www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 376-379
e-ISSN : 2583-1062
Impact Factor : 7.001

SATELLITE MAP CLASSIFICATION AND IMAGE CAPTIONING USING DEEP LEARNING

Danish Ahmed N F[1], Dr. M. Praneesh[2]

[1]UG Student, Computer Science With Data Analytics, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamilnadu, India.

[2]Assistant Professor, Computer Science With Data Analytics, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamilnadu, India.

DOI: https://www.doi.org/10.58257/IJPREMS39666

ABSTRACT

The presented research here proposes an end-to-end model to perform semantic segmentation and autopillar satellite image captioning in order to research land cover categories effectively. Building upon DeepLabV3+ for pixel classification and BLIP (Bootstrapped Language Image Pretraining) for captioning natural languages, the model classifies various categories of land such as agricultural lands, urban land, water body, and industrial lands through detection and marking.

The segmentation model was trained on the DeepGlobe dataset with a class-based color map, captioning was enhanced by two methods:(1) model-generated captions via BLIP and (2) land distribution-grounded text summaries of segmented masks. The work was completed within a low-resource environment on Google Colab and was optimized to support training checkpoints and memory-conserving data management. Evaluation metrics were IoU for segmentation performance and qualitative analysis for captioning. Results demonstrate successful segmentation with clear class boundaries and relevant context captions improving interpretability of satellite data. The study presents a cost-effective and scalable application for environmental observation, urban development, and geographical studies.

Keywords: Satellite Segmentation, DeepLabV3+, BLIP, Image Captioning, Land Cover Analysis, Remote Sensing

1. INTRODUCTION

Satellite images have proved to be an indispensable tool in environmental monitoring, urban planning, and land use mapping. Deep learning has made possible much enhanced extraction of informative content from high-resolution satellite images. Semantic segmentation makes it possible to detect the land cover classes with precision at the pixel level and offer insight into geographical processes and human activity.

Parallel to this, image captioning models have also been found to be effective aids for generating human-readable visual descriptions of content, ensuring accessibility and understanding of complex data. Satellite image analysis research today is focused more on segmentation accuracy improvement and interpretability improvement through AI-based summarization techniques.

Deep frameworks like DeepLabV3+ have shown excellent results in semantic segmentation when paired with properly annotated datasets. Similarly, transformer-based caption models like BLIP can output contextual descriptions from visual input. This study attempts to combine these two powerful approaches—semantic segmentation and image captioning—into one shared pipeline for the analysis of satellite images.

The goal is not merely to appropriately classify land cover regions but also to generate informed captions that reflect land use pattern. Through extension of existing research and tailoring it for low-resource environments, this project contributes to the newly arising corpus of automatic geospatial analysis

2. METHODOLOGY

The suggested system for image segmentation from satellite images and generation of captions aims to label land cover areas and create informative text descriptions based on satellite images. The project works with a two-component architecture: DeepLabV3+ for semantic segmentation and BLIP for image captioning. The process is structured as dataset preparation, model choice, training steps, and captioning logic.

2.1 Dataset Preparation and Preprocessing-

DeepGlobe Land Cover Classification dataset is used for segmentation task. It consists of paired satellite images and manually annotated masks specifying land type as Urban Land, Agriculture Land, Forest, Water, Industrial, and Unknown. The images are preprocessed by resizing, normalization, and transformation into PyTorch tensors. For segmentation masks, the pixels' values are also converted to class-wise categorical labels.

2.2 Model Architecture and Training-

For segmentation, the DeepLabV3+ model with a ResNet-101 backbone is used because it is effective in processing multi-scale contextual information. CrossEntropyLoss is used for training the model along with the Adam optimizer. Training is conducted using Google Colab with support for checkpointing to handle disconnections. For captioning, BLIP (Bootstrapped Language Image Pretraining) is used to produce natural language captions. Land-distribution-based captions are also supported by computing the proportion of each land class in the segmented image.

