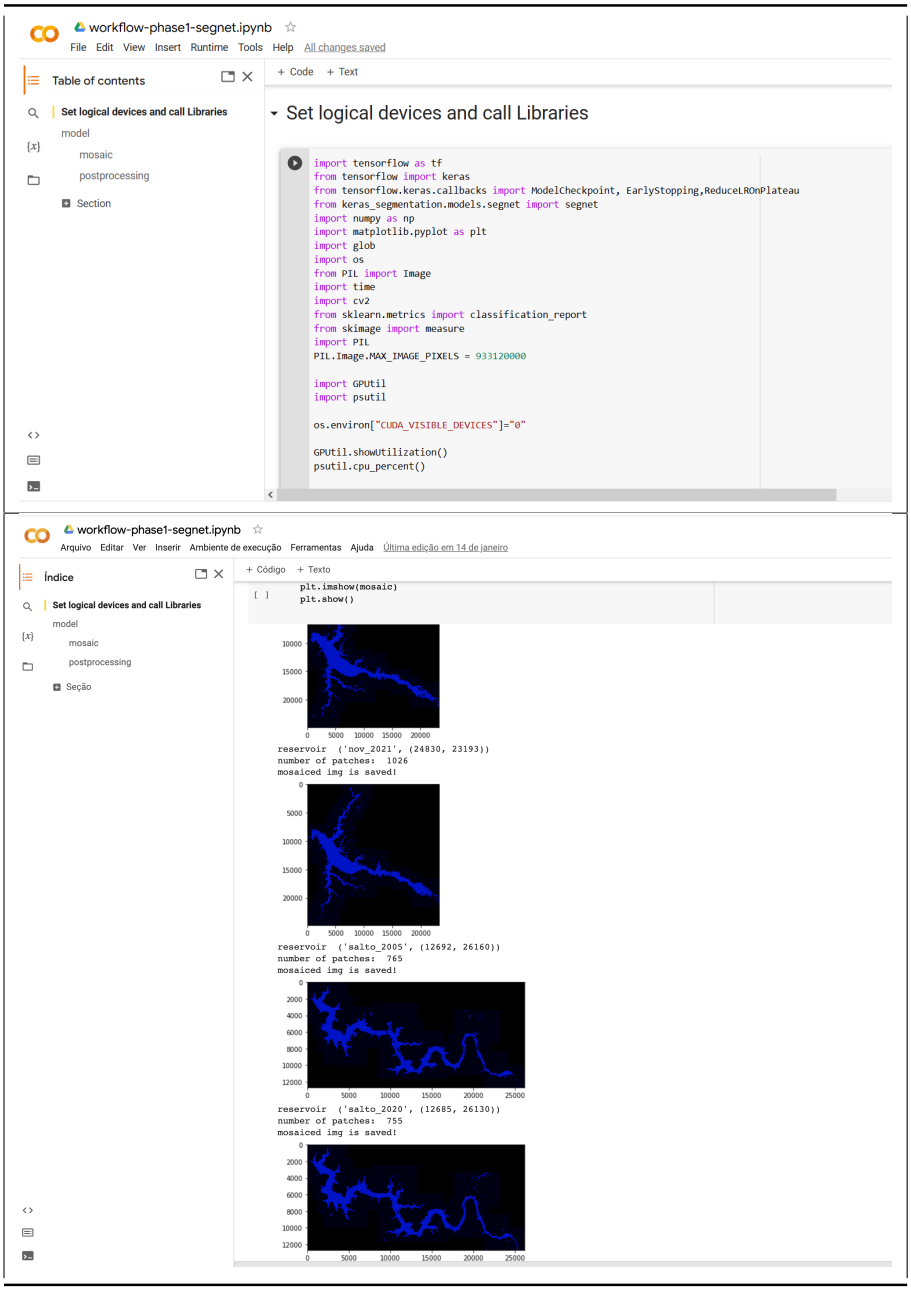

Figure 4.9: *Jupyter Notebook (Google Colaboratory) created to prepare the proposed method.*

Table 4.4: *The performance of models with healthy learning curves for Phase-1 semantic segmentation on the validation set.*

	Precision		Recall		F1-score	
Model	**non-reservoir**	**reservoir**	**non-reservoir**	**reservoir**	**non-reservoir**	**reservoir**
U-Net_p	98.18	93.72	98.71	91.27	98.44	92.48
U-Net_s	98.63	93.85	98.71	93.49	98.67	93.67
SegNet_d	98.79	94.39	98.82	94.24	98.81	94.32

processing the network outputs is an essential task. Figure 4.11 illustrates examples of non-interesting water bodies in the collected dataset and their segmentation results in generated reservoir maps.

Table 4.5: *SegNet_d performance on the test set.*

Class	Precision	Recall	F1-score	Support (N.pixels)
Non-reservoir	98.82	98.87	98.85	4211777759
Reservoir	94.33	94.11	94.22	84172385

Morphological operations post-processing is highly effective in removing minor errors, as described above. Each reservoir's applied structuring element size equals 100/(spatial resolution). For example, if the spatial resolution of a mosaic image is one meter, the structuring element size is 100x100.

As the reservoirs contain branches, applying morphological operations with large kernel sizes increases FP and FN objects. Accordingly, significant errors are removed by applying the two rules to objects in the produced segmentation maps. Post-processing using only rules is time-consuming because of the high number of FP and FN objects in prediction maps, whereas morphological operations speed up this process. Anta-2014 and Nova-2021 mosaic images, besides their ground truths, model outputs, and post-processing outputs, are illustrated in Figure 4.12. Moreover, SegNet_d performance in segmenting these two reservoirs besides post-processing performance are presented in Table 4.6. Applying the proposed post-processing improves the accuracy of produced reservoir maps except for two of the 16 studied cases.

Table 4.6: *Prediction and refinement performance metrics for Anta-2014 an Nova-2021.*

		Model		Post-processing	
Reservoir	Class	Precision	Recall	Precision	Recall
Anta-2014	non-Reservoir	98.45	98.95	99.15	99.56
	Reservoir	89.75	85.53	95.72	92.07
Nova-2021	non-Reservoir	98.33	98.03	98.67	99.13
	Reservoir	92.56	93.65	96.63	94.90

4.3.3 Phase-2 Experimental Results

As discussed above, VGG-16, ResNet-50, and ResNet-101 are the most frequented backbones Ghanbari et al. [2021]. In this study, these three backbones besides EfficientNet-B4 have been experimented. All backbones are initialized with weights trained on the ImageNet dataset. The Adam optimizer is used as the optimizer in all models. The initial learning rate is set to 0.0001 or 0.001, which is automatically reduced by a factor of 0.2 after every five epochs with no reduction in validation loss down to 10^{-7}. The mini-batch size is set to two and power of two (up to the possible size based on the model's size and available memory). The number of epochs in training all models is set to 80, and the early stopping is not implemented. Vertical and horizontal flips are two augmentation methods implemented on different portions of images (up to 0.7). We added dropout regularization ($<$0.3) to

