STM JOURNALS

Journal of
Image Processing & Pattern Recognition Progress

ISSN: 2394-1995
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/JoIPPRP

Research JoIPPRP

Change Detection in Aerial Imagery

Mandira M.[1], D. Karan Sai Reddy[1], Kesanapalli Lakshmi Priyanka[1,*],
K. Vamshi Krishna[2]

Abstract

The development of the Multi U-Net engineering marks an essential headway in geospatial question location inside ethereal symbolism investigation. The altered U-Net addresses the complexities of multi-class division in assorted geospatial settings. Leveraging the inalienable growing and contracting pathways inside U-Net plans, the Multi U-Net exceeds expectations in capturing complicated spatial data, in this manner setting up a vigorous establishment for exact division. The extend envelops an advanced picture handling pipeline joining Min-Max scaling for information normalization and Patchify for fine-grained conservation of spatial subtle elements, optimizing profound learning show preparation. Assessment measurements just like the Jaccard coefficient give exact bits of knowledge into spatial coherence, and misfortune capacities amalgamating Central Misfortune and Dice Misfortune guide effective model training. The results emphasize Multi U-Net's adeptness in taking care of particular question categories, minimizing untrue positives and negatives, and illustrating versatility over different datasets. Changes in division assessment exactness imply its potential benefits over applications such as natural observing, calamity administration, and urban arranging. This investigate contributes essentially to the domain of geographic question recognizable proof by showing an arrangement with striking upgrades to demonstrate execution, inventive engineering, and mastery in multi-class division.

Keywords: Geospatial object detection, semantic segmentation, Multi-UNet, satellite imagery

INTRODUCTION

Geospatial location utilizing ethereal symbolism has advanced with the integration of cutting-edge profound learning strategies. Among these strategies, the convolutional neural organize (CNN) has demonstrated to be profoundly compelling in picture division. This investigation venture looks to form a critical commitment to this rising field by presenting and investigating an unused CNN engineering called Multi U-Net, which stands in stark contrast to the conventional U-Net demonstration. Multi U-Net has been carefully outlined to address the complexities of multidimensional division issues related with differing geospatial situations. This investigation effort, which presents and altogether investigates a progressive CNN design called Multi U-Net, has the potential to contribute to this rapidly evolving field. Multi U-Net could be a critical disparity from the conventional U-Net worldview, carefully created to handle the complex multi-class division issues that are omnipresent in numerous geographic situations. This comprehensive presentation points to allow a comprehensive diagram of the extent and an in-depth understanding of the engineering subtleties of Multi U-Net and the

***Author for Correspondence**
Kesanapalli Lakshmi Priyanka
E-mail: klpriyanka.cs20@rvce.edu.in

[1]Student, Department of Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering (RVCE), Bengaluru, Karnataka, India
[2]Associate Professor, Department of Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering (RVCE), Bengaluru, Karnataka, India

Received Date: June 05, 2024
Accepted Date: July 03, 2024
Published Date: July 09, 2024

Citation: Mandira M., D. Karan Sai Reddy, Kesanapalli Lakshmi Priyanka, K. Vamshi Krishna. Change Detection in Aerial Imagery. Journal of Image Processing & Pattern Recognition Progress. 2024; 11(2): 22–28p.

importance of the Jaccard coefficient inside the system of the investigation. The Jaccard coefficient (IoU) could be an execution estimation that evaluates the spatial cover of the genuine names and the anticipated names. Measuring the proportion of IoU to the crossing point of these sets gives us a nuanced and profitable knowledge into the model's segmentation precision. This metric is particularly critical within the discovery of geospatial objects, where precision in characterizing question boundaries is of the most extreme significance.

RELATED WORK

Change detection in aerial imagery has gained popularity in recent years due to the increasing availability of high-resolution satellite images for various applications in organizations and businesses. However, the cost and complexity of implementation have hindered wider adoption of change detection systems, leading some businesses to rely on manual methods or existing systems for monitoring changes in landscapes. This section provides a glimpse of various significant studies conducted by different researchers and their approaches to designing more effective change detection systems. There has been remarkable work and extensive literature available in the field of remote sensing, particularly focusing on change detection in aerial imagery.

Dabra and Kumar have discussed a way to develop an effective approach for identifying salient regions where changes have occurred between two time periods [1]. The proposed algorithm, referred to as BIC2 (Bi-temporal Hyperspectral Image Iterative Clustering and Classification), cascades clustering and classification through iterative learning. The method first spatially segments the bi-temporal hyperspectral images into super pixels using simple linear iterative clustering (SLIC). It then iteratively performs clustering and classification on the super pixels to separate the salient regions where changes have occurred from the unchanged regions.

