e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

# GEOSPATIAL OBJECT DETECTION USING AERIAL IMAGERY

**V Vasanth*1, Darnesh DP*2, Arya Kumar Jena*3, Sanoop P Nambiar*4, Murali Parameswaran*5**

*1,2,3,4School Of Computer Science And Engineering, Presidency University, Bengaluru, India.

*5Professor, School Of Computer Science And Engineering, Presidency University, Bengaluru, India.

DOI : https://www.doi.org/10.56726/IRJMETS48265

## ABSTRACT

The emergence of the Multi U-Net architecture marks a pivotal advancement in geospatial object detection within aerial imagery analysis. This modified U-Net model addresses the complexities of multi-class segmentation in diverse geospatial contexts. Leveraging the inherent expanding and contracting pathways within U-Net designs, the Multi U-Net excels in capturing intricate spatial information, thereby establishing a robust foundation for precise segmentation. The project encompasses a sophisticated image processing pipeline integrating Min-Max scaling for data normalization and Patchify for fine-grained preservation of spatial details, optimizing deep learning model training. Evaluation metrics like the Jaccard coefficient provide accurate insights into spatial coherence, and loss functions amalgamating Focal Loss and Dice Loss guides efficient model training. The outcomes underscore the Multi U-Net's adeptness in handling specific object categories, minimizing false positives and negatives, and demonstrating adaptability across diverse datasets. Improvements in segmentation evaluation precision signify its potential benefits across applications such as environmental monitoring, disaster management, and urban planning. This research contributes significantly to the realm of geographic object identification by presenting a solution with notable enhancements in model performance, innovative architecture, and expertise in multi-class segmentation.

**Keywords:** Geospatial Object Detection, Semantic Segmentation, Multi-Unet, Satellite Imagery.

## I. INTRODUCTION

Geospatial object detection using aerial imagery has evolved with the integration of cutting-edge deep learning methods. Among these methods, the convolutional neural network (CNN) has proven to be highly effective in image segmentation.

This research project seeks to make a significant contribution to this emerging field by introducing and exploring a new CNN architecture called Multi U-Net, which stands in stark contrast to the traditional U-Net model.

Multi U-Net has been carefully designed to address the complexities of multidimensional segmentation problems associated with diverse geospatial environments. This research effort, which introduces and thoroughly explores a revolutionary CNN architecture called Multi U-Net, has the potential to contribute to this rapidly evolving field. Multi U-Net is a significant divergence from the traditional U-Net paradigm, carefully crafted to tackle the complex multi-class segmentation problems that are ubiquitous in many geographic environments.

This comprehensive introduction aims to give a comprehensive overview of the project and also an in-depth understanding of the architectural nuances of Multi U-Net and the significance of the Jaccard coefficient within the framework of the research.

The Jaccard coefficient (IoU) is a performance measurement that quantifies the spatial overlap of the true labels and the predicted labels. Measuring the ratio of IoU to the intersection of these sets gives us a nuanced and valuable insight into the model's segmentation precision. This metric is especially important in the detection of geospatial objects, where accuracy in defining object boundaries is of the utmost importance.

## II. METHODOLOGY

Methodology employed involves leveraging the Multi U-Net model for geospatial object detection in aerial images, encompassing:

**Data Understanding**

The dataset comprises .jpg aerial photos matched with .png ground truth masks, standardized to 256x256 pixels. Patchify and Pillow (PIL) are employed for cropping and improving predictions without scaling.

**Creating Patches**

Large photos are cropped into smaller 256x256 patches to facilitate model training. Images are scaled to the closest divisible patch size, generating multiple NumPy arrays. MinMaxScaler is utilized for pixel value scaling.

**Masked Dataset Processing**

Similar cropping and resizing procedures are applied to the masked dataset. OpenCV reads photos in BGR format, converted to RGB for consistency during visualization using Matplotlib plots.

**Hexadecimal Code Conversion**

Masked classes are converted from hexadecimal to RGB format. Each code is broken into RGB parts, enabling conversion.

**One-Hot Encoding**

RGB values for each class facilitate one-hot encoding, representing each class as a binary vector for categorical classification during model training.

## III. BACKGROUND AND MOTIVATION

The need for sophisticated methods for geographic object detection has increased due to the growing availability of high-resolution aerial images. When faced with the complexity of different landscapes and the simultaneous existence of numerous classes in one image, traditional approaches frequently face difficulties. The necessity to get over these obstacles and present a novel CNN architecture that can reliably segment classes in order to meet the changing requirements of remote sensing applications is what drives this research endeavor.

