SEMANTIC SEGMENTATION FOR AERIAL IMAGES

Dr. B. Harika[1], *M. Bharath*[2] *and K. Himneesh*[3]
[1]*Department of IT, MGIT(A), Gandipet, Hyderabad, 500075, Telangana, India.*
[2]*Department of IT, MGIT(A), Gandipet, Hyderabad, 500075, Telangana, India.*
[3]*Department of IT, MGIT(A), Gandipet, Hyderabad, 500075, Telangana, India.*

Abstract: *Semantic segmentation of aerial imagery plays a critical role in modern urban planning, environmental monitoring, and the development of smart cities. This project presents an interactive web-based application that performs semantic segmentation on high-resolution aerial images using a deep learning-based U-Net model. The system is developed using Python and integrated into a Streamlit framework to provide a seamless user experience through a browser-based interface.*

The application allows users to upload aerial or satellite images and visualizes pixel-wise segmentation results across six predefined classes: Buildings, Roads, Land, Vegetation, Water, and Unlabeled. It goes beyond basic segmentation by offering advanced interactive features such as zooming, class mask toggling, and real-time class-wise statistical analysis, including area coverage.

The model is trained and evaluated using the "Semantic Segmentation of Aerial Imagery – Dubai, UAE" dataset, which contains pixel-annotated satellite imagery. The proposed system addresses limitations in existing solutions, such as lack of interactivity, low accuracy, and absence of class-wise analytics and toggling class maks. By streamlining the segmentation workflow and offering rich visualization and analytical tools, the system enhances accessibility for non-technical users and supports data-driven decision-making in geospatial analysis.

Keywords: Aerial Imagery, Deep Learning, Geospatial Intelligence, Image Processing, Interactive Visualization, Land Cover Analysis, Pixel-wise Classification, Remote Sensing, Semantic Segmentation, U-Net, Urban Planning, Web Application

1. Introduction

Geospatial analysis has transformed significantly in the recent years, primarily due to the satellite imagery and remote sensing technology revolution. Semantic segmentation of aerial imagery is among the most significant tools that have emerged to enable us to comprehend and even manage city spaces better. Semantic segmentation is the act of labeling each pixel in an image to a fixed category, such as buildings, roads, vegetation, and water, to better comprehend complex city configurations. However, conventional segmentation methods are unable to handle the high detail and complexity of new satellite imagery. The methods are unable to support accurate, real-time analysis and do not contain the interactive features that urban planners, ecologists, and decision-makers require.

In order to solve these problems, the current project builds a web application using deep learning methods that process aerial images by classifying them into different classes and showing dynamic visualizations. The system uses a pre-trained U-Net model to effectively segment high-resolution satellite images and display the output on an easy-to-use web interface. Based on a well-designed Streamlit platform, the system offers interactive features like zooming and unzooming, panning, class switching among different classes, and real-time updating of object numbers and area coverage. This solution bridges the gap between deep learning methods and real-world deployment, providing an easy-tounderstand and easy-to-use analysis tool for technical and non-technical users of cities. The system is especially useful for smart city planning, disaster relief efforts, and green city planning, where rapid and precise spatial assessment is important.

2. System Architecture

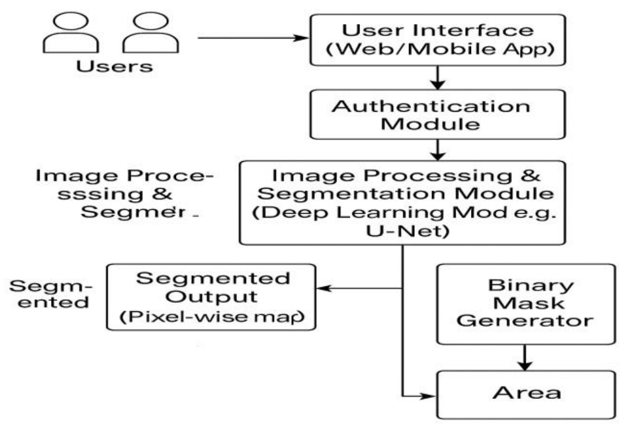

Figure 1: System Architecture:

The system architecture for the automated detection and classification of Diabetic Retinopathy using retinal fundus images. The process begins with the input of raw retinal images, which undergo pre-processing steps including color normalization, edge enhancement, and color space conversion to improve the quality and consistency of the images for further analysis. Following this, the system extracts key retinal structures such as the optic disc, blood vessels, macula, and fovea to identify relevant anatomical features necessary for disease detection. The next phase involves the detection of pathological features associated with Diabetic Retinopathy, including exudates, microaneurysms, maculopathy, and hemorrhages. These extracted features are then passed through a Convolutional Neural Network (CNN) model that classifies the severity of Diabetic Retinopathy into five levels, ranging from no apparent signs to proliferative stages. Finally, the system outputs the results, displaying the diagnostic classification to aid in clinical decision-making. This architecture is designed to enhance the accuracy, speed, and accessibility of Diabetic Retinopathy screening and diagnosis.

Dataset Preparation

Before developing and training the semantic segmentation models, it was crucial to meticulously prepare the dataset to ensure accurate, high-quality input for effective learning. This project primarily used publicly available aerial and satellite imagery datasets sourced from platforms such as Kaggle, Google Earth Engine, and open-access remote sensing repositories. The images contained diverse landscape elements such as buildings, roads, vegetation, water bodies, and open land, which were essential for object-level segmentation and classification.

Once gathered, the dataset underwent several preprocessing steps to standardize and enhance the images for deep learning-based segmentation:

- **Standardization**: The images varied in resolution and format. To maintain uniformity, all images were resized to 256×256 pixels and converted to RGB format. This standard size was chosen to balance computational efficiency and spatial detail.

● **Normalization**: Pixel values were scaled to the range [0, 1] to accelerate model convergence and ensure numerical stability during training.

● **Noise Removal**: Preprocessing techniques such as Gaussian filtering and histogram equalization were applied to reduce noise and enhance image clarity, especially in regions with poor lighting or cloud cover.

● **Annotation and Mask Generation**: For each aerial image, corresponding ground-truth segmentation masks were prepared. These masks labeled each pixel according to object classes (e.g., building, road, vegetation), either using manually labeled datasets or tools like Labelbox and VGG Image Annotator (VIA).

● **Class Balancing**: Datasets were reviewed for class imbalance. Oversampling, undersampling, or data augmentation techniques (rotation, flipping, cropping) were applied to ensure balanced representation across all classes.

The final dataset was split into training, validation, and testing subsets. This structure ensured that the model could learn segmentation patterns from one part of the dataset, fine-tune hyperparameters on another, and be evaluated fairly on unseen data. This division played a key role in building a robust and generalizable segmentation model capable of accurately identifying objects in various aerial image conditions.

3. Model Training

Once the dataset was preprocessed and properly structured, the next critical phase involved training deep learning models to accurately perform semantic segmentation on aerial images. This stage focused on teaching the model to label each pixel in an image with its corresponding object class—such as buildings, roads, vegetation, or water bodies—based on spatial and contextual patterns learned during training.

- The U-Net model was trained on the preprocessed dataset with a categorical cross-entropy loss functionand an Adam optimizer.
- Training is done in multiple rounds with the correct batch sizes. Validation of performance is done with metrics such as accuracy, IoU (Intersection over Union), and dice coefficient for each class.
- Validation data is utilized to evaluate generalization performance and to optimize hyperparameters.
- Model checkpoints and early stopping mechanisms are employed to avoid overfitting and ensure convergence.

Web Application Development:

- The U-Net model that is trained is embedded into a Streamlit web app with real-time user input.
- The users can upload aerial images through the web-site.
- The backend processes the images to generate segmentation masks.
- It is developed with JavaScript, CSS, and HTML.

- It supports interactive visualization such as panning, turning class masks on and off, zooming, and inspecting pixels in detail.
- Other functionalities are displaying the percentage coverage of every class, retrieving objects by area, and determining their size in pixels.