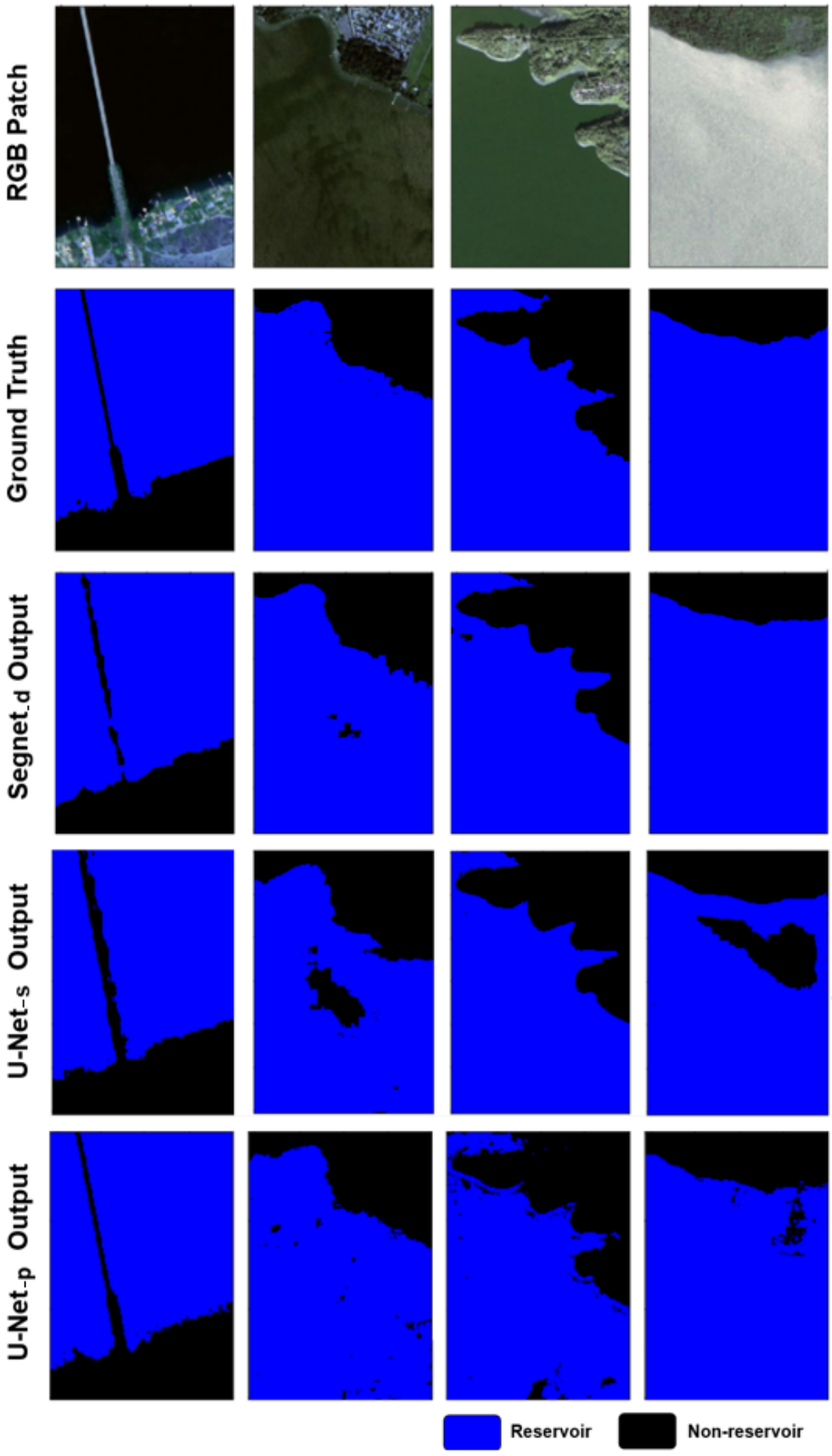

Figure 4.10: *Examples of the test set patches beside their corresponding ground truths and segmentation outputs.*

Figure 4.11: *Examples of non-reservoir water bodies in the collected dataset and corresponding segmentation results.*

the models with overfitting. Furthermore, experiments are on two train sets, a train set containing 70% of data or the over-sampled train set. The over-sampled images are images with at least 200 man-made objects pixels.

The evaluation metrics for the highest performance model constructed using each architecture are presented in Table 4.7. These models are all trained on the over-sampled train set with a learning rate of 0.0001. Furthermore, the augmentation rate in these models is set to 0.7 for each augmentation method, and the dropout regularization is set to 0.0, 0.3, 0.3, and 0.0, respectively. Regarding the F1-score, the best performance belongs to FPN; however, the differences are insignificant. The utilized backbones for each model in the Table are ResNet50, VGG-16, VGG-16, and EfficientNet-B4, respectively. Utilizing the VGG-16 as the PSPNet backbone could significantly improve the model's performance (2.34%). However, the performances of the rest models are slightly affected by changing their backbones ($<$0.73%). This indicates the efficiency of the EfficientNets architectures (the numbers of trainable parameters of EfficientNet-B4, VGG-16, ResNet50, and ResNet101 are about 19, 135, 26, and 45 million, respectively). In our experiments, oversampling images with more than 200 man-made object pixels improved the performances. Moreover, despite the expectation, increasing batch size did not increase the performance metrics in all cases. Adding the dice loss to the focal loss function significantly improved the models' performances. Although increasing the augmentation rate prevented overfitting in some cases, in other cases increasing dropout and augmentation rates were both essential. Though the FPN outperforms the PSPNet, each epoch training time of PSPNet is less than one-third of the FPN. FPN performance in the segmenting test set is presented in Table 4.8.

Moreover, the FPN performance in segmenting RoI of reservoirs located in the countryside and urban areas are computed separately and shown in Table 4.9. Some examples of patches besides their ground truths and segmentation outputs are illustrated in Figure 4.13. This figure illustrates examples of different types of roads, rooftops, and urban and countryside

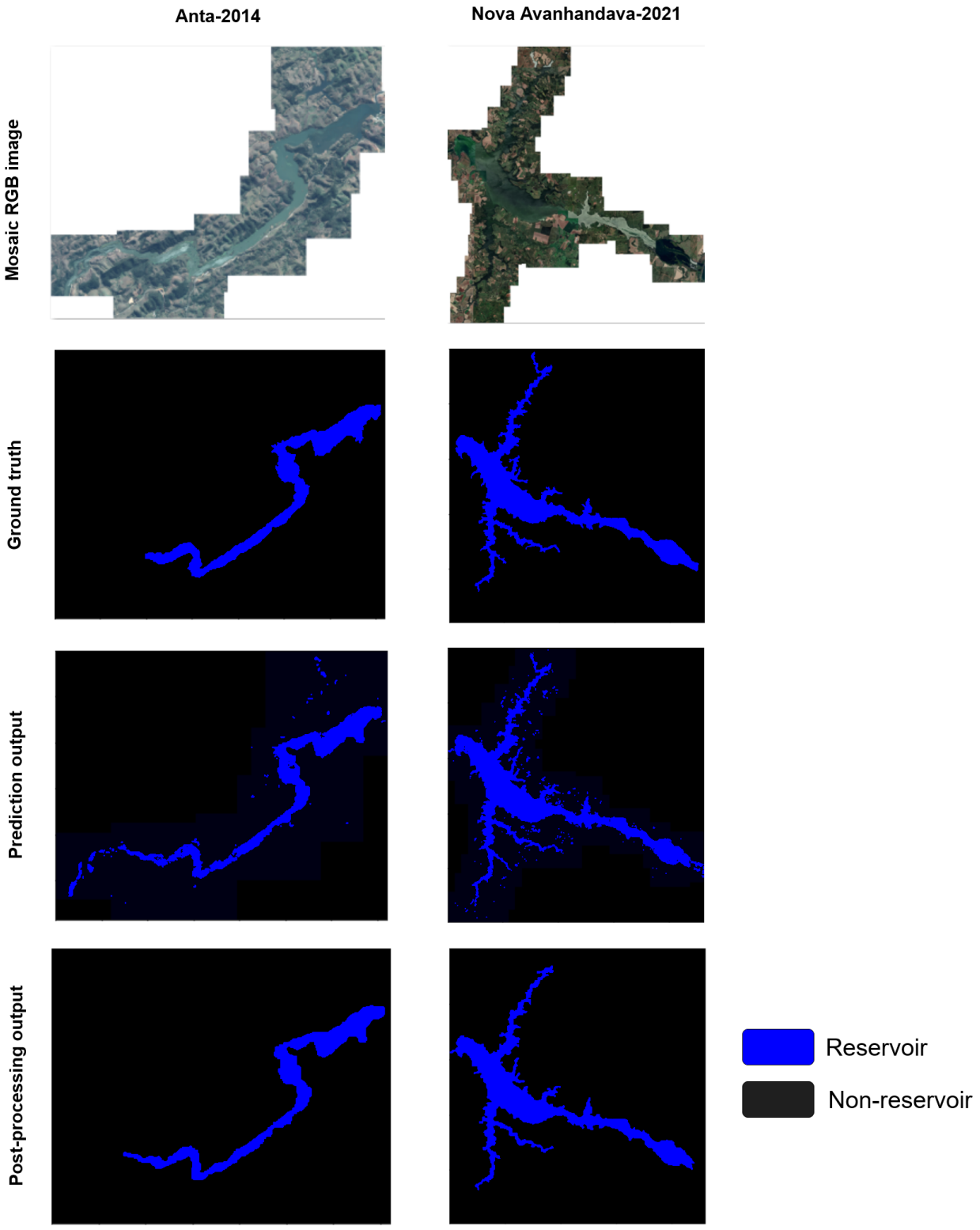

Figure 4.12: *Two examples of produced mosaic images, corresponding ground truths, prediction outputs, and post-processing outputs. Anta-2014 with 11687 × 14430 pixel size, and Nova-2021 with 24830 × 23193 pixel size are depicted in the first and second columns, respectively.*

constructions with different density levels.