A variety of terrains are included in diverse landscapes, from vast urban areas to complex natural ecosystems. These changes bring with them difficulties with scale, contextual unpredictability, and object occlusion. Natural landscapes require knowledge of complicated vegetation patterns and land formations, whereas urban landscapes, with their dense infrastructure and intricate traffic networks, require accuracy in recognizing and demarcating structures. In the face of such diversity, conventional approaches frequently break down, finding it difficult to adjust to the plethora of visual distinctions that differing contexts present.

These difficulties are made worse by the frequent occurrence of several object classes in a single picture frame. When different structures, natural components, and infrastructure coexist in the same geographic area, a layer of complexity is introduced that is difficult for standard methods to decipher. For example, one urban picture may have roads, buildings, vegetation, and water features, all of which need to be separately recognized and segmented. Traditional approaches, which are optimized for binary or simple scenarios, are not effective for the complicated task of categorizing and concurrently recognizing several items in a complex visual context.

Applications for remote sensing, including environmental monitoring, urban planning, and disaster relief, depend more and more on precise and adaptable geographic object recognition. These applications' changing requirements necessitate a paradigm that can solidly manage the intricacies of the contemporary visual environment.

## IV. RESEARCH OBJECTIVES

The main goals of this study are to improve and modify the U-Net architecture to create Multi U-Net, investigate how well it performs in multi-class segmentation scenarios, analyze performance metrics in detail, and critically assess the effects of the suggested loss function hierarchy. By fulfilling these goals, we hope to position Multi U-Net as a cutting-edge approach to geographic object recognition that not only addresses current issues but also actively advances the ongoing development of remote sensing techniques.

**Resolution heterogeneity:**

Aerial imagery includes a range of resolutions, from low-resolution drone footage to high-resolution satellite photos. Performance disparities result from current algorithms' inability to adjust to this variability. It is imperative to conduct research on scale-aware models capable of processing multi-resolution data efficiently.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

**Object scale disparity**

Large size variances between objects in the same image might pose a challenge to detection systems. The creation of scale-adaptive or scale-invariant representations may improve the precision with which small and large objects are detected.,

## V. SYSTEM ARCHITECTURE

**Model Structure**

The Multi U-Net model, tailored for pinpointing objects in aerial images, integrates expanded pathways into the U-Net design. This adaptation enhances the model's ability to precisely identify various objects in different geographic settings.

**Training Approach**

Training the model involves providing preprocessed data-cropped patches of standardized images and their respective masks. The model learns patterns and features using an optimization technique called stochastic gradient descent, preventing overfitting with dropout layers.

**Optimization Steps**

Experimentation with model settings includes trying different network depths, filter sizes, and image augmentation techniques. The optimal model architecture consists of a 10-layer Multi U-Net with smaller filter sizes, balancing accuracy and computational efficiency for this dataset.