Table 1. Model Comparison

Model	Accuracy	Speed	Ideal Use Case
UNet (Custom)	High	Medium	Segmentation of small-to-medium sized aerial datasets.
UNet with VGG16 Backbone	Higher	Medium	Detailed building and land cover segmentation.
UNet with ResNet50 Backbone	Very High	Slower	Complex object segmentation with high spatial consistency.
DeepLabV3+ with ResNet101	Very High	Slower	Large-scale, high-resolution semantic segmentation tasks.

4. Deployment and Results

After training and validating the segmentation models, the best-performing model was deployed using *Streamlit*, a lightweight and user-friendly Python framework for building interactive web applications. This allowed real-time inference directly within a browser, enabling users to upload aerial images and visualize the segmented outputs instantly without any need for backend or frontend development.

The application presents segmentation results with *color-coded overlays* that distinguish between various land cover types such as *buildings, roads, vegetation, and water bodies, making the results easily interpretable for stakeholders like **urban planners, environmental analysts, and survey agencies*.

Deployment Details:

Platform: Streamlit Web App

User Interface: Built entirely with Streamlit widgets (e.g., file uploader, sliders, buttons)

Model Format: .h5 or ONNX for efficient loading

Performance: Inference time \~2–4 seconds per high-resolution image on a mid-range GPU

Features:

- Real-time image upload and segmentation
- Interactive display of output with overlaid segmentation masks
- Downloadable results for further analysis

This streamlined deployment ensured accessibility and ease of use, with no need for separate backend or frontend infrastructure.

System Development

• **App Features and Functionality:** The aerial photo segmentation web application that would offer a a user-friendly and intuitive interface with these main attributes:

- **Image Upload and Preview:** Image Upload and Preview: Clients are able to upload high-high-resolution aerial imagery that is subsequently displayed to Segmentation.
- **Semantic Segmentation Output:** The backend model classifies the image into categories such as Buildings, Roads, Vegetation, Water, Land, and Unlabeled regions. Segmented masks are apparent overlayed on the original picture.
- **Class Mask Toggling:** Switch the visibility of specific topics (e.g., Structures, Highways) for enhanced interpret some objects within the image.
- **Pixel Area Calculator:** Users can click on segmented objects to identify their class and calculate their pixel area.
- **Class Coverage Percentage:** The application calculates and displays the percentage area each class occupies in the uploaded image.

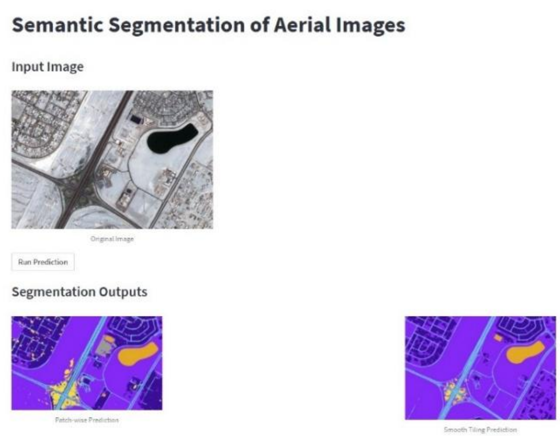

FIGURE 5.1 Segmentation Output: An example of a segmented image showing different land cover

Figure 1: Illustrates the semantic segmentation process on an aerial image. The top portion is the original image, which indicates a city area with a body of water and surrounding buildings. When you press the ”Run Prediction” button, it generates two kinds of outputs: Patch-wise Prediction and Smooth Tiling Prediction. These outputs indicate how various land types—such as water, roads, and buildings—are

detected using color-coded masks. The Patch-wise Prediction indicates the model's output on individual image patches, which can cause boundary issues, whereas the Smooth Tiling these issues by blending overlapping tiles for improved flow and continuity in the final segmented output.