Table 4.7: *The highest achieved performances using trained models for Phase-2 semantic segmentation on train and test sets.*

	F1-score	
Model	**Train set**	**Test set**
U-Net	91.64	90.13
PSPNet	91.29	89.58
FPN	92.16	90.32
LinkNet	91.95	90.15

Table 4.8: *FPN performance in segmenting test set to man-made, and non-man-made.*

Class	Precision	Recall	F1-score	Support (N.pixels)
non-man-made	99.52	99.56	99.54	327065669
man-made	81.79	80.43	81.10	8101819

Table 4.9: *FPN performance in segmenting countryside and urban man-made objects. C and U are the abbreviations for countryside and urban.*

	Precision		Recall		F1-score	
Class	C	U	C	U	C	U
non-man-made	99.68	99.39	99.73	99.26	99.71	99.33
man-made	78.70	86.62	75.78	88.75	77.21	87.67

4.3.4 Workflow evaluation

We evaluated the proposed workflow using a dataset collected from the Barra Grande reservoir (Barra). Barra is located in Santa Catarina and Rio Grande do Sul states in Brazil. The collected images belong to 2021, and their spatial resolution is two meters. To evaluate the proposed workflow using the collected data, first, patches with 416 x 608-pixel size are constructed from the mosaic RGB image of Barra. Next, patches are fed to the trained SegNet_d to be segmented into the reservoir and non-reservoir (it took 103 seconds). The SegNet_d performance is evaluated by comparing model outputs with manually produced ground truths. In the next step, the SegNet_d outputs are assembled to be refined using the proposed post-processing stage (it took 100 seconds). The refined reservoir map is used to detach the RoI around Barra. The covered distance from the border of the reservoir is 200 meters. This process takes 219 seconds utilizing the mathematical morphology-based approach and 21 seconds utilizing the polygon approximation-based approach. Finally, patches of detached RoI are fed to the phase-2 pre-trained model to be segmented into man-made and non-man-made (it took 61 seconds). In Table 4.10, the performances of the phase-1 semantic segmentation stage, besides the performance of proposed post-processing, are reported. Table 4.11 shows the evaluation metrics for the phase-2 semantic segmentation stage. Furthermore, some samples of phase-1 and phase-2 semantic segmentation outputs

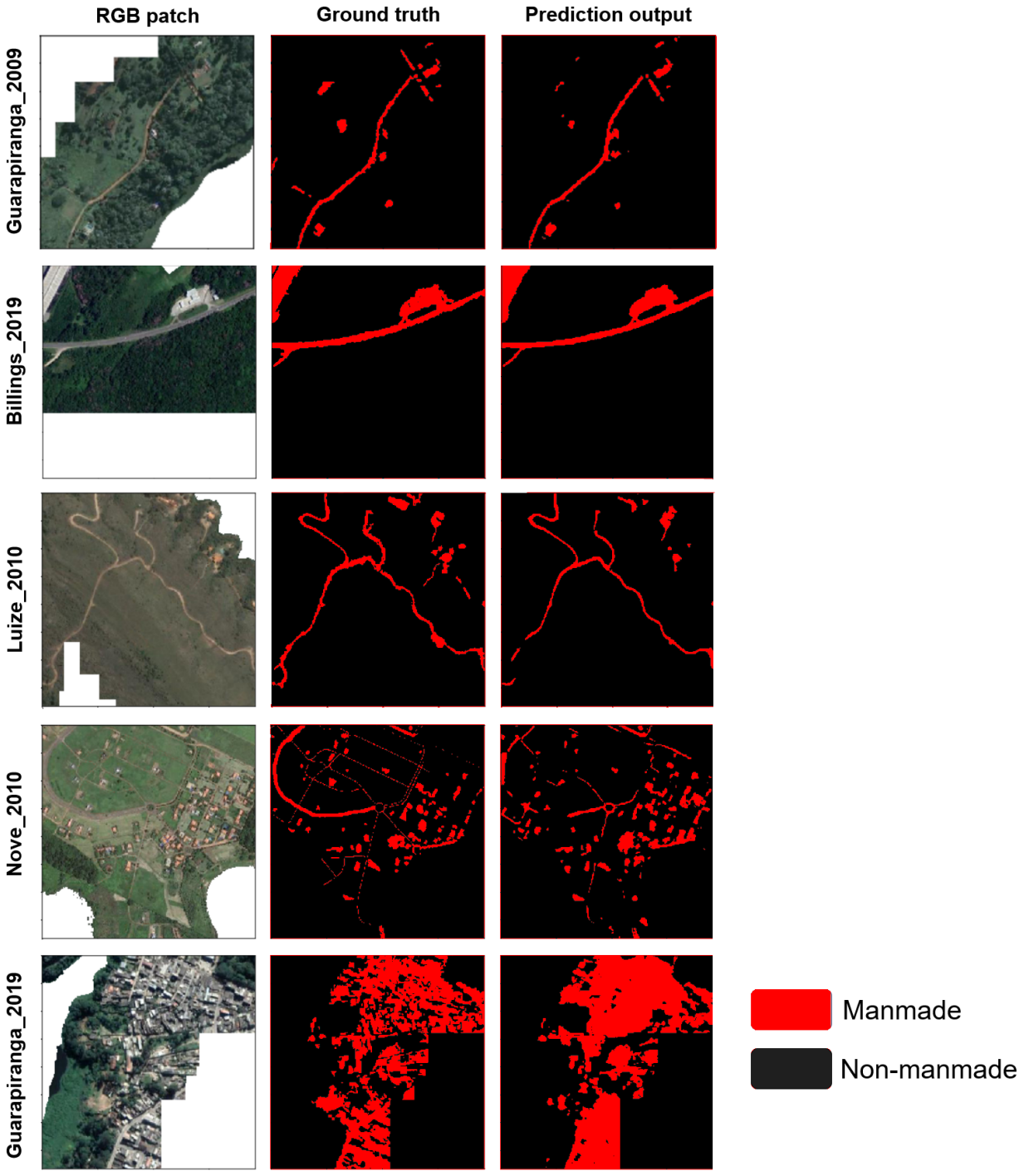

Figure 4.13: *Samples of four studied reservoirs RoI patches beside their corresponding ground truths and prediction outputs.*

are illustrated in Figures 4.14 and 4.15, respectively.

Table 4.10: *Performance of phase-1 semantic segmentation and post processing stages in segmenting Barra dataset to reservoir and non-reservoir.*

	Model			Post-processing		
class	Precision	Recall	F1-score	Precision	Recall	F1-score
non-Reservoir	98.39	96.86	97.62	98.38	98.36	98.37
Reservoir	84.00	91.21	87.45	90.92	91.04	90.98

Table 4.11: *Performance of phase-2 semantic segmentation stage in segmenting Barra RoI to man-made and non-man-made.*

Class	Precision	Recall	F1-score
non-man-made	99.99	99.99	99.99
man-made	73.29	79.43	76.23