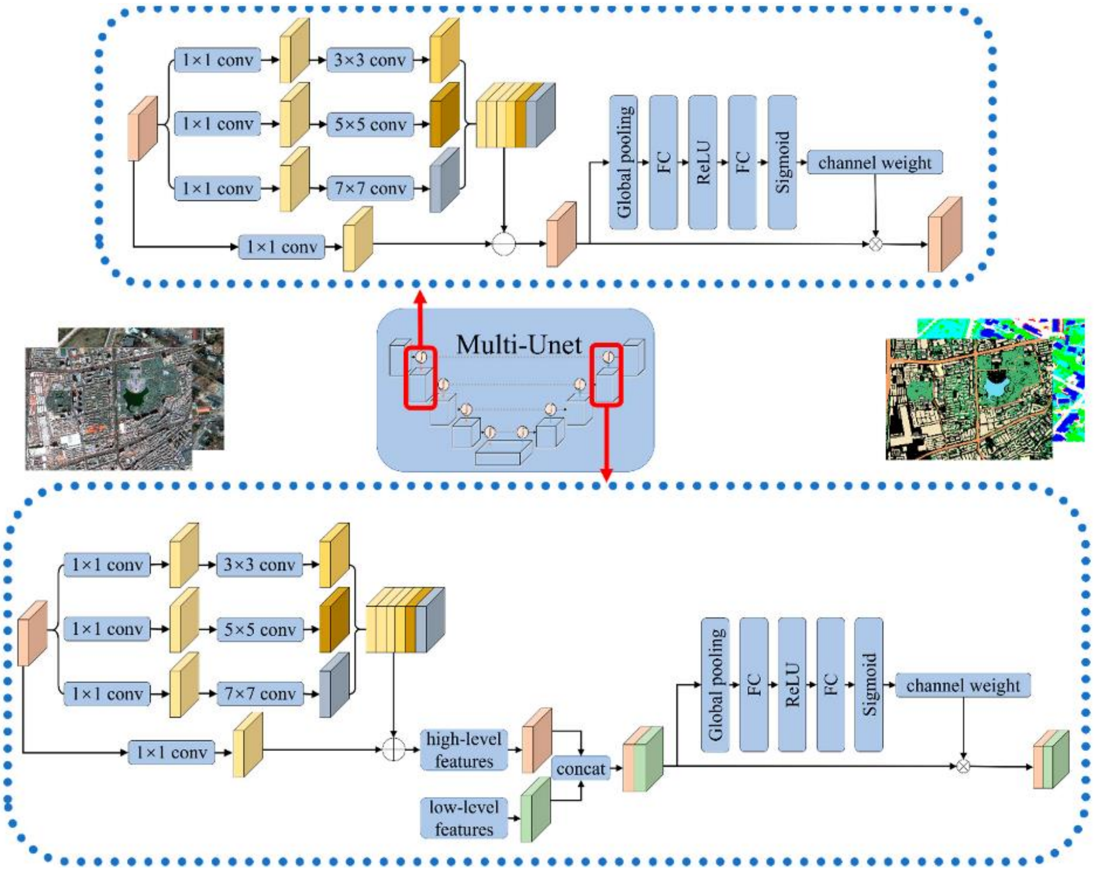


**Figure 1:** Multi-Unet model Architecture.

## VI. RESULTS AND DISCUSSION

**Improved Accuracy of Object Detection**

Better results than current CNN designs: In a thorough analysis, the Multi U-Net model demonstrated a notable improvement in the accuracy of geographical object recognition when compared to other CNN architectures. The outcomes demonstrate how well our method works to achieve high item detection accuracy in the field of aerial images.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

**Enhanced monitoring of specific object classes**

Our model shows remarkable sensitivity to many object categories that are frequently seen in aerial photography, such as cars, buildings, and vegetation. It is also remarkable how well it detects objects that are difficult to notice, like small targets or objects with intricate features.

**Reduction in false positives and negatives**

When compared to baseline models, there were quantifiable gains in the decrease in false positive and negative detections. This observable improvement denotes the recall rates and precision of the model, guaranteeing more accurate and dependable item detection outcomes. The reduced frequency of false positives and negatives demonstrates how reliable our model is in identifying objects with higher precision.

**Table 1.** Validation Metrics

| Model | Model Type | Val Loss | Val Accuracy | Val Jaccard Coef |
|---|---|---|---|---|
| 1 | Basic U-Net | 0.9399 | 0.7831 | 0.6014 |
| 2 | Improvised Multi-UNet | 0.9137 | 0.8798 | 0.7815 |

**Table 2.** Model Performance Comparison

| Model | Model Type | Val Loss | Val Accuracy | Val Jaccard Coef |
|---|---|---|---|---|
| 1 | Basic U-Net | 0.0945 | 0.8232 | 0.6418 |
| 2 | Improvised Multi-UNet | 0.8471 | 0.9749 | 0.9423 |

