Improving Land Cover Segmentation Using Multispectral Dataset

by Herlawati Herlawati

Submission date: 13-Nov-2023 10:35PM (UTC+0700)
Submission ID: 2226769984
File name: turnitin-icic_aptikom_2023-Herlawati_tim_promotor.docx (1.04M)
Word count: 1888
Character count: 10525

Improving Land Cover Segmentation Using Multispectral Dataset

***Abstract*—Computer vision has been used in many areas such as medical, transportation, military, geography, etc. The fast development of sensor devices inside camera and satellite provides not only red-greed-blue (RGB) images but also multispectral dataset with some channels including RGB, infrared, short-wave, and thermal wave. Most of the dataset is panchromatic (black and white) and RGB, for example Google Map and other satellite-based map applications. This study examines the effects of multispectral dataset for semantic segmentation of land cover. The comparison between RGB with band 2 to band 7 of Landsat 8 Satellite shows an improvement of accuracy from 90.283 to 94.473 for U-Net and from 91.76 to 95.183 for DeepLabV3+. In addition, this research also compares two well-known semantic segmentation methods, namely U-Net and DeepLabV3+, that shown that DeepLabV3+ outperformed U-Net regarding to speed and accuracy. Testing was conducted in the Karawang Regency area, West Java, Indonesia.**

Keywords—DeepLabV3+, Semantic Segmentation, Landsat, MATLAB, Deep Learning

I. INTRODUCTION

Computer vision is a branch of Artificial Intelligence (AI) that aims to mimic human capabilities in recognizing image patterns. Neural Networks, the fundamental building blocks of Deep Learning, have been utilized extensively in pattern recognition, gradually replacing models based on mathematical and statistical principles, such as Local Binary Pattern Histogram [1], Adaptive Feature Fusion [2], and others for face recognition. Rapid advancements in hardware have facilitated pattern recognition computations based on neural networks. Graphics Processing Units (GPUs) are quite effective in supporting neural networks with many layers, which are more commonly known as Deep Learning.

However, there exists a gap between the field of computer science and Remote Sensing and Geographic Information Systems (RS-GIS), where the use of multispectral satellite imagery is still limited among computer science researchers. This is evident from the scarcity of multispectral images in benchmarking sites. Most researchers still rely on Red, Green, and Blue (RGB) and grayscale images.

Deep Learning is very practical considering the preprocessing combined within a single neural network model. These preprocessing leverages convolution and has proven to be quite effective in its implementation under the name Convolutional Neural Networks (CNN). This method utilizes a set of filters to extract features from the images to be processed [3].

Research continues to develop CNN with several modifications to address issues that arise in the method [4]–[8], one of which is the vanishing gradient problem due to the large number of filtering effects on the convolution side. This issue can be addressed with the Residual Networks (ResNet) method, allowing layers to be increased to hundreds in a ResNet [9].

There are three problems in computer vision, namely classification, object detection, and segmentation. If classification aims to categorize images according to the training data, object detection will provide notifications (usually in the form of bounding boxes) around the detected objects. The research we are conducting focuses on segmentation, where its classification results are at the pixel level of the image with multiple colors representing their classes. All three of these problems still use the same method, which is convolution-based (CNN) known as U-Net [10]. Furthermore, other modifications for semantic segmentation have been researched [11]–[15].

This research aims to contribute to other researchers who want to apply multispectral datasets and determine which bands are suitable for use in semantic segmentation. This research also can serve as a reference for remote sensing researchers to leverage Deep Learning in the segmentation process. Additionally, the comparison between the U-Net and DeepLabV3+ models can serve as a reference for their implementation in the field, along with their pros and cons.

The rest of the paper is organized as follows. After detailing the dataset and methods, the results are analyzed and discussed, with a focus on the multispectral dataset used. Subsequently, the conclusion provides insights from the findings of this research.