Giang *et al.* used the VGG16-UNet which empowers high-resolution satellite image analysis by performing semantic segmentation, allowing for detailed understanding of cityscapes [2]. By training three customized Convolution Neural Network (CNN) models: VGG16-UNet, MobileNetV2-UNet, and DeepLabV3, on manually labeled high-resolution satellite images for the city of Mumbai, this paper seeks to close this gap. For the classes under consideration, all three models demonstrated exceptional performance, attaining the cutting edge of overall accuracy, which hovers around 95%. Since VGG16-UNet outperformed the other two models, it was used to create green, open, and green and open indexes for the entire city.

Kilic and Ozturk used UNet Convolutional Network, which is utilized for land cover classification based on multispectral UAV images in a mining area and were trained using different optimizer functions [3]. They trained three modified Convolutional Neural Network (CNN) models: VGG16-UNet, MobileNetV2-UNet, and DeepLabV3, on high-resolution satellite imagery of Mumbai. All three models achieved high accuracy, with VGG16-UNet performing the best. This model was then applied citywide to develop indices measuring green and open spaces.

Li *et al.* primarily used the support vector machine algorithm in machine learning to update the multi-target recognition and image segmentation algorithm [4]. Simulation experiments are used to confirm the validity of the newly proposed algorithm, and the authors concluded that the multi-target recognition and image segmentation algorithm based on machine learning can process images more effectively. Lastly, experiments were conducted to assess the performance of the image segmentation and multiclass object recognition algorithms suggested in this study to that of traditional algorithms.

Dalal *et al.* provided a straightforward and efficient single-shot detector model for counting and detecting automobiles in aerial photos [5]. Target automobile instances' heatmaps are predicted using the heatmap learner convolutional neural network (HLCNN), a suggested model. We have upgraded CNN architecture by adding three convolutional layers as adaptation layers instead of fully connected layers in order to learn the target cars' heatmap. The suggested model makes use of the VGG-16 as its backbone convolutional neural network. The suggested technique accurately locates the center of the target cars and counts the number of cars.

Ardeshir *et al.* employed a Density-Map guided object identification network (DMNet) model in their study [6]. It is intended to solve issues with object detection in high-resolution aerial photos. The wide range of object sizes and the uneven distribution of things in the pictures present two major challenges. To illustrate how items are dispersed across the image in terms of pixel intensity, DMNet creates a density map. This map facilitates object detection by breaking up the big aerial image into smaller, easier-to-manage crops.

Yu and Ji employed the Single Shot Multibox Detector (SSD) model used in this study [7]. This is made up of the following: Module for Deconvolution, Module for Super-Resolution, and Module for Feature Fusion in Shallow Layer. Satellite photos can be super-resolved using the algorithm covered in the sources. This algorithm is used to improve low-resolution satellite picture resolution, enabling a crisper and more detailed depiction of the imagery.

Vasanth *et al.* focused on handling the intricacies of multidimensional segmentation issues related to various geographic situations, Multi U-Net was meticulously created [8]. This research endeavor has the potential to make a significant contribution to this quickly developing subject as it presents and thoroughly investigates the breakthrough CNN architecture known as Multi U-Net. The primary objectives of this work are to refine and adapt the U-Net architecture to generate Multi U-Net, examine its performance in multi-class segmentation situations, conduct a thorough analysis of performance metrics, and evaluate the implications of the proposed loss function hierarchy.

Zhang *et al.* focused on the online and offline tracking methods which are the two categories under which the traffic object tracking techniques are examined [9]. Moreover, there are four main categories into which online tracking techniques fall: CF-based, TBD, DL-based, and optical flow-based techniques. Lastly, talks on the examined traffic tracking techniques are presented. This research examines five primary tracking objects' object tracking using satellite video analysis. High-level classification of the tracking objects and benchmark dataset into artificial (ships and traffic objects) and natural (typhoons, fire, and ice motion) targets is done. The motion velocity and target size are the primary distinctions between artificial and natural targets. As a result, multiple tracking methods and datasets with varying spatial and temporal resolutions are produced. To be more precise, tracking automobiles requires high FPS and spatial resolution films. Additionally, tracking ships and automobiles requires the successful integration of multimodality data, like AIS and SAR, with optical images. As a result, the appropriate and accessible datasets for natural targets differ depending on the object's size.