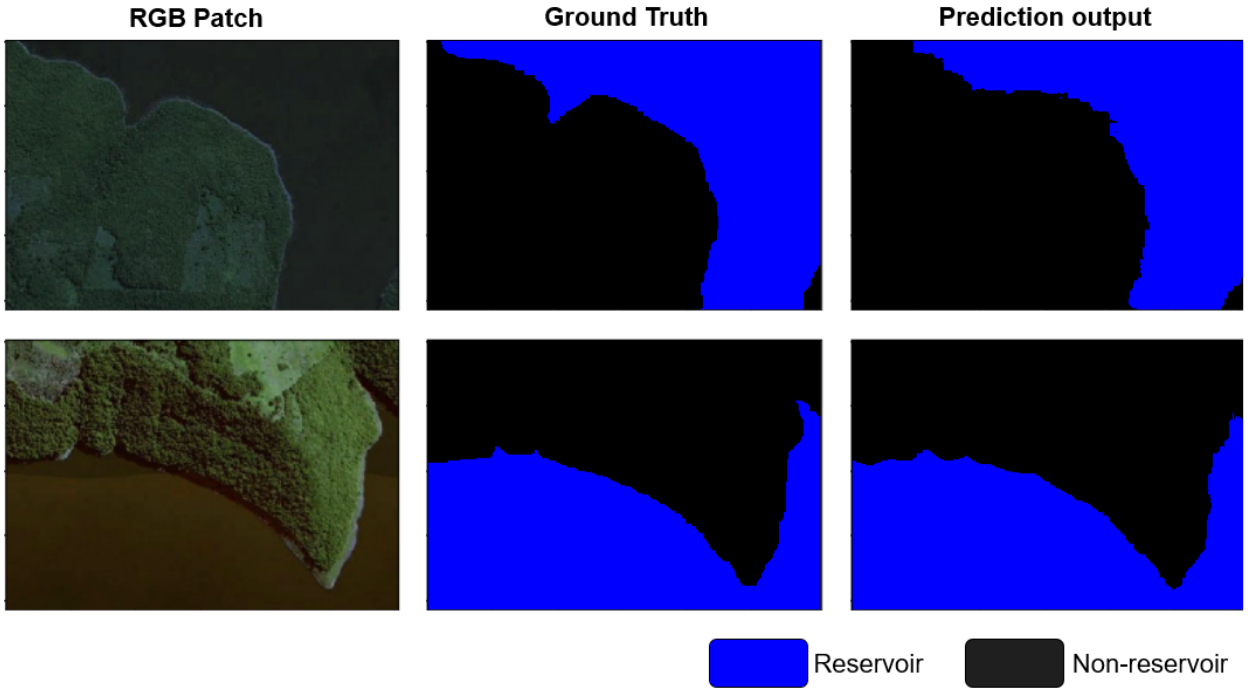

Figure 4.14: *Two samples of Barra phase-1 patches, besides their corresponding ground truths and semantic segmentation results.*

4.3.5 Benchmark

In order to show the effectiveness of our proposed two-phase approach, we applied a single-phase network for semantic segmentation of reservoir, man-made, and else, as the baseline. In this model, the VGG-16 is used as the backbone, the learning rate is set to 0.0001, the number of epochs is set to 150, the early stopping is not applied, and the summation of Dice and Focal losses is used as the loss function. The learning rate is reduced by a factor of 0.2 after every five epochs with no reduction in validation loss down to 10^{-7}. Like the phase-2 training phase, we constructed patches with 384 x 384-pixel size and split them into two sets: train and test.

Since man-made objects inside RoIaR are annotated as man-made and outside as non-man-made (because they are not around the reservoir), the baseline performance is poor (see Table 4.12), as expected. This simple baseline approach illustrates the importance of our proposed two-phase approach.

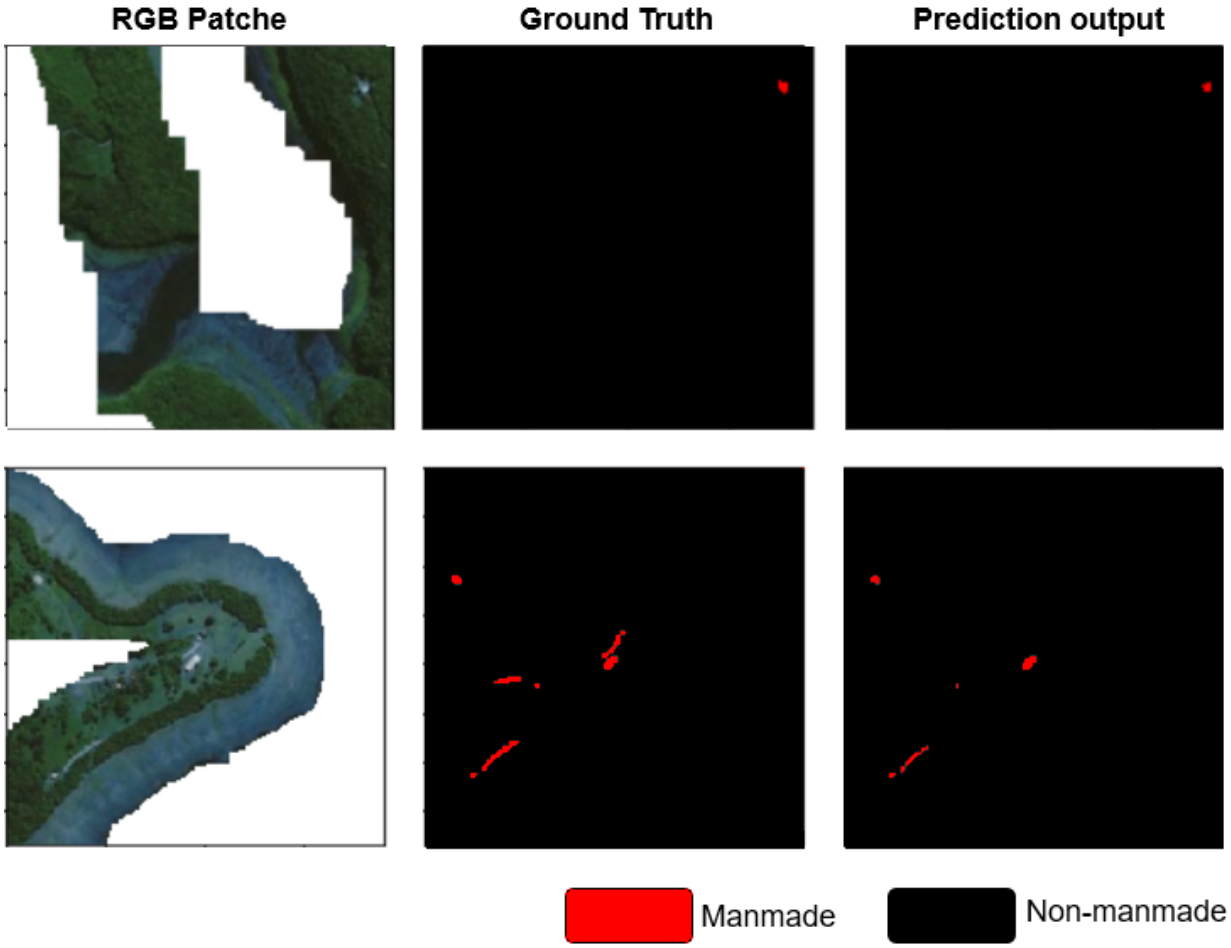

Figure 4.15: *Two samples of Barra RoI patches, their corresponding ground truths, and semantic segmentation results.*

Table 4.12: *U-Net performance in segmenting train and test sets into reservoir, man-made, and else.*

	Precision		Recall		F1-score	
Class	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Reservoir	96.58	96.15	96.52	95.72	96.55	95.94
Man-made	62.17	59.02	50.78	49.13	55.90	53.62
Else	98.64	98.37	98.88	98.69	98.76	98.53

4.4 Discussion

The experimental performance evaluation has addressed the results of phases 1 and 2 of the proposed workflow, workflow validation by an external testing dataset, and the single-phase segmentation benchmark result.

Reservoir segmentation is addressed in phase-1 of the workflow. We trained three U-Net-based models in this phase. The vanilla U-Net was changed to keep the size of each output equal to the corresponding input to produce a pixel-wise classification. Besides, a U-Net with VGG-16 as the backbone was trained. The model over-fitted highly to the train set. Decreasing the number of feature maps in the model (named U-Net_s) caused

performance improvement and fixed the over-fitting issue, as shown in Table 4.3. A SegNet-based architecture was also trained to examine its ability to enhance segmentation outputs. However, it outperformed the U-Net_s slightly (1.23% in F1-Score).

The reservoirs are considered in a broad class called water bodies. In this study, a post-processing stage is proposed to eliminate errors caused by floating vegetation, and delete FP and FN objects caused by spectral similarities between reservoirs and other objects. The proposed post-processing improved the overall accuracy and provided a clear map of the reservoirs, as shown by the examples in Table 4.6 and Figure 4.12.

Phase-2 restricts the segmentation of man-made objects in the RoIaR. Four DL architectures have been evaluated to segment the man-made objects: U-Net, FPN, LinkNet, and PSPNet. This problem typically involves imbalanced data because of government policies to protect such areas besides difficulty in segmenting countryside man-made objects.