FIGURE 5.2 Individual Class Masks

Figure 2: The above image represents the binary masks of various land cover classes obtained from the aerial image. Each mask distinctly isolates one particular class, which can be better understood and where the features are. Water Mask: White marks bodies of water. It covers 104,188 pixels, 3.28% of the entire image area. Building Mask: Displays buildings within the area, occupying 551,032 pixels or 17.34% of the area. Road Mask: Masks the road network distinctly, using 326,490 pixels, or 10.27% of the map. Vegetation Mask: Reveals areas of green vegetation, covering 28,831 pixels, or about 0.91% of the image. These masks are employed separately to study the land with great detail. They help in viewing certain details like buildings, resources, and the evolution of cities.

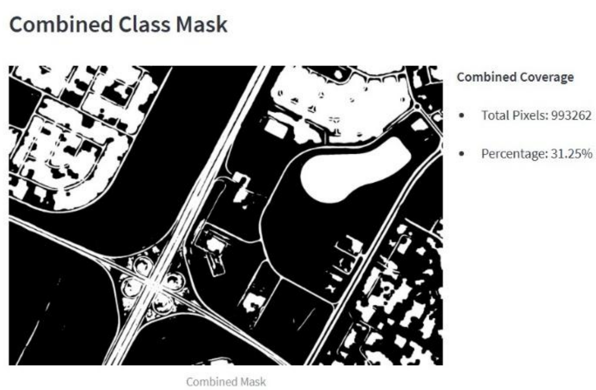

FIGURE 5.3 Combined Class Mask

Figure 3: Combined Class Mask The image above illustrates a combined binary mask generated by merging the selected individual class specifically, water, buildings, roads, and vegetation. In this combined representation, all detected regions belonging to any of the selected classes are shown in white, while the background remains black. This combined mask simplifies visualization when the overall land cover distribution or performing calculations that require the union of multiple classes. The combined coverage consists of Total Pixels: 993,262 Percentage of Image Area: 31.25%Such a composite view is especially useful for tasks like monitoring urban expansion, estimating resource distribution, or preparing inputs for further spatial analysis.