3. MODELING AND ANALYSIS

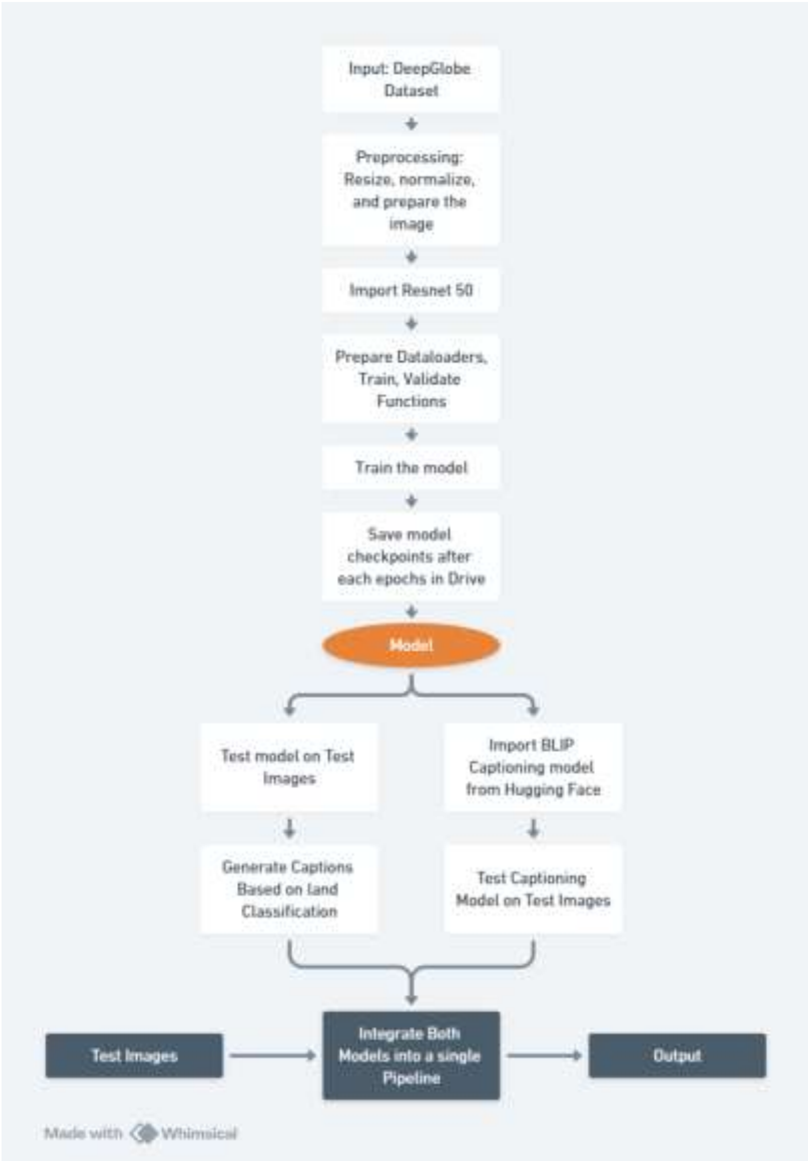

Figure 1: Pycnometer Test Procedure.

4. RESULTS AND DISCUSSION

The system was evaluated on a subset of the DeepGlobe dataset to evaluate its performance in semantic segmentation and captioning under autopilot mode. The DeepLabV3+ model worked well in classifying main land use classes, including agriculture, urban, forest, water, and industrial. The Intersection over Union (IoU) metric was used to evaluate segmentation accuracy per class, with good results achieved using clearly defined boundaries and minimal noise in most cases. Sample outputs reveal that segmented masks track the ground truth annotations closely, especially in regions with well-defined land boundaries. In the captioning component, BLIP generated natural language captions that accurately described visible details such as "a patch of cultivated land beside an urban settlement." At the same time, the land distribution-based captions provided percentage-wise distributions such as "This image contains 42.3% Agriculture, 31.7% Industrial, 20.5% Urban, and 5.5% Water," providing a quantitative overview of land cover.