II. MATERIALS AND METHODS

A. Datasets

For the dataset, all available bands were downloaded from the United States Geological Survey (USGS) [16] website for the Jakarta Metropolitan Region (JABOTABEK), which were then cropped to cover the entire area of Karawang Regency. Figure 1 shows the capture area (referred to as a 'tile').

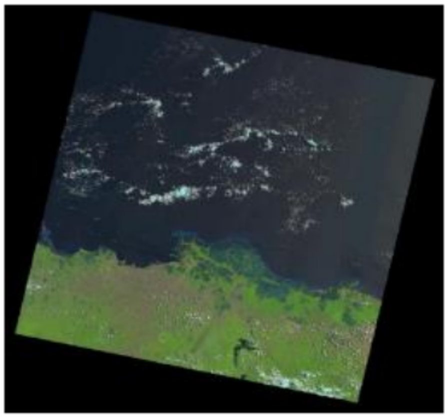

Fig. 1. The region downloaded (Tile) of Jakarta Metropolitan Region

To produce clear satellite captures, it is necessary to find a date during the dry season with minimal cloud cover. Figure 2 shows the Geo TIFF files for the required channels, which were taken in July 2021.

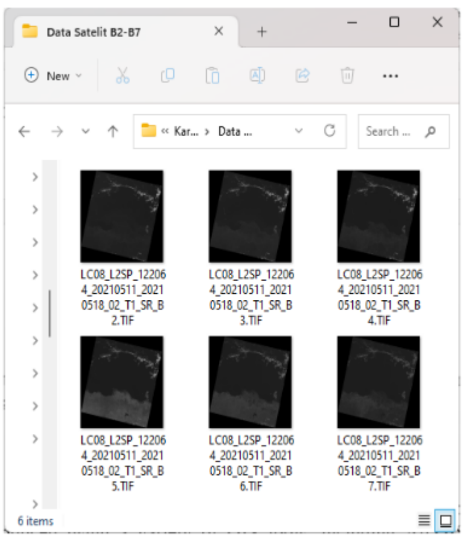

Fig. 2. Landsat capture data for July 2021

Landsat 8 has a resolution of 30 meters and is considered as Level 1 according to the standards [16], [17]. Considering that this research only investigates the effects of multispectral with RGB, the land cover to be segmented is urban, vegetation, and water. To calculate accuracy, ground truth data is required, representing the actual conditions of the urban, vegetation, and water segment classes, using the TerrSet software, which utilizes a semi-automatic process, namely Iterative Self-Organizing Clustering (ISOCLUST) and the RECLASS function to manually reclassify them into three classes [18]. This research uses MATLAB as the programming language and Graphic User Interface (GUI) to facilitate experiments.

B. U-Net

U-Net is a semantic segmentation model that utilizes convolutional methods in CNN. The principle is to perform encoding processes with multiple filtrations followed by decoding, which is the reverse of convolution (deconvolution), where at each level, a copy and crop process is performed between the encoder and decoder sides to refine the resulting image into segments that match the training data (Figure 3).

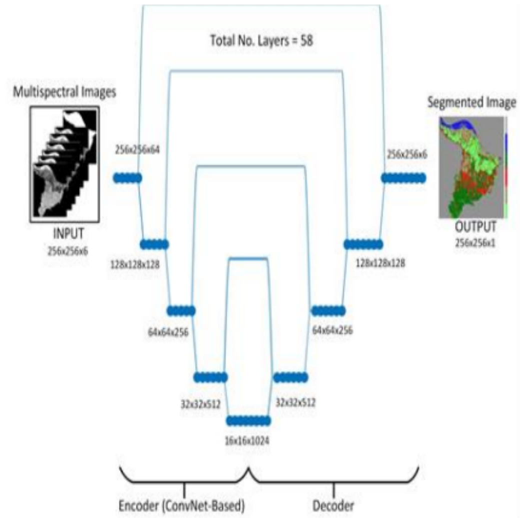

Fig. 3. U-Net Architecture

Equation 1 illustrates the principle of convolution with a 3x3 square filter. G, H, and F are convolution operator, image matrix, and filter, respectively. Variable k is the size of filter.