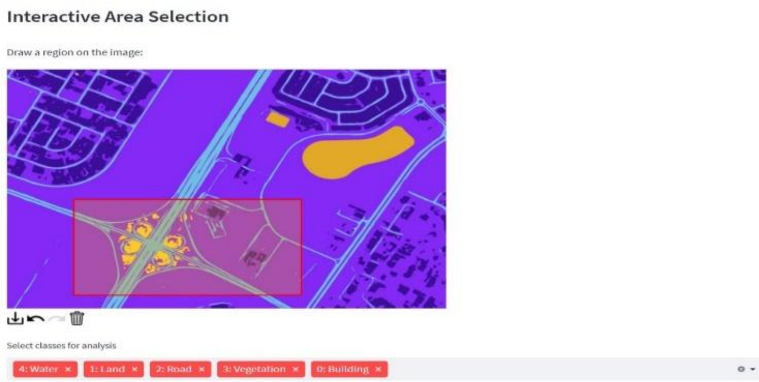

FIGURE 5.4 Interactive Area Selection

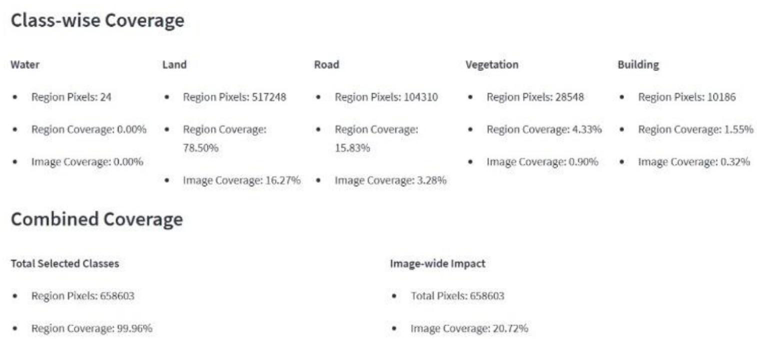

FIGURE 5.5 Area covered

Figures 4 and 5 show the Interactive Area Selection

Fig. 4. Interactive Area Selection Fig. 5. Area covered tool, A particular rectangular region has been found in a segmented aerial photograph to analyze the pattern of various land covers. The selected site has principal characteristics of infrastructure, i.e., a road the intersection and surrounding land plots. Five have has been chosen to be researched in this field: Water (class 4), Land (class1), Road (class 2), Vegetation (class 3), and Building (class 0). Class-wise coverage statistics indicate that the Land class has the selected area, which includes 517,248 pixels account for 78.50% of the region's area and conmakes up 16.27% of the total image area. The category of Road are listed below, 104,310 pixels (15.83% region coverage, 3.28% image coverage, whereas Vegetation and Buildings Classes occupy 28,548 pixels (4.33%) and 10,186 pixels. (1.55%) of land, respectively. What they do to the image coverage are 0.90% and 0.32%. The Water class is not so prevalent in the selected region, with a mere 24 pixels, which is virtually nothing at 0.00% in both regions. and photo coverage. Combined, the selected categories have 658,603 pixels. in the region addressed here, making a common space 99.96% coverage. From a general point of view, this the chosen region occupies 20.72% of the whole image, emphasizing its leading position in the group analysis.

5. Conclusion

This project demonstrates the powerful capabilities of deep learning models, particularly semantic segmentation architectures such as UNet and UNet++, in accurately analyzing and interpreting aerial imagery. By leveraging high-resolution satellite datasets, the system was able to distinguish and classify various land cover types—including buildings, roads, vegetation, and water bodies—with high precision and reliability.

The entire workflow, from dataset preparation and preprocessing to model training and deployment, was meticulously designed to ensure scalability, efficiency, and accuracy. The implementation of advanced techniques like transfer learning, data augmentation, and fine-tuning of pretrained models significantly improved the model's generalization ability across different geographical terrains.

Moreover, the development of a web-based user interface made the system accessible and user-friendly, enabling real-time image segmentation for end-users such as urban planners, environmental researchers, and disaster response teams. The model's robust performance on diverse aerial images validates its potential for integration into larger geospatial analysis platforms.

In conclusion, this project not only highlights the practicality of deep learning in remote sensing applications but also paves the way for future advancements in automated geographic feature extraction. With further improvements—such as integrating multi-spectral data, real-time satellite feeds, and more refined labeling techniques—the system can evolve into a comprehensive solution for smart mapping, land use monitoring, and infrastructure planning on a global scale.

6. References

[1]. T. Bakirman, I. Komurcu, and E. Sertel, "Comparative analysis of deep learning based building extraction methods with the new VHR Istanbul dataset," Expert Systems with Applications, vol. 2022, 2022.

[2]. B. Amirgan and A. Erener, "Semantic segmentation of satellite images with different building types using deep learning methods," Remote Sensing Applications: Society and Environment, vol. 2024, 2024.

[3]. Z. Wang, S. Sun, X. Que, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[4]. Panuntun, Y.-N. Chen, I. Jamaluddin, and T. L. C. Tran, "Evaluation of Deep Learning Semantic Segmentation for Land Cover Mapping on Multispectral, Hyperspectral and High Spatial Aerial Imagery," arXiv preprint arXiv:2406.14220, 2024.

[5]. R. Doghmane and K. Boukari, "Enhanced U-Net model for accurate aerial road segmentation," Machine Graphics & Vision, vol. 33, no. 3/4, pp. 71–96, 2024.

[6]. Z. Wang, S. Sun, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[7]. Y. Li, Y. Zhang, and J. Wang, "A new framework for improving semantic segmentation in aerial images," Frontiers in Remote Sensing, vol. 2024, 2024.

[8]. B. Amirgan and A. Erener, "Semantic segmentation of satellite images with different

building types using deep learning methods," Remote Sensing Applications: Society and Environment, vol. 2024, 2024.

[9]. Z. Wang, S. Sun, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[10]. Panuntun, Y.-N. Chen, I. Jamaluddin, and T. L. C. Tran, "Evaluation of Deep Learning Semantic Segmentation for Land Cover Mapping on Multispectral, Hyperspectral and High Spatial Aerial Imagery," arXiv preprint arXiv:2406.14220, 2024.

[11]. R. Doghmane and K. Boukari, "Enhanced U-Net model for accurate aerial road segmentation," Machine Graphics & Vision, vol. 33, no. 3/4, pp. 71–96, 2024.

