Semantic Segmentation of Land Cover in Multisource Aerial Imagery using U-Net

1st Sayid Rayhan Mulachela
School of Computing
Telkom University
Bandung, Indonesia
rayhanmulachela@student.telkomuniversity.ac.id

2nd Erwin Budi Setiawan
School of Computing
Telkom University
Bandung, Indonesia
erwinbudisetiawan@telkomuniversity.ac.id

3rd Gamma Kosala
School of Computing
Telkom University
Bandung, Indonesia
gammakosala@telkomuniversity.ac.id

***Abstract*— Accurate land cover segmentation in aerial imagery is vital for environmental monitoring and greenhouse gas (GHG) emissions assessment. This study applies the U-Net model for semantic segmentation of land cover types, including buildings, forests, water bodies, and roads, to analyze their impact on GHG emissions. Using the LandCover.ai dataset (1,200 images) and a supplementary dataset from Java, Indonesia (400 images), the research evaluates U-Net's performance at the pixel level. The dataset was split into 70% training (840 images), 15% validation (180 images), and 15% testing (180 images). Metrics such as Intersection over Union (IoU) and Dice Coefficient were used for evaluation. The model achieved a mean IoU of 0.81 and a Dice Coefficient of 0.80 on the primary dataset, while performance declined with the Java dataset (mIoU 0.72, Dice 0.70), indicating generalization challenges. Data augmentation improved results to an mIoU of 0.82 and Dice 0.81. These findings highlight U-Net's potential in remote sensing for environmental monitoring and climate change mitigation strategies.**

Keywords*— aerial imagery, computer vision, semantic segmentation, remote sensing, U-Net.

I. Introduction

Accurate LULC monitoring is critical for effective carbon stock management and reducing GHG emissions, where nations in United Nations have international agreement in Kyoto Protocol [1], [2], [3]. In addition to being time-consuming and labor-intensive, traditional methods that rely on manually classifying satellite photos are also prone to human error, which makes them unsuitable for the high precision required in extensive environmental evaluations.[4]. A potential and more effective alternative is offered by recent developments in remote sensing and sophisticated image analysis methods, which is semantic segmentation. Semantic segmentation enables precise pixel-level classification of land cover types—such as buildings, forests, water bodies, and roads—facilitating comprehensive environmental analysis and informed decision-making[5], [6].

However, semantic segmentation of aerial imagery faces challenges due to high variability in lighting conditions, diverse vegetation types, intricate structures, and varied geographical features. Accurate categorization is made more difficult by elements like noise, occlusions, and uneven image resolutions[7]. Traditional pixel-based classification methods, including Support Vector Machines (SVM), Random Forests (RF), and Maximum Likelihood Estimation (MLE), often yield mean Intersection over Union (mIoU) values below 30% and overall accuracy (OA) below 75%, which are inadequate for environmental monitoring needs [7]. This gap highlights the need for more advanced segmentation techniques, such as deep learning-based models like U-Net, which have demonstrated superior performance in handling complex segmentation tasks [8].

Deep learning models, particularly Convolutional Neural Networks (CNNs) like U-Net, excel in semantic segmentation by capturing both local and global contextual information through their encoder-decoder architectures with skip connections [8]. U-Net's ability to extract hierarchical features and maintain spatial information makes it highly effective for pixel-level classification in remote sensing applications [9]. Comparative studies have shown that U-Net outperforms traditional segmentation models, achieving higher mIoU and OA values [9], [10].

This research is aiming to leverage the U-Net model's capabilities in pixel-based segmentation, by using in our aerial imagery dataset. The main objective includes assessing the model's performance in classifying land cover types from high-resolution imagery, comparing its effectiveness with another imagery but with less resolution (our secondary imagery) to know what the effect will be on different source of dataset. This research will also conduct different scenarios like different optimizer, data augmentations scenario, and different use of dataset to get insight of where is the most important aspect that affecting the model performance. Detailed and precise LULC maps contribute to understanding how different land cover types influence GHG emissions, informing policymakers and environmental managers in developing targeted strategies for land use planning, conservation efforts, and emission reduction [11]. The performance evaluation of this model across diverse datasets and scenarios will provide valuable insights for future research and practical applications in environmental monitoring systems.

The remainder of this paper is organized as follows: Chapter 2 : the results and discussion, including segmentation performance metrics and time evaluations. The research will then concluded in Chapter 3 with a summary of the main conclusions, the research limitations, and recommendations for further research.

II. Methodology

The following flowchart illustrates the system flow that the author implements in this research. The process flow can be seen in Fig. 1, and the explanation of the steps is provided below.