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v]F[i-u,j-v] \qquad (1)$$

B. DeepLabV3+

One of the main challenges of U-Net, based on its foundation in CNN, is the issue of vanishing gradients. This occurs when attempts to improve accuracy by adding layers result in a slow convolution process. This can be especially time-consuming for large-sized images. Therefore, other models utilize ResNet, which performs convolution processes in its residual path rather than the main path. As a result, the process is faster, and the addition of many layers does not lead to vanishing gradients. The ResNet versions have many layers, ranging from 50 to 101, compared to CNN, which can only be developed up to 19 layers in the Visual Geometry Group (VGG19).

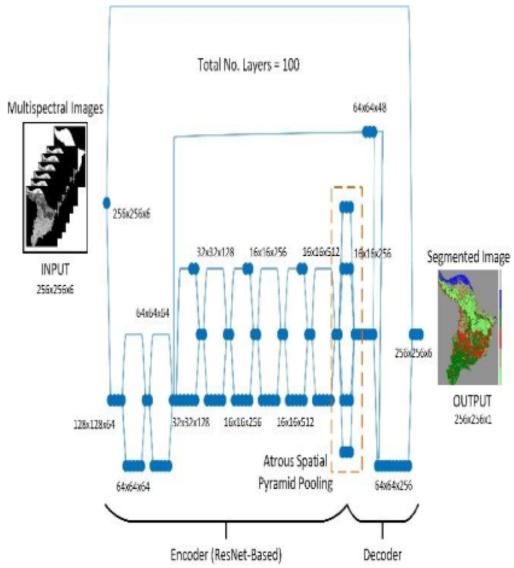

Fig. 4. DeepLabV3+ Architecture

DeepLab uses ResNet on both the encoder and decoder sides. Subsequently, several versions of DeepLab emerged, including DeepLabV3+, which implements Atrous Spatial Pyramid Pooling (ASPP) to expedite the process and address the classification issues of objects with significantly different sizes from one another.

The training process is conducted using a MATLAB live script. The process takes several days. Figure 5 shows the live script for both U-Net and DeepLabV3+ training processes.

The live script has the capability to run program code alongside the display of output and other information, both in text and image formats. The training results are U-Net and DeepLabV3+ models ready for land cover segmentation. A graphical user interface (GUI) was created to facilitate both experiments and user usage.

The GUI used for inference contains buttons to input both RGB and multispectral images, followed by ground truth images for accuracy calculations. The output includes land cover segmentation as well as accuracy and percentage for each segment class.

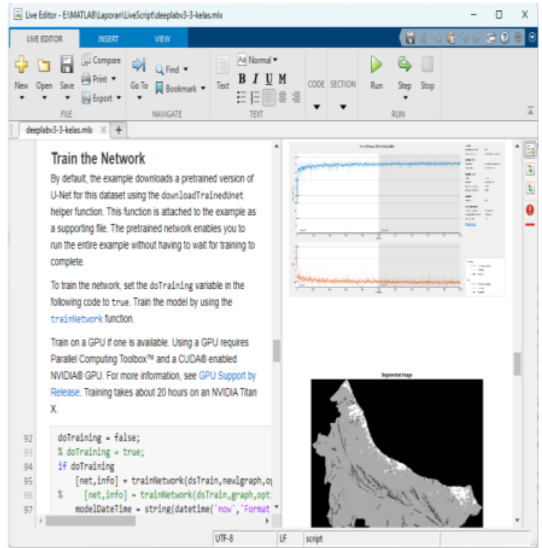

Fig. 5. Live Script for Training

Two trained models are generated, namely U-Net and DeepLabV3+, which are then used for inference on two types of data, namely RGB and multispectral data, before being checked for accuracy based on their respective ground truth datasets.

III. Result and Discussion

The training of the U-Net and DeepLabV3+ models took approximately 4 hours. The accuracy obtained for U-Net and DeepLabV3+ is 94% and 95%, respectively.