[12]. Z. Wang, S. Sun, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[13]. Y. Li, Y. Zhang, and J. Wang, "A new framework for improving semantic segmentation in aerial images," Frontiers in Remote Sensing, vol. 2024, 2024.

[14]. B. Amirgan and A. Erener, "Semantic segmentation of satellite images with different building types using deep learning methods," Remote Sensing Applications: Society and Environment, vol. 2024, 2024.

[15]. Z. Wang, S. Sun, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[16]. Panuntun, Y.-N. Chen, I. Jamaluddin, and T. L. C. Tran, "Evaluation of Deep Learning Semantic Segmentation for Land Cover Mapping on Multispectral, Hyperspectral and High Spatial Aerial Imagery," arXiv preprint arXiv:2406.14220, 2024.

[17]. R. Doghmane and K. Boukari, "Enhanced U-Net model for accurate aerial road segmentation," Machine Graphics & Vision, vol. 33, no. 3/4, pp. 71–96, 2024.

[18]. Z. Wang, S. Sun, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[19]. Y. Li, Y. Zhang, and J. Wang, "A new framework for improving semantic segmentation in aerial images," Frontiers in Remote Sensing, vol. 2024, 2024.

[20]. B. Amirgan and A. Erener, "Semantic segmentation of satellite images with different building types using deep learning methods," Remote Sensing Applications: Society and Environment, vol. 2024, 2024.

[21]. Z. Wang, S. Sun, and X. Ma, "Interactive segmentation in aerial images: a new benchmark and an open access web-based tool," arXiv preprint arXiv:2308.13174, 2023.

[22]. Panuntun, Y.-N. Chen, I. Jamaluddin, and T. L. C. Tran, "Evaluation of Deep Learning Semantic Segmentation for Land Cover Mapping on Multispectral, Hyperspectral and High Spatial Aerial Imagery," arXiv preprint arXiv:2406.14220, 2024.

[23]. R. Doghmane and K. Boukari, "Enhanced U-Net model for accurate aerial road segmentation," Machine Graphics & Vision, vol. 33, no. 3/4, pp. 71–96, 2024.

[24]. Khan, B.A. and Jung, J.W., 2024. Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions. Applied Sciences, 14(9), p.3712.

[25]. Amirgan, B. and Erener, A., 2024. Semantic segmentation of satellite images with different building types using deep learning methods. Remote Sensing Applications: Society and Environment, 34, p.101176.

[26]. Ramos, L. and Sappa, A.D., 2024. Multispectral Semantic Segmentation for Land Cover Classification: An Overview. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

[27]. Gui, S., Song, S., Qin, R. and Tang, Y., 2024. Remote sensing object detection in the deep learning era—a review. Remote Sensing, 16(2), p.327.

[28]. Meng, X., Zhu, L., Han, Y. and Zhang, H., 2023. We Need to Communicate: Communicating Attention Network for Semantic Segmentation of High-Resolution Remote Sensing Images. Remote Sensing, 15(14), p.3619.

[29]. Li, X., Jiang, Y., Peng, H. and Yin, S., 2019, May. An aerial image segmentation approach based on enhanced multi-scale convolutional neural network. In 2019 IEEE international conference on industrial cyber physical systems (ICPS) (pp. 47-52). IEEE.

[30]. Hua, Y., Marcos, D., Mou, L., Zhu, X.X. and Tuia, D., 2021. Semantic segmentation of remote sensing images with sparse annotations. IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.

[31]. Shan, L., Wang, W., Lv, K. and Luo, B., 2022. Class-incremental semantic segmentation of aerial images via pixel-level feature generation and task-wise distillation. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-17.

[32]. Yang, N. and Tang, H., 2021. Semantic segmentation of satellite images: A deep learning approach integrated with geospatial hash codes. Remote Sensing, 13(14), p.2723.

[33]. StarUML Documentation. n.d. Class Diagram – StarUML Documentation. [online] Available at: https://docs.staruml.io/working-with-uml-diagrams/class-diagram.