In order to address these issues, we tried out the capability of two recommended loss functions (Dice and Focal losses) and the over-sampling strategy. Although Focal loss was reported as the best loss function for segmenting unbalance data, adding Dice loss to the Focal loss significantly improved the performances. The oversampling caused a slight improvement in the performance of some models. We trained each architecture with four different backbones, ResNet50, ResNet101, VGG-16, and EfficientNet-B4. The highest improvement caused by changing the backbone belongs to VGG-16 in PSPNet, 2.24%, whereas changing the backbone in other architectures had a low contribution.

Workflow validation has been carried out using data not seen by the model during training (Barra reservoir, see Section 4.3.4). The validation data included realistic noise and difficulties such as clouds. Despite this, the phase-1 model achieved to 92.54% average F1-score that was even improved to 94.68% by applying post-processing techniques (see Table 4.10). Additionally, the reservoir is in the countryside. The majority of roads are not asphalted, and man-made objects present different visual features from urban areas. Also, there are fewer samples of them in the training data. Accordingly, segmenting them is more complicated compared to urban areas. Nonetheless, the phase-2 model could gain an acceptable performance, as seen in Table 4.11.

We increased the feature maps in the phase-2 trained U-Net-based model and trained that to segment collected data into the reservoir, man-made and non. The data was split into the train and test sets, and no early stopping was applied. Nonetheless, the model man-made F1-score was 35.74 % less than the phase-2 U-Net model.

Chapter 5

Concluding remarks

5.1 Conclusion

The application of OB and DL techniques for mapping the RS imagery into different land cover classes is an important topic in the RS community due to their significant impact on professional urban management. Unauthorized construction close to the reservoirs as important water sources is an urban management issue. In this study, we analyzed monitoring this social phenomenon utilizing OB and DL approaches.

In chapter 3, the OB approach is studied, and a novel pipeline is proposed for segmenting high-resolution RGB images into the reservoir, man-made object, vegetation, and shadow. The suggested approach adopts a multi-phase segmentation. The advantages of the proposed segmentation method are diminishing the need to tune various parameters of the time-consuming segmentation setup and enabling the classifying of reservoir segments just using their spatial information. Spectral similarity analysis and feature extraction steps are proposed to be implemented before defining the rules of the rule-based classifier. These two steps indicate the suitable parameters and thresholds to be used for defining the rules. We calibrated and tested the pipeline utilizing data collected from three zones around the Guarapiranga reservoir. However, despite the OB approach producing promising results, its parameters and rules calibration is still empirical and time-consuming. Therefore, our experiments have shown the need for an end-to-end approach with less user dependency. This conclusion led us to utilize the DL approach for a more generalizing approach.

In chapter 4, we proposed a two-phase workflow to segment man-made objects around reservoirs in an end-to-end procedure. Modifying the U-Net architecture increased the performance of the model in segmenting images into reservoir and non-reservoir. In order to improve produced reservoir maps, a post-processing stage is proposed that, besides increasing the precision metric, its effect is noted by visual evaluation. A small portion of images belongs to the class of man-made objects, specially countryside man-made objects. Nonetheless, we obtained promising results by collecting images of reservoirs mainly located in the countrysides and defining a suitable loss function. The collected RS images have high spatial

resolutions, contain reservoirs with different spectral properties, are located in urban areas as well as the countrysides, and are acquired from different states and seasons. These factors increase the reliability and robustness of constructed models and the proposed workflow. The trained workflow was evaluated with an external testing dataset. Although the collected images are noisy in some areas, the RoIaR is in the countryside, and the RoIaR contains difficult man-made objects for prediction (e.g., roads that are partially covered with high or low density of vegetation), the average F1-scores of phase-1 and phase-2 outputs show the reliability of the prepared workflow. The workflow outperformed significantly in man-made object segmentation compared to the single-phase semantic segmentation benchmark.

5.2 Future work

The availability of larger datasets plays an important role in constructing a higher-performance model. Therefore, collecting data from reservoirs with more geographical and temporal diversity, especially from the countrysides, and adapting available annotated RS images (such as ISPRS Potsdam data [1]) can help in constructing a higher performance and a more generalized model.

The need of pixel-level annotated data in the supervised approach is a challenge in utilizing our proposed method. The semi-supervised approach addresses problems with partially unlabelled data Oliveira et al. [2020]. This can be useful to explore already annotated data and new unlabeled data.

As illustrated in Figure 4.13, the boundary between two objects from two different classes is not segmented well in some cases. In order to generate more accurate segmentation maps, preserving objects' boundary information is important. This issue may be addressed by different approaches in the future works, such as:

- Integrating OB and DL approaches.
- Utilizing the OB approach as a post-processing step for refining DL outputs.
- Applying a sensitive loss function to the border pixels.
- DL architecture modification to embed boundary-aware feature maps.

An important possible application of RoIaR man-made segmentation is the timely detection of unauthorized constructions around the reservoirs. This is a social problem that might lead to serious consequences such as reservoir contamination and dangerous situations for communities living in such places. Unfortunately, if such constructions are not detected in their first stages, and local communities start to live there, it becomes more and more difficult for public services to move such communities. Hence, timely change detection in the

[1] http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html

RoIaR man-made objects is an important application that might rely on the segmentation procedure described in this thesis. Change detection in RS images is an issue that deserves a research work Canty [2019].

Finally, a key issue of remote sensing imagery analysis regards the challenges in analyzing data from different locations and dates. Geographical and atmospheric variations affect the images, and domain adaptation approaches can be developed. This problem has been circumvented in this work by sparse annotation of all considered reservoirs, reflected by our sampling strategy. We are considering other possible domain adaptation approaches, such as few-shot and self-supervised learning Oliveira et al. [2020]. Of course, context-aware networks could be adopted as a possible alternative. This is left as future work since it involves its own challenges.

Appendix A

Paper: Object-based Method for Identifying New Constructions around Water Reservoirs: Preliminary Results

Parts of this thesis are based on the publication Hamidishad and Marcondes Cesar Jr [2019].

Object-based Method for Identifying New Constructions around Water Reservoirs: Preliminary Results

Nayereh Hamidishad
Institute of Mathematics and Statistics
University of Sao Paulo
Email: nhamidi@ime.usp.br

Roberto M. Cesar-Jr
Institute of Mathematics and Statistics
University of Sao Paulo
Email: rmcesar@usp.br

***Abstract*— Identifying new constructions in large cities can be done simply, quickly, and at low cost by applying image processing techniques on time-series remote sensing (RS) images and producing land cover maps. In recent years, object-based (OB) image classification has attracted significant attention as a method for land cover mapping. This method consists of two steps: segmentation and classification. In this research, we will develop a new approach based on image processing techniques to be utilized in the OB classification method for the analysis of urban growth. In this approach, we propose a multi-phase segmentation for the segmentation step and a rule-based method for the classification step. Besides speeding up the process of OB classification, the accuracy of the final preliminary results is another advantage of the proposed approach. Moreover, for collecting RS images, a two-zoom level data collection is adopted using an open source RGB RS database. An important application of analyzing RS images is the detection of non-authorized communities formation around water reservoirs. Therefore, in our preliminary experiments, we selected three different regions around Guarapiranga reservoir in Sao Paulo, Brazil, for collecting our RS images.**

***Index Terms*— Image processing, land cover mapping, remote sensing, object-based method.**

I. INTRODUCTION

RS is a technology that involves the use of space-imaging systems for monitoring earth resources and obtaining information from a target through the analysis of acquired data [1]. By development of RS technology, a large number of satellite and aerial images with high quality have been created. The spatial resolution of some of these RS images is improved to centimeters [2].

One way for obtaining satellite images is by utilizing Google Earth Pro© software[1]. The database of this software contains high-resolution images of many different areas on the earth [4]. Besides containing an open database of RS images, the availability of historical images and flexibility in selecting different zoom levels and spatial resolutions are the other advantages of this software.

Utilizing RS images for land cover mapping has been widely accomplished using pixel-based approaches, where each pixel is independently classified [5]. Even so, with the increase in spatial resolution of satellite images, a single pixel does not capture well the characteristics of targeted objects and it causes a reduction in the accuracy of classification using pixel-based methods [3].