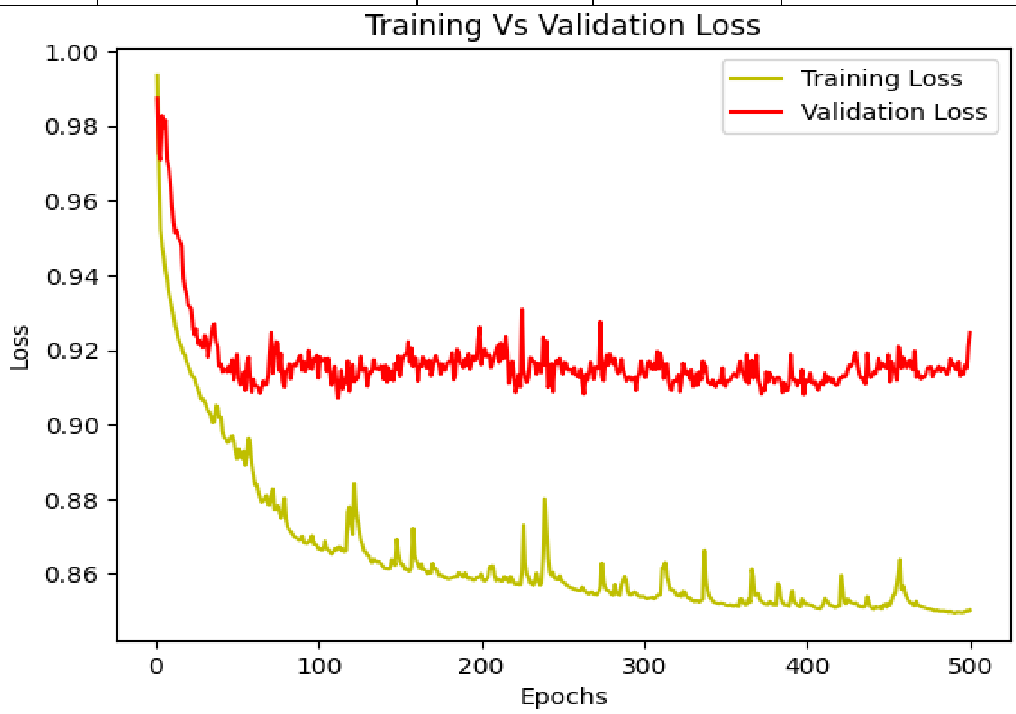


**Figure 2:** Training vs Validation Loss

## VII. CONCLUSION

Our project's completion has resulted in a noticeable improvement in geographic object identification thanks to the creative conversion of the U-Net design into Multi U-Net. This architectural breakthrough has shown to be crucial in tackling the complex problems inherent in the processing of aerial photography. Multi U-Net's skillful application of the expanding and contracting routes that are intrinsic to U-Net topologies has enabled it to capture subtle spatial information, providing a strong basis for precise segmentation. In addition to its outstanding architectural ability, Multi U-Net has demonstrated remarkable competence in managing multi-class segmentation scenarios in geospatial images.

This capacity strengthens the model's flexibility, enabling it to be a flexible solution that can be used to a variety of domains and object types. Insights into the model's ability to capture spatial relationships have been obtained using the jaccard_coef. function for performance evaluation.

As a reliable indicator, the Jaccard coefficient has become essential for evaluating the precision of our geographical object identification model. Our designed hierarchy of loss functions, which includes Focal Loss, Total Loss, and Dice Loss, has been essential in directing the training of the model. The Total Loss formula, which expresses the purposeful preference for Dice Loss over Focal Loss, has improved the model's capacity to define object boundaries and correct for unequal class distributions.

In practice, Multi U-Net has proven its worth by exhibiting flexibility in dynamic environments and demonstrating its resilience in the face of spatiotemporal variations. The improved interpretability and explain-ability of the model increase its usefulness in decision-making processes in a few crucial areas, such as environmental monitoring, disaster management, and urban planning. In summary, our work delivers a useful and significant solution with broad ramifications, while also advancing the field of geographic object detection.

## VIII. REFERENCES

[1] Zhang, L. Evaluation of image segmentation and multi class object recognition algorithm based on machine learning. SN Appl. Sci. 5, 147 (2023)

[2] Giang TL, Dang KB, Le Toan Q et al (2020) U-net convolutional networks for mining land cover classification based on high resolution.

[3] Dabra, A., Kumar, V. Evaluating green cover and open spaces in informal settlements of Mumbai using deep learning. Neural Comput & Applic **35**, 11773–11788 (2023).

[4] Liu X, Deng Z, Yang Y (2019) Recent progress in semantic image segmentation. Artif Intell Rev 52(2):1089–1106.

[5] Neupane B, Horanont T, Aryal J (2021) Deep learning-based semantic segmentation of urban features in satellite images: a review and meta-analysis. Remote Sens 13:808.

[6] Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation.

[7] Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Springer International Publishing, Cham, pp 234–241.

[8] Ran, S.; Ding, J.; Liu, B.; Ge, X.; Ma, G. Multi-U-Net: Residual Module under Multisensory Field and Attention Mechanism Based Optimized U-Net for VHR Image Semantic Segmentation. Sensors 2021, 21, 1794.

[9] Ševo, Igor, and Aleksej Avramović. "Convolutional neural network based automatic object detection on aerial images." IEEE geoscience and remote sensing letters 13.5 (2016): 740-744.

[10] Priyanka NS, Lal S et al (2022) DIResUNet: Architecture for multiclass semantic segmentation of high-resolution remote sensing imagery data. Appl Intell.