Combining visual segmentation with text summaries enhanced satellite image interpretability and better suited the system for use in GIS, environmental monitoring, and urban planning.

Figure 1: Example Output of the Satellite Image classification with Captioning

By Uploading a sample image, the model was able to classify the land with pixel segmentation close to the ground truth. After that the segmentation mask is overlayed on original image for better understanding. After this captions are generated for better understanding. A BLIP caption is generated too so that a brief caption with all image features is generated.

5. CONCLUSION

This research suggests an end-to-end deep learning pipeline merging semantic segmentation and image captioning to further strengthen analysis of satellite imagery. DeepLabV3+-segmentation module accurately predicted diverse land covers with pixel-level prediction. However, the BLIP caption model produced descriptive captions for the visual data, and both contextual phrases and quantitative land distribution-based captions.

By means of structured preprocessing, model training, and testing, the system proved to have capability to provide accurate segmentation maps and perceptive textual observations. Such an ability of visual as well as textual understanding to a considerable extent helps in easing accessibility and comprehension of advanced geospatial information.

The suggested approach is promising for real-world applications across different domains such as geographic information systems, environmental monitoring, and urban planning. Moreover, the architecturally extensible approach is possible to extend in the future with improved models or more data to enable finer land classification and multilingual descriptions.

6. REFERENCES

[1] Land cover clustering and classification of satellite images [Available at: https://www.itm-conferences.org/articles/itmconf/pdf/2023/06/itmconf_icdsac2023_02004.pdf]

[2] Land Cover Classification from Time Series Satellite Images

[3] [Available at: https://ieeexplore.ieee.org/document/9456315]

[4] Land use classification of satellite images with convolutional neural networks (CNNs) [Available at: https://www.iacis.org/iis/2024/2_iis_2024_267-276.pdf]

[5] Land Cover Classification using Satellite Imagery and Deep Learning [Available at: https://developers.arcgis.com/python/latest/samples/land-cover-classification-using-unet/]

[6] Land Cover Classification From RGB and NIR Satellite Images Using Modified U-Net Model. [Available at: https://ieeexplore.ieee.org/document/10530973]

[7] Land use and land cover classification from satellite images based on ensemble machine learning and crowdsourcing data verification [Available at: https://www.tandfonline.com/doi/full/10.1080/23729333.2023.2166252]

[8] CNN architecture for land classification using satellite images [https://www.researchgate.net/publication/371379293_CNN_ARCHITECTURE_FOR_LAND_CLASSIFICATION_USING_SATELLITE_IMAGES]

[9] Land Classification Using Satellite Imagery Authors: Krish Pathak, Tanvi Thopte, Romir Maji, Jaishnu Kotian, Mrs. Shweta Yadav [Available at: https://www.ijraset.com/research-paper/land-classification-using-satellite-imagery]

[10] Phiri D, Simwanda M, Salekin S, Nyirenda VR, Murayama Y, Ranagalage M. Sentinel-2 data for land cover/use mapping: a review. Remote Sens. 2020;12(14):2291. [Available at: https://www.mdpi.com/2072-4292/12/14/2291]

[11] Zhang J, Chaoquan L, Li X, Kim H-J, Wang J. A full convolutional network based on DenseNet for remote sensing scene classification. Math Biosci Eng. 2019;16(5):3345–67. [Available at: https://www.aimspress.com/MBE/article/3611/fulltext.html]

[12] Helber P, Bischke B, Dengel A, Borth D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE J Sel Top Appl Earth Observ Remote Sens. 2019;12(7):2217–26. [Available at: https://ieeexplore.ieee.org/abstract/document/8736785]

[13] Ouchra, H., Belangour, A., & Erraissi, A. (2022). Machine learning for satellite image classification: A comprehensive review. 1–5. [https://doi.org/10.1109/ICDABI56818.2022.10041606]

[14] Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification: Priit Ulmas, Innar Liiv [https://arxiv.org/abs/2003.02899]