PROBLEM IDENTIFICATION

In order to ensure sustainable urban planning, it is essential to monitor urban development by means of aerial and spatial imagery, which enables the detection of changes in the landscape over time. Ensuring a high degree of clarity, the management of large datasets and integrating diverse sources of data for precise analysis are among the challenges faced by this process [10].

PROPOSED SYSTEM

Model Structure System Architecture

The Multi U-Net model incorporates extended routes into the U-Net architecture and is specifically designed for object detection in aerial photos. This modification improves the model's capacity to accurately detect different items in diverse geographical contexts.

Training Methodology

To train the model, preprocessed data-cropped regions of reference images together with the corresponding masks are supplied. Stochastic gradient descent is an optimization strategy that the model uses to learn patterns and features, avoiding overfitting using dropout layers.

Optimization Steps

Experimentation with various network depths, filter widths, and image augmentation approaches are part of the model settings' experimentation process. For this dataset, a 10-layer Multi U-Net with reduced filter sizes is the best model architecture since it balances computational efficiency and accuracy as shown in Figure 1.

RESEARCH METHODOLOGY

Data Understanding

The dataset consists of up of .jpg files representing aerial photos and matching .png files serving as ground truth masks. Each file has roughly 72 images in total, and a .json file provides information on several classes. All images are reduced in size to 256×256 pixels so that the data is consistent. Cropping is preferred over scaling in order to increase the number of data patches for the model and enhance predictions. Patchify and Pillow (PIL) are used for this.

Create patches

Large photos are cropped into locations to facilitate model training. Reading an image, scaling it to the nearest shareable web size (256), and then cropping it to 256×256 web is resizing. This process splits one huge image into multiple images, each of which is then converted to a NumPy array. The pixel values are then scaled using Min Max Scaler.

Masked Dataset Processing

Similar photos, cropping and resizing are also used in masked dataset. Although RGB format is preferred, OpenCV can read photos in BGR format. When reading the image, a new line of code is added to convert the BGR to RGB. Alignment of photos and corresponding masks is checked during visualization using Matplotlib graphs.