[1]https://www.google.com/earth/desktop/

In contrast with pixel-based methods, OB classification methods are less sensitive to spectral variances within objects and can make use of both object features and spatial relations between objects. Because of these reasons, it has become the main approach in dealing with high-resolution satellite imagery [6]. Many studies compared these two classification methods and concluded that the OB method is more precise in comparison with the pixel-based method [7]. The process of OB classification which is implemented in many studies [8], [9], consists of two steps: segmentation and classification. One of the most commonly used algorithms for segmenting and classifying RS images are multi-resolution segmentation and rule-based classification respectively. However, the multi-resolution segmentation method requires users to determine a set of proper segmentation parameters. In addition, the most important parameter in the multi-resolution segmentation (called scale) can differ between various objects in an image. Moreover, many studies have concluded that to achieve an accurate segmentation result, the multi-scale segmentation method is necessary (i.e., defining a different scale parameter for each class in the scene) [8], [10]. Among the different methods for assessing the values of these parameters, visual assessment is the most common one employed [11]. Therefore, determining and assessing these parameters utilizing the most commonly used methods are very time-consuming. On the other hand, implementing the rule-based method for classifying, especially when the dataset consisting of several large RS images, is also time-consuming. However, utilizing this method may lead to the best results [12]. Additionally, the reservoir objects have similar spectral properties to other objects that cause misclassification of this land cover type. This paper addresses these issues by proposing a new pipeline for the OB classification. It avoids setting a different scale parameter for each class type by implementing a multi-phase segmentation method instead of the multi-scale method. The proposed pipeline also avoids selecting segmentation parameters and classification criteria and threshold values through trial and error on large RS images (that cover larger ground regions and consequently, consist of more objects). Moreover, it classifies the scene in two phases: the reservoir classification and the remained objects classification.

In order to illustrate the performance of the proposed approach, the preliminary results of its application in land

Figure A.1

Appendix B

Paper: An End-to-End Two-Phase Deep Learning-Based Workflow to Segment Man-made Objects Around Reservoirs

Parts of this thesis are based on the publication Hamidishad and Marcondes Cesar Jr [2023].

An End-to-End Two-Phase Deep Learning-Based workflow to Segment Man-made Objects Around Reservoirs

Nayereh Hamidishad [1,*], and Roberto Marcondes Cesar Junior [1]
[1] University of São Paulo, São Paulo, SP, Brazil

February 8, 2023

Abstract

Reservoirs are fundamental infrastructures for the management of water resources. Constructions around them can negatively impact their quality. Such unauthorized constructions can be monitored by land cover mapping (LCM) remote sensing (RS) images. In recent years, deep learning (DL) has attracted considerable attention as a method for LCM the RS imagery and has achieved remarkable success. In this paper, we develop a new approach based on DL and image processing techniques for man-made object segmentation around the reservoirs. In order to segment man-made objects around the reservoirs in an end-to-end procedure, segmenting reservoirs and identifying the region of interest (RoI) around them are essential. In the proposed two-phase workflow, the reservoir is initially segmented using a DL model. A post-processing stage is proposed to remove errors such as floating vegetation. Next, the RoI around the reservoir (RoIaR) is identified using the proposed image processing techniques. Finally, the man-made objects in the RoIaR are segmented using a DL architecture. To illustrate the proposed approach, our task of interest is segmenting man-made objects around some of the most important reservoirs in Brazil. Therefore, we trained the proposed workflow using collected Google Earth (GE) images of eight reservoirs in Brazil over two different years. The U-Net-based and SegNet-based architectures are trained to segment the reservoirs. To segment man-made objects in the RoIaR, we trained and evaluated four possible architectures, U-Net, FPN, LinkNet, and PSPNet. Although the collected data has a high diversity (for example, they belong to different states, seasons, resolutions, etc.), we achieved good performances in both phases. The highest achieved F1-score for the test sets of phase-1 and phase-2 semantic segmentation stages are 96.53% and 90.32%, respectively. Furthermore, applying the proposed post-processing to the output of reservoir segmentation improves the precision in all studied reservoirs except two cases. We validated the prepared workflow with a reservoir dataset outside the training reservoirs. The F1-scores of the phase-1 semantic segmentation stage, post-processing stage, and phase-2 semantic segmentation stage are 92.54%, 94.68%, and 88.11%, respectively, which show high generalization ability of the prepared workflow.

keywords: land cover mapping; deep learning; Google Earth imagery

1 Introduction

Reservoirs reduce the effects of interseasonal and interannual streamflow fluctuations and hence facilitate water supply, hydroelectric power generation, and flood control, to name a few [GBL12]. There is a significant interaction between the environment and reservoirs as essential water resource management tools. For example, reservoirs affect the quality of the water downstream of their dams, and human activities affect the quality of the reservoir's water as well as the chemical and biological processes in it [VB89].

Unauthorized constructions around reservoirs can be considered destructive activities that can be monitored by LCM of RS images. The purpose of this study is to segment man-made objects around reservoirs. However, to reach this aim using an end-to-end workflow, we have to segment the reservoirs and detect the RoI around them besides segmenting the man-made objects.

The pixel-based, object-based (OB), and, recently, DL methods are three different approaches that can be implemented for LCM RS images. Pixel-based methods (e.g., SVM) rely on the spectral signatures of individual pixels, and each pixel is independently classified [Alm12]. With the increase in the spatial resolution of satellite images by improving in RS systems, a single pixel does not

Figure B.1

Bibliography

J. Geng, H. Wang, J. Fan, and X Ma. Deep supervised and contractive neural network for sar image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(4): 2442–2459, 2017. ix, 2

D. C. Duro, S. E. Franklin, and M. G Dubé. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using spot-5 hrg imagery. *Remote sensing of environment*, 118: 259–272, 2012a. ix, 13

A. K. Bhandari, A. Kumar, and G. K Singh. Feature extraction using normalized difference vegetation index (ndvi): A case study of jabalpur city. *Procedia technology*, 6:612–621, 2012. 1

N. E. Rozanda, M. Ismail, and I Permana. Segmentation google earth imagery using k-means clustering and normalized rgb color space. In *Computational Intelligence in Data Mining-Volume 1*, pages 375–386. Springer, 2015. 1

J. B. Campbell and R. H Wynne. *Introduction to remote sensing*. Guilford Press, 2011. 1

Y. T Liau. Hierarchical segmentation framework for identifying natural vegetation: a case study of the tehachapi mountains, california. *Remote Sensing*, 6(8):7276–7302, 2014. 1

J. R Jensen. *Introductory digital image processing: a remote sensing perspective*. Pearson, 2015. 1, 5, 9, 15, 30

L. Zhu, Y. Chen, P. Ghamisi, and J Benediktsson. Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5046–5063, 2018. 1

F. Qi, J. Z. Zhai, and G Dang. Building height estimation using google earth. *Energy and Buildings*, 118:123–132, 2016. 2

J. Zheng, Y. Xi, M. Feng, X. Li, and N Li. Object detection based on bing in optical remote sensing images. In *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on*, pages 504–509. IEEE, 2016. 2

E. Abdelrahim and R. F Mansour. Shadow detection and geo-tagged image information based strategic infrastructure characterization. *arXiv preprint arXiv:1709.08375*, 2017. 2

M. Huang, N. Chen, W. Du, Z. Chen, and J Gong. Dmblc: An indirect urban impervious surface area extraction approach by detecting and masking background land cover on google earth image. *Remote Sensing*, 10(5):766, 2018. 2

R. A Schowengerdt. *Remote sensing: models and methods for image processing*. Elsevier, 2006. 2

H. Gao, C. Birkett, and D. P Lettenmaier. Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research*, 48(9), 2012. 2

L. Votruba and V Broža. *Water management in reservoirs*. Elsevier, 1989. 2

Q. Hu, W. Wu, T. Xia, Q. Yu, P. Yang, Z. Li, and Q Song. Exploring the use of google earth imagery and object-based methods in land use/cover mapping. *Remote Sensing*, 5 (11):6026–6042, 2013a. 2, 8

T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. Van der Meer, H. Van der Werff, F. Van Coillie, et al. Geographic object-based image analysis–towards a new paradigm. *ISPRS journal of photogrammetry and remote sensing*, 87: 180–191, 2014a. 3

L. Qu, Z. Chen, M. Li, J. Zhi, and H Wang. Accuracy improvements to pixel-based and object-based lulc classification with auxiliary datasets from google earth engine. *Remote Sensing*, 13(3):453, 2021. 3, 7, 9