A. U-Net Segmentation Result

RGB satellite images are images that can be seen by the human eye. In Landsat 8, RGB is a combination of Band 4, Band 3, and Band 2. Figure 6 shows the segmentation accuracy based on RGB images and multispectral images in the U-Net model.

The testing results with U-Net show that multispectral images yield an accuracy that is 4 percent better than RGB. The urban and water segments, based on the experimental results, had the most significant impact, resulting in a decrease in accuracy. Table 1 shows the recapitalization of U-Net testing using RGB and Multispectral dataset.

TABLE I. U-Net Segmentation

No.	Dataset	Accuracy (%)	Speed
1.	Multispectral	94.473	5 minutes
2.	RGB	90.283	5 minutes

RGB is slightly less accurate in detecting urban and water areas. The accuracy of RGB images is more than 4% lower than that of multispectral datasets.

The training results can be applied to other regions in Indonesia. For more segment classes such as wetlands, barren land, agriculture, forest, and others, a training process is required. This also includes regions with a subtropical climate and four seasons.

The GUI facilitates the input and output processes. Conversion into an executable program (EXE) can be done using the deployment module in MATLAB.

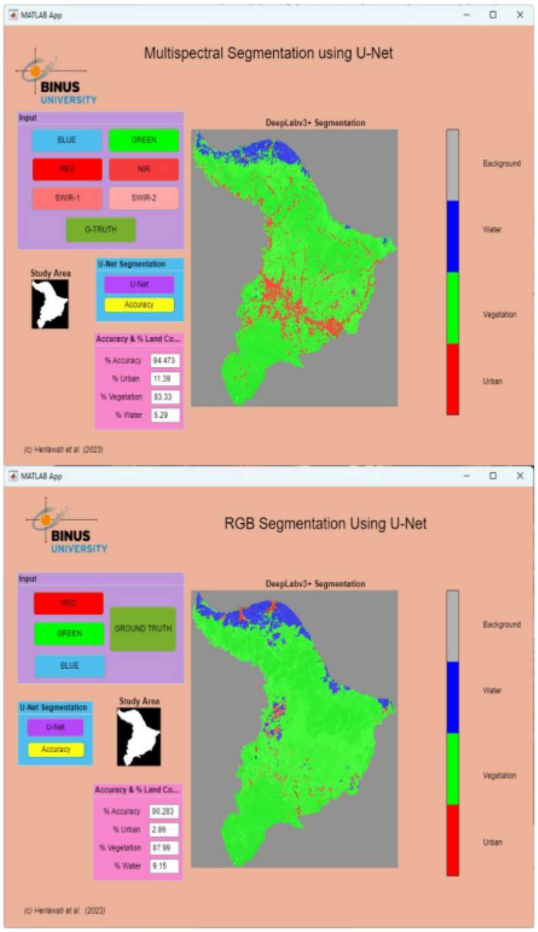

Fig. 6. U-Net Segmentation Result: Multispectral (above) and RGB (below)

B. *DeepLabV3+ Segmentation Result*

The testing results (Figure 7) with DeepLabV3+ show that multispectral images yield an accuracy that is 4 percent better than RGB. The urban and water segments, based on the experimental results, had the most significant impact, resulting in a decrease in accuracy.

With the same GUI, the process button directs to a MAT file, which is the result of training DeepLabV3+. For multispectral images, bands 2, 3, 4, 5, 6, and 7 represent blue, green, red, near-infrared, short-wave infrared-1 (SWIR-1), and short-wave infrared-2 (SWIR-2) images, respectively.

For RGB data, bands 4, 3, and 2 represent the red, green, and blue channels. Unlike U-Net, which takes about 5 minutes, DeepLabV3+ is much faster, taking less than 1 minute. Unlike U-Net, DeepLabV3+ with RGB is still able to detect most of the urban classes. In addition, DeepLabV3+ shows 1% higher accuracy compared to U-Net.