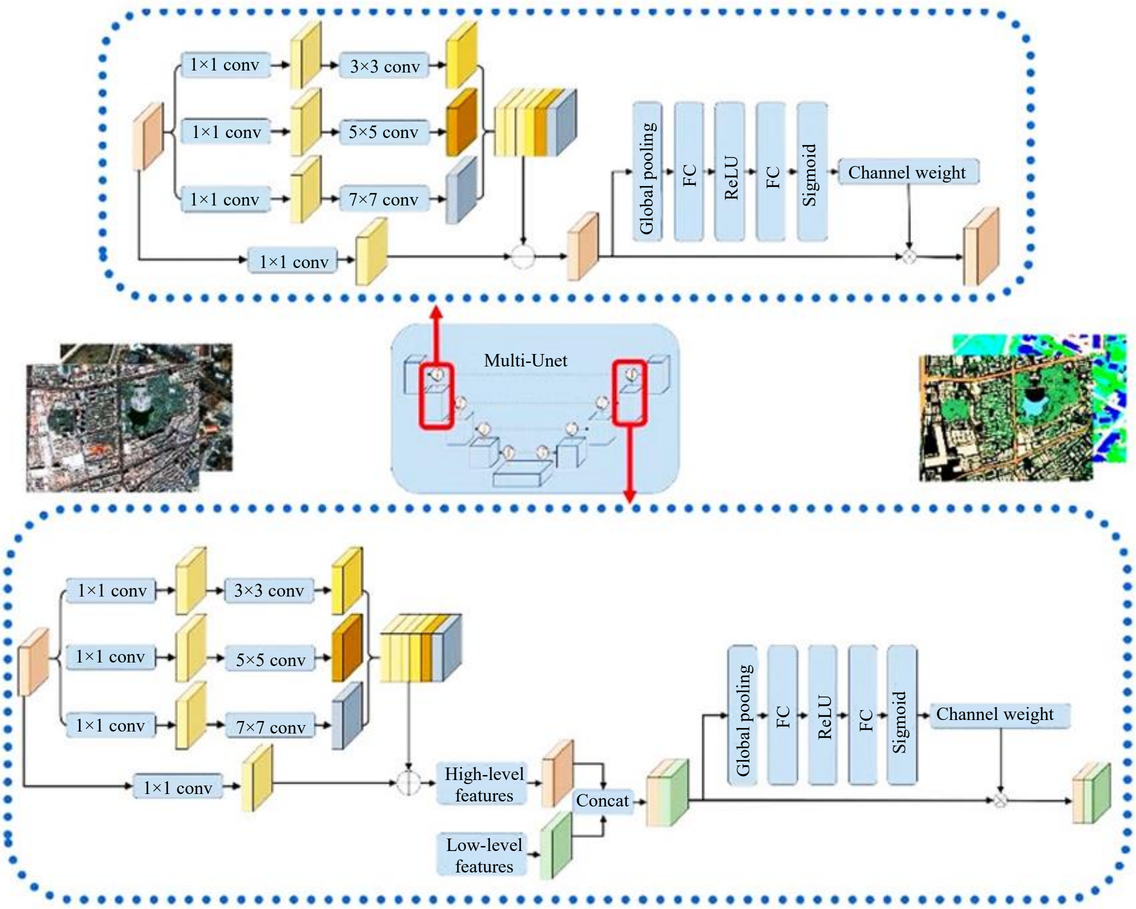

Figure 1. Architecture of Multi U-Net.

Hexadecimal Code Conversion

Converting the masked classes from hexadecimal to RGB format is the first step in the process. All hexadecimal codes are decomposed into RGB components, and the conversion process is explained. The RGB value for the "buildings" class, for instance, with hex code 3C1098 is (R=60, G=10, B=152). This method is carried out for each class in the dataset.

One-Hot Encoding

After obtaining the RGB values for each class, one-hot encoding is used in the model for classification. To aid in categorical categorization during model training, a binary vector is utilized to represent every class.

IMPLEMENTATION

Data Acquisition and Preprocessing

- *Dataset Selection:* Choose a diverse and representative dataset of aerial images.
- *Data Annotation:* Annotate the dataset with bounding boxes around objects of interest.
- *Data Splitting:* Divide the dataset into training, validation, and test sets to ensure robust model evaluation.

Image Preprocessing

- *Normalization:* Standardize pixel values to a common scale to ensure consistency across images.
- *Augmentation:* Apply data augmentation techniques (rotation, flipping, zooming) to increase the diversity of the training set and improve model generalization.

Model Selection

- *Base Architecture:* Considered using Multi U-net model.
- *Transfer Learning:* Utilization of pre-trained models on large datasets as a starting point.

Model Adaptation

- *Architecture Modification:* Adjust the selected model architecture to match the input characteristics and challenges of aerial imagery.
- *Hyperparameter Tuning:* Fine-tune hyperparameters such as learning rate, batch size, and anchor box sizes to optimize model performance.

Training

- *Loss Function:* Choose an appropriate loss function for object detection, such as the combination of classification and localization losses.

Evaluation

- *Metrics:* Employ evaluation metrics like precision, recall, F1 score, and Intersection over Union (IoU) to assess the model's performance.
- *Test Set Evaluation:* Evaluate the trained model on the test set to ensure unbiased performance assessment.
- *Post-Processing:* Non-Maximum Suppression: Implement post-processing techniques, non-maximum suppression, to filter redundant bounding boxes and improve localization accuracy.

RESULTS AND DISCUSSION

Better Results than Current CNN Designs

Detailed research showed that, in comparison to existing CNN designs, the Multi U-Net model significantly improved the accuracy of geographical object detection. The results show how well our approach performs in the field of aerial photos to obtain high item detection accuracy as shown in Table 1.

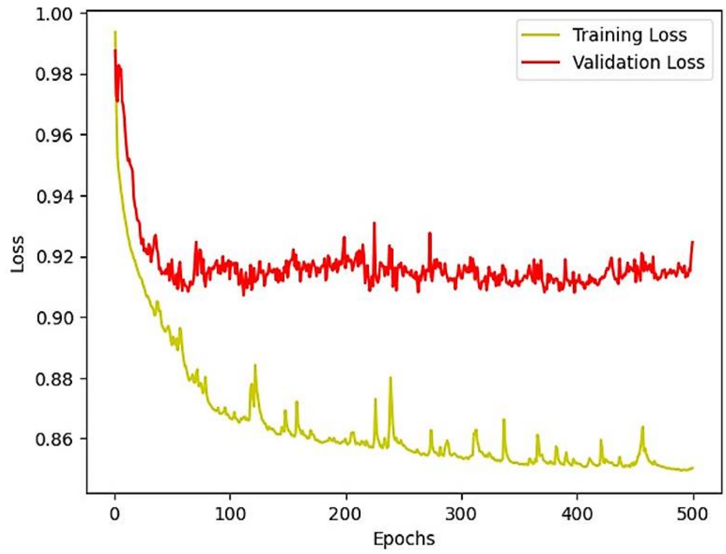

Figure 2. Training vs. Validation loss.