M. D. Hossain and D Chen. Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:115–134, 2019. 3

L. Drăguţ, O. Csillik, C. Eisank, and D Tiede. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88:119–127, 2014. 4, 8

B. A. Johnson and S. E Jozdani. Identifying generalizable image segmentation parameters for urban land cover mapping through meta-analysis and regression tree modeling. *Remote Sensing*, 10(1):73, 2018. 4, 8

E. M. Silveira, S. H. G. Silva, F. W. Acerbi-Junior, M. C. Carvalho, L. M. T. Carvalho, J. R. S. Scolforo, and M. A Wulder. Object-based random forest modelling of aboveground forest biomass outperforms a pixel-based approach in a heterogeneous and mountain tropical environment. *International Journal of Applied Earth Observation and Geoinformation*, 78:175–188, 2019. 4

N. Wu, L. G. T. Crusiol, G. Liu, D. Wuyun, and G Han. Comparing machine learning algorithms for pixel/object-based classifications of semi-arid grassland in northern china using multisource medium resolution imageries. *Remote Sensing*, 15(3):750, 2023. 4

H. Ghanbari, M. Mahdianpari, S. Homayouni, and F Mohammadimanesh. A meta-analysis of convolutional neural networks for remote sensing applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3602–3613, 2021. 4, 10, 48

M. Wurm, A. Droin, T. Stark, C. Geiß, W. Sulzer, and H Taubenböck. Deep learning-based generation of building stock data from remote sensing for urban heat demand modeling. *ISPRS International Journal of Geo-Information*, 10(1):23, 2021. 4

M. E. Malerba, N. Wright, and P. I Macreadie. A continental-scale assessment of density, size, distribution and historical trends of farm dams using deep learning convolutional neural networks. *Remote Sensing*, 13(2):319, 2021. 4, 11

M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N Paragios. Building detection in very high resolution multispectral data with deep learning features. In *2015 IEEE international geoscience and remote sensing symposium (IGARSS)*, pages 1873–1876. IEEE, 2015. 4

B. Neupane, T. Horanont, and J Aryal. Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. *Remote Sensing*, 13(4):808, 2021. 5, 10

B. Hou, Q. Liu, H. Wang, and Y Wang. From w-net to cdgan: Bitemporal change detection via deep learning techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 58 (3):1790–1802, 2019. 5

X. Zhang, L. Han, L. Han, and L Zhu. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Remote Sensing*, 12(3):417, 2020. 5, 10

A. Jacobson, J. Dhanota, J. Godfrey, H. Jacobson, Z. Rossman, A. Stanish, H. Walker, and J Riggio. A novel approach to mapping land conversion using google earth with an application to east africa. *Environmental Modelling & Software*, 72:1–9, 2015. 5, 15

M. H. Almeer. Vegetation extraction from free google earth images of deserts using a robust bpnn approach in hsv space. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(5), 2012. 5, 15, 36

N. Hamidishad and R Marcondes Cesar Jr. Object-based method for identifying new constructions around water reservoirs: Preliminary results. In *Anais Estendidos da XXXII Conference on Graphics, Patterns and Images*, pages 172–175. SBC, 2019. 5, 13, 61

N. Hamidishad and R Marcondes Cesar Jr. An end-to-end two-phase deep learning-based workflow to segment man-made objects around reservoirs. *arXiv preprint arXiv:2302.03282*, 2023. 5, 35, 63

T Blaschke. What's wrong with pixels? some recent developments interfacing remote sensing and gis. *GeoBIT/GIS*, 6:12–17, 2001. 7

S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, 115(5):1145–1161, 2011. 7, 9

Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly, and D Schirokauer. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering & Remote Sensing*, 72(7):799–811, 2006. 7

P. Wang, X. Feng, S. Zhao, P. Xiao, and C Xu. Comparison of object-oriented with pixel-based classification techniques on urban classification using tm and ikonos imagery. In *Geoinformatics 2007: Remotely Sensed Data and Information*, volume 6752, pages 892–899. SPIE, 2007. 7

Y. Zhou, G. Yao, and M Li. Object-oriented technology research of high resolution remote sensing image classification. In *Computer and Computational Sciences (ICCCS), 2015 International Conference on*, pages 119–123. IEEE, 2015. 7

T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. Van der Meer, H. Van der Werff, F. Van Coillie, et al. Geographic object-based image analysis–towards a new paradigm. *ISPRS journal of photogrammetry and remote sensing*, 87: 180–191, 2014b. 7

M Baatz. Multi resolution segmentation: an optimum approach for high quality multi scale image segmentation. In *Beutrage zum AGIT-Symposium. Salzburg, Heidelberg, 2000*, pages 12–23, 2000. 8

E. Guirado, S. Tabik, D. Alcaraz-Segura, J. Cabello, and F Herrera. Deep-learning versus obia for scattered shrub detection with google earth imagery: Ziziphus lotus as case study. *Remote Sensing*, 9(12):1220, 2017. 8

M. Liu, H. Li, L. Li, W. Man, M. Jia, Z. Wang, and C Lu. Monitoring the invasion of spartina alterniflora using multi-source high-resolution imagery in the zhangjiang estuary, china. *Remote Sensing*, 9(6):539, 2017. 8, 9, 16

C. Zhang, G. Li, and W Cui. High-resolution remote sensing image change detection by statistical-object-based method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018a. 8

H. Tong, T. Maxwell, Y. Zhang, and V Dey. A supervised and fuzzy-based approach to determine optimal multi-resolution image segmentation parameters. *Photogrammetric Engineering & Remote Sensing*, 78(10):1029–1044, 2012. 8

Y. Zhang, T. Maxwell, H. Tong, and V Dey. *Development of a supervised software tool for automated determination of optimal segmentation parameters for ecognition*. na, 2010. 8

C. Witharana and D. L Civco. Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure euclidean distance 2 (ed2). *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 108–121, 2014. 8

S. E. Jozdani, M. Momeni, B. A. Johnson, and M Sattari. A regression modelling approach for optimizing segmentation scale parameters to extract buildings of different sizes. *International Journal of Remote Sensing*, 39(3):684–703, 2018. 8

L. Drăguţ, D. Tiede, and S. R Levick. Esp: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6):859–871, 2010. 8

H. Grybas, L. Melendy, and R. G Congalton. A comparison of unsupervised segmentation parameter optimization approaches using moderate-and high-resolution imagery. *GIScience & Remote Sensing*, 54(4):515–533, 2017. 8, 18

M. Belgiu and L Drăguţ. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96:67–75, 2014. 8, 18

Y. Gao, J. F. Mas, N. Kerle, and J. A Navarrete Pacheco. Optimal region growing segmentation and its effect on classification accuracy. *International journal of remote sensing*, 32(13):3747–3763, 2011. 8

D. C. Duro, S. E. Franklin, and M. G Dubé. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using spot-5 hrg imagery. *Remote Sensing of Environment*, 118: 259–272, 2012b. 8

N. Zaabar, S. Niculescu, and M. M Kamel. Application of convolutional neural networks with object-based image analysis for land cover and land use mapping in coastal areas: A case study in ain témouchent, algeria. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:5177–5189, 2022. 8

N. Baghdadi and M Zribi. *Land Surface Remote Sensing in Agriculture and Forest*. Elsevier, 2016. 9

Z. Ziaei, B. Pradhan, and S. B Mansor. A rule-based parameter aided with object-based classification approach for extraction of building and roads from worldview-2 images. *Geocarto International*, 29(5):554–569, 2014. 9, 20

L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y Liu. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277–293, 2017. 9

R. Chen, X. Li, and J Li. Object-based features for house detection from rgb high-resolution images. *Remote Sensing*, 10(3):451, 2018a. 9

X. Lv, Z. Shao, D. Ming, C. Diao, K. Zhou, and C Tong. Improved object-based convolutional neural network (iocnn) to classify very high-resolution remote sensing images. *International Journal of Remote Sensing*, 42(21):8318–8344, 2021. 9

C. D. Rittenhouse, E. H. Berlin, N. Mikle, S. Qiu, D. Riordan, and Z Zhu. An object-based approach to map young forest and shrubland vegetation based on multi-source remote sensing data. *Remote Sensing*, 14(5):1091, 2022. 9

Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4): 541–551, 1989. 9

A. Krizhevsky, I. Sutskever, and G. E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pages 1097–1105, 2012. 10, 11

K. Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 10

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 10

K. He, X. Zhang, S Ren, and J Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 10

J. Long, E. Shelhamer, and T Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 10

V. Badrinarayanan, A. Handa, and R Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv:1505.07293*, 2015. 10, 40

O. Ronneberger, P. Fischer, and T Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 10, 39

R. A. Emek and N Demir. Building detection from sar images using unet deep learning method. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2020. 10

H. B. Kashani, A. Jodeiri, M. M. Goodarzi, and I. S Rezaei. Speech enhancement via deep spectrum image translation network. In *2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME)*, pages 1–7. IEEE, 2019. 10

T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017a. 10, 43

B. Bischke, P. Helber, J. Folz, D. Borth, and A Dengel. Multi-task learning for segmentation of building footprints with deep neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1480–1484. IEEE, 2019. 10

H. Zhao, J. Shi, X. Qi, X Wang, and J Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 10, 43

X. Yuan, Z. Chen, N. Chen, and J Gong. Land cover classification based on the pspnet and superpixel segmentation methods with high spatial resolution multispectral remote sensing imagery. *Journal of Applied Remote Sensing*, 15(3):034511, 2021. 10

R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z Maningo. Object detection using convolutional neural networks. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 2023–2027. IEEE, 2018. 11

L. Deng, J. Li, J. T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, et al. Recent advances in deep learning for speech research at microsoft. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 8604–8608. IEEE, 2013. 11

T Pham. Semantic road segmentation using deep learning. In *2020 Applying New Technology in Green Buildings (ATiGB)*, pages 45–48. IEEE, 2021. 11

A. Abdollahi, B. Pradhan, and A. M Alamri. An ensemble architecture of deep convolutional segnet and unet networks for building semantic segmentation from high-resolution aerial images. *Geocarto International*, pages 1–16, 2020. 11

X. Xie, Y. Zhou, Y. Xu, Y. Hu, and C Wu. Openstreetmap data quality assessment via deep learning and remote sensing imagery. *IEEE Access*, 7:176884–176895, 2019. 11

S. Song, J. Liu, Y. Liu, G. Feng, H. Han, Y. Yao, and M Du. Intelligent object recognition of urban water bodies based on deep learning for multi-source and multi-temporal high spatial resolution remote sensing imagery. *Sensors*, 20(2):397, 2020. 11

M. Li, P. Wu, B. Wang, H. Park, H. Yang, and Y Wu. A deep learning method of water body extraction from high resolution remote sensing images with multisensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3120–3132, 2021. 11

Y. Chen, R. Fan, X. Yang, J. Wang, and A Latif. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water*, 10(5):585, 2018b. 11

X. Zhang, B. Wu, L. Zhu, F. Tian, M. Zhang, et al. Land use mapping in the three gorges reservoir area based on semantic segmentation deep learning method. *arXiv preprint arXiv:1804.00498*, 2018b. 11

B. Aryal, S. M. Escarzaga, S. A. Vargas Zesati, M. Velez-Reyes, O. Fuentes, and C Tweedie. Semi-automated semantic segmentation of arctic shorelines using very high-resolution airborne imagery, spectral indices and weakly supervised machine learning approaches. *Remote Sensing*, 13(22):4572, 2021. 11

A. Van Soesbergen, Z. Chu, M. Shi, and M Mulligan. Dam reservoir extraction from remote sensing imagery using tailored metric learning strategies. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. 11

K. Makantasis, K. Karantzalos, A. Doulamis, and K Loupos. Deep learning-based man-made object detection from hyperspectral data. In *International symposium on visual computing*, pages 717–727. Springer, 2015. 11

M. Papadomanolaki, M. Vakalopoulou, and K Karantzalos. A novel object-based deep learning framework for semantic segmentation of very high-resolution remote sensing data: Comparison with convolutional and fully convolutional networks. *Remote Sensing*, 11(6): 684, 2019. 12

A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A Alamri. Multi-object segmentation in complex urban scenes from high-resolution remote sensing data. *Remote Sensing*, 13(18):3710, 2021. 12

J. Yu, P. Zeng, Y. Yu, H. Yu, L. Huang, and D Zhou. A combined convolutional neural network for urban land-use classification with gis data. *Remote Sensing*, 14(5):1128, 2022. 12

E. Manos, C. Witharana, M. R. Udawalpola, A. Hasan, and A. K Liljedahl. Convolutional neural networks for automated built infrastructure detection in the arctic using sub-meter spatial resolution satellite imagery. *Remote Sensing*, 14(11):2719, 2022. 12

G. J. Hay and G Castilla. Geographic object-based image analysis (geobia): A new name for a new discipline. In *Object-based image analysis*, pages 75–89. Springer, 2008. 13

H. M. A. Van der Werff and F. D Van der Meer. Shape-based classification of spectrally identical objects. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2):251–258, 2008. 13

T Blaschke. Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16, 2010. 13

H. Zhang, J. E. Fritts, and S. A Goldman. Image segmentation evaluation: A survey of unsupervised methods. *computer vision and image understanding*, 110(2):260–280, 2008. 14

V. Visser, B. Langdon, A.D. Pauchard, and David M Richardson. Unlocking the potential of google earth as a tool in invasion science. *Biological invasions*, 16(3):513–534, 2014. 15, 36

Z. Guo, X. Shao, Y. Xu, H. Miyazaki, W. Ohira, and R Shibasaki. Identification of village building via google earth images and supervised machine learning methods. *Remote Sensing*, 8(4):271, 2016. 15

D Potere. Horizontal positional accuracy of google earth's high-resolution imagery archive. *Sensors*, 8(12):7973–7981, 2008. 15

U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M Heynen. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information. *ISPRS Journal of photogrammetry and remote sensing*, 58(3-4):239–258, 2004. 16, 18

eCognition Developer 9.3.1 User Guide. Trimble Germany GmbH, 2018. 18

R. Pande-Chhetri, A. Abd-Elrahman, T. Liu, J. Morton, and V. L Wilhelm. Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing*, 50(1):564–576, 2017. 20

Q. Hu, W. Wu, T. Xia, Q. Yu, P. Yang, Z. Li, and Q Song. Exploring the use of google earth imagery and object-based methods in land use/cover mapping. *Remote Sensing*, 5 (11):6026–6042, 2013b. 20

M Osava. Guarapiranga. https://web.archive.org/web/20080528160544/http://www.ipsnews.net/news.asp?idnews=38094, 2008. 22

Z. Miller, J. B. Sharma, J. NolanZac Miller, and B Rumiser. eCognition segmentation and supervised classification. http://faculty.ung.edu/zmiller/4350k/labs/ecognitionsegmentationandsupervisedclassification.html, 2013. 23

R. G. Congalton and K Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2008. 30

T. Lillesand, R. W. Kiefer, and J Chipman. *Remote sensing and image interpretation*. John Wiley & Sons, 2014. 30

J. Qian, M. Xia, Y. Zhang, J. Liu, and Y Xu. Tcdnet: Trilateral change detection network for google earth image. *Remote Sensing*, 12(17):2669, 2020. 36

D. H. Douglas and T. K Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, cartographica. In *The International Journal for Geographic Information and Geovisualization*, pages 112–122, 1973. 38

M Cubes. A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques. New York: Association for Computing Machinery*, pages 163–69, 1987. 38

L. D. F. D. Costa and R Marcondes Cesar Jr. *Shape analysis and classification: theory and practice*. CRC press, Inc., 2000. 38

S. Ioffe and C Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 40

L. Mou and X. X Zhu. Rifcn: Recurrent network in fully convolutional network for semantic segmentation of high resolution remote sensing images. *arXiv preprint arXiv:1805.02091*, 2018. 42

A. Chaurasia and E Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2017. 43

T. Y. Lin, P. Goyal, R. Girshick, K. He, and P Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017b. 44

T. A Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. page 1–34, 1948. 44

L. R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297–302, 1945. 44

S Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7. IEEE, 2020. 44

D. P. Kingma and J Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 46

H. N. Oliveira, E. Ferreira, and J. A Dos Santos. Truly generalizable radiograph segmentation with conditional domain adaptation. *IEEE Access*, 8:84037–84062, 2020. 58, 59

M Canty. *image analysis, classification and change detection in remote sensing: with algorithms for python*. CRC press, 2019. 59