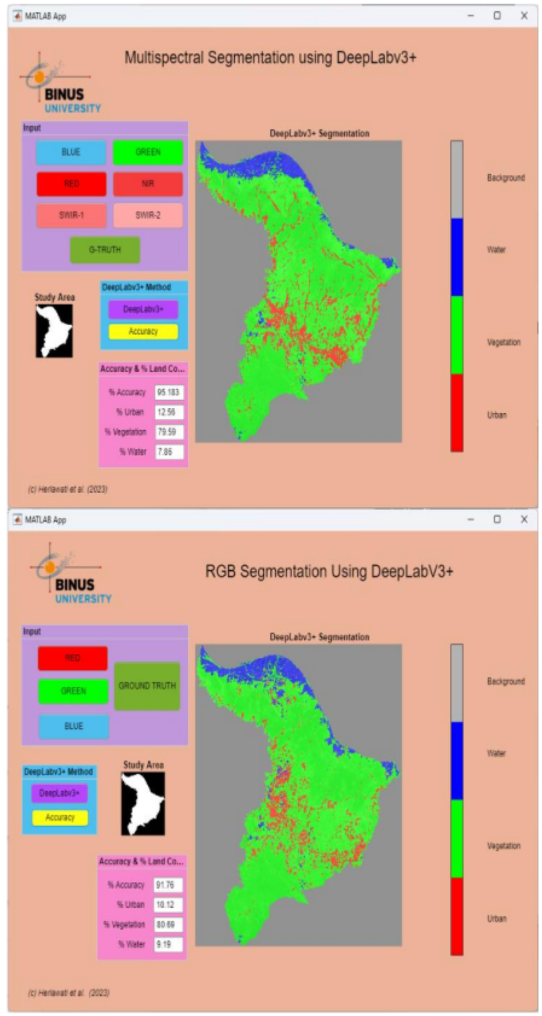

Fig. 7. DeepLabV3+ Segmentation Result: Multispectral (above) and RGB (below)

TABLE II. DEEPLABV3+ SEGMENTATION

No.	Dataset	Accuracy (%)	Speed
1.	Multispectral	95.183	30 second
2.	RGB	91.76	30 second

Although DeepLabV3+ has a slight advantage in terms of accuracy, the inference process shows that U-Net is four to five times slower than DeepLabV3+, depending on the user's processor capacity. Therefore, DeepLabV3+ is more suitable for satellite images implementation compared to U-Net, which is indeed suitable for small-sized images such as medical images (X-ray, MRI, etc.).

IV. CONCLUSIONS

The availability of satellite imagery should be utilized for land management. One of the advantages of satellite imagery is its sensors, which have many specific functions. Several channels in Landsat 8 each have the capability to capture specific frequencies, not only the frequency range visible to the human eye but also infrared, short-wave infrared, and thermal. RGB has an advantage because most cameras can capture this kind of images, but experimental results show that multispectral images have better accuracy than RGB images. Additionally, DeepLabV3+, not only having better accuracy than U-Net, also has the advantage of inference speed. Future research needs to produce proposed models or additional features that outperform both DeepLabV3+ and U-Net.

ACKNOWLEDGMENT

The authors thank to Universitas Bhayangkara Jakarta Raya for funding this research, Universitas Islam 45 Bekasi, and reviewers for suggestion and comments.

Improving Land Cover Segmentation Using Multispectral Dataset

ORIGINALITY REPORT

4%	3%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	rua.ua.es Internet Source	1%
2	export.arxiv.org Internet Source	1%
3	www.researchgate.net Internet Source	1%
4	Fata Nidaul Khasanah, Rahmadya Trias Handayanto, Herlawati Herlawati, Djuni Thamrin, Prasojo Prasojo, Erik Saut H Hutahaean. "Decision Support System For Student Scholarship Recipients Using Simple Additive Weighting Method with Sensitivity Analysis", 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020 Publication	1%
5	mdpi.com Internet Source	1%
6	mdpi-res.com Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		