Table 1. Results.

Model	Model Type	Val Loss	Val Accuracy
1	Basic U-Net	0.0945	0.8232
2	Improvised Multi U-Net	0.8471	0.9423

Enhanced Monitoring of Specific Object Classes

When it comes to numerous object categories that are commonly visible in aerial imagery, like cars, buildings, and plants, our model has exceptional sensitivity. Its ability to identify subtle or intricately featured objects, as well as small targets, is particularly impressive.

Reduction in False Positives and Negatives

There were measurable improvements in the reduction of false positive and negative detections as compared to baseline models. This observable improvement indicates the model's recall rates and precision, ensuring more reliable and accurate item detection results. Our model's increased precision in detecting objects is demonstrated by the decreased occurrence of false positives and negatives as shown in Figure 2.

CONCLUSION

With the end of our project, the imaginative transformation of the U-Net architecture into Multi U-Net has led to a discernible improvement in geographic object recognition. It has become evident that this architectural innovation is essential to solving the intricate issues associated with processing aerial photos. Through the deft use of the expanding and contracting routes inherent in U-Net topologies, Multi U-Net has been able to collect sensitive spatial information, offering a solid foundation for accurate segmentation. Apart from its exceptional architectural skills, Multi U-Net has proven to be exceptionally adept at handling multi-class segmentation scenarios in geographic pictures.

The model's versatility is strengthened by this capacity, making it a versatile solution applicable to a wide range of domains and object kinds. By evaluating the model's performance, the jaccard_coef function has provided insights into how well it captures spatial relationships.

The Jaccard coefficient has become a crucial metric for assessing the accuracy of our geographical object recognition model because it is a dependable indicator. The training of the model has been guided by our constructed hierarchy of loss functions, which consists of Focal Loss, Total Loss, and Dice Loss.

The Total Loss formula has enhanced the model's ability to determine object boundaries and adjust for mismatched class distributions by expressing the deliberate preference for Dice Loss over Focal Loss.

Multi U-Net has demonstrated its worth in practice by displaying robustness against spatiotemporal variations and flexibility in dynamic contexts. The model's enhanced interpretability and explainability make it more beneficial for application in decision-making processes in a few critical domains, including urban planning, disaster management, and environmental monitoring. In conclusion, our study advances the field of geographic object detection and provides a major and practical solution.

REFERENCES

1. Dabra A, Kumar V. Evaluating green cover and open spaces in informal settlements of Mumbai using deep learning. Neural Comput Appl. 2023 Jun; 35(16): 11773-88.
2. Giang TL, Dang KB, Le QT, Nguyen VG, Tong SS, Pham VM. U-Net convolutional networks for mining land cover classification based on high-resolution UAV imagery. IEEE Access. 2020 Oct 19; 8: 186257–73.
3. Kilic E, Ozturk S. An accurate car counting in aerial images based on convolutional neural networks. J Ambient Intell Human Comput. 2023 Feb 1; 14(8): 1259–1268.
4. Li C, Yang T, Zhu S, Chen C, Guan S. Density map guided object detection in aerial images. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020; 190–191.
5. Dalal AA, Shao Y, Alalimi A, Abdu A. Mask R-CNN for geospatial object detection. International Journal of Information Technology and Computer Science (IJITCS). 2020; 12(5): 63–72.
6. Ardeshir S, Zamir AR, Torroella A, Shah M. GIS-assisted object detection and geospatial localization. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13. Cham: Springer International Publishing. 2014; 602–617.
7. Yu D, Ji S. A new spatial-oriented object detection framework for remote sensing images. IEEE Trans Geosci Remote Sens. 2021 Nov 10; 60: 1–16.
8. Vasanth V, Darnesh DP, Arya Kumar Jena. Geospatial Object Detection Using Aerial Imagery. Int Res J Mod Eng Technol Sci. 2024; 6(1): 860–864.
9. Zhang Z, Wang C, Song J, Xu Y. Object tracking based on satellite videos: A literature review. Remote Sens. 2022 Jul 31; 14(15): 3674.
10. Shen Y, Liu D, Chen J, Wang Z, Wang Z, Zhang Q. On-board multi-class geospatial object detection based on convolutional neural network for High Resolution Remote Sensing Images. Remote Sens. 2023 Aug 10; 15(16): 3963.

