WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Semantic Segmentation of Remote Sensing Images of Urban Areas using Deep Learning Methods

V. Pranathi[1], D. Vignan[2], B. Akshay[3], B. Yaswanth[4], Dr. S. Akila Agnes[5]

[1,2,3,4] Computer Science and Engineering GMR Institute of Technology Srikakulam, India
[5]Assistant Professor, Computer Science and Engineering, GMR Institute of Technology Srikakulam, India akila.s@gmrit.edu.in
Email: [1]21345A0501@gmrit.edu.in, [2]20341A0548@gmrit.edu.in, [3]20341A0533@gmrit.edu.in, [4]20341A0531@gmrit.edu.in
Doi: https://doi.org/10.55248/gengpi.5.0324.0779

ABSTRACT—

As our world is attempting to step into a better innovative future, the need for smart cities is demanding. The major aspect of the smart city consists of environmental monitoring, innovative navigation, urban area planning and traffic management etc. Semantic segmentation of Remote Sensing images have a vital role in the development and the growth of smart cities. Remote sensing has been shown to be a very useful and efficient technique for mapping regions. In the pursuit of accurately detecting and segmenting specific regions on land, the collection of relevant data is as important as it sounds. Satellite imagery emerges as a paramount source, providing detectable insights for the detection and segmentation of areas. By harnessing the rich details in satellite images, precise information can be drawn. This approach allows us for a good analysis of the urban area. Given that the U-NET is currently a leading approach for accurate detection and segmentation, there is a growing trend to enhance its MIOU and overall performance by incorporating transformers into the model. This integration attempts to increase the model's precision and aid in its general improvement.

Keywords—Semantic segmentation, Remote sensing images, Transformer, Segmentation, Deep Learning

I. INTRODUCTION

Remote sensing image segmentation is an invaluable tool in our current context and for the future due to its precision and accuracy in identifying and classifying features within captured imagery. Its utility lies in early detection and continuous monitoring of specific areas, enabling predictive analyses. Ranging from disaster management and traffic control to construction planning, there are many situations were one need segmentation of the area. Acknowledging the growing demand and rapid advancements in image segmentation, the technology has been progressing at a commendable pace. We have expanded our model-building approach beyond the U-net architecture by incorporating two additional models, namely, attention Unet, Deeplab v3 plus and Swim-Unet transformers, for segmentation. The comparison of these models will ultimately determine which one performs better.. Pleiades-1A true-color, high-resolution satellite imagery was used to construct the collection.

An Airbus product called Pleiades offers imagery at several spectral combinations with a resolution of 0.5 meters. A total of 110 patches, each measuring 600 by 600 pixels, were chosen by visually spotting random spots throughout the city that had a range of urban features, including slums, vegetation, developed, highways, etc. The photos were divided into six distinct classes: (1) greenery; (2) built-up areas; (3) informal settlements; (4) impervious surfaces (streets, parking lots, highways, and regions resembling roadways around buildings); (5) bare; and (6) water. The dataset has six primary classifications, plus an additional class called "Unlabelled," which accounts for 0.08% of the total. Literature survey

Zhang, C., Jiang, W.,[1] and others presented a novel deep neural network for precisely segmenting high-resolution remote sensing pictures. It blends transformer and convolutional neural network (CNN) models. In order to enhance segmentation outcomes, the network employs a variety of strategies, including attention blocks and skip connections. In benchmark testing, the network obtains good accuracy. The decoder module uses a variety of building blocks and techniques to improve border detection accuracy, preserve local details, increase feature extraction, and restore feature map size. The International Society for Photogrammetry and Remote Sensing (ISPRS) Vaihingen and Potsdam benchmarks were the focus of the paper's extensive investigations.

The research paper [2] emphasizes the usage of the Modified Dingo Optimizer with Deep Learning (MSCRSI- MDODL) Approach for Remote Sensing Imagery-Based Multi-Label Scene Classification. In his discussion of the shortcomings of traditional techniques for classifying remote sensing images (RSI), Ragab, M. [2] presents a fresh strategy that makes use of techniques as well as an altered optimization algorithm. Using an optimum hyperparameter tuning algorithm, the suggested MSCRSI-MDODL technique merges attention Squeeze and Excitation (SE) with DenseNet model for feature extraction. The methodology entails hyperparameter optimization of the upgraded DenseNet architecture via the MDO algorithm. For scene categorization, the

method also makes use of the stacked dilated convolutional autoencoders (SDCAE) model. The MSCRSI- MDODL method yielded an average precision of 98.41% and an accuracy of 99.84%.

This study [3] by Zhang, L., Lu intends to provide a partially unlabeled remote-sensing picture segmentation method using a semisupervised convolution neural network. The objective is to enhance the quality of pseudo-labels by contrastive loss and self-training techniques, and to lessen the dependence on manually annotated pixel-level labels in order to improve the performance of semantic segmentation algorithms for remote sensing images. In comparison to earlier approaches, the study seeks to obtain a greater Mean

Intersection over Union (mIOU). The pixel-level and region- level contrastive loss approaches are the ones covered in this section. The suggested method for semi-supervised segmentation of remote-sensing photos outperforms other current semi-supervised algorithms on the POTSDAM and Vahingen datasets, according to the experimental results described in the paragraph. High segmentation accuracy is achieved by the method even with a limited number of annotated examples.reducing reliance on labeled data.

The purpose of the paper [4] is to assess FCNs' ability to transfer knowledge in connection to slum mapping in various satellite images. Data from Sentinel-2 and TerraSAR-X are fed into a model developed with very high resolution optical satellite imagery from QuickBird. The results of segmentation are significantly improved when a pre- trained network is transferred from QuickBird images to Sentinel-2 images. This makes it possible to use medium resolution sensors at 10 m GSD to map slums over very huge areas, even entire countries or subcontinents. The optical data utilized in this article demonstrates very high accuracies for mapping slums: Sentinel-2 applying transfer learning shows a considerable gain (from 38 to 55% and from 79 to 85% for PPV and sensitivity), while the QuickBird image achieves 86– 88% (PPV and sensitivity).

The main objective of Hua, Y., Marcos, D., [5] to introduce a method called FESTA to leverage these sparse annotations and improve the performance of semantic segmentation while reducing labeling costs. The objective of the paper is to propose a framework and method for semantic segmentation of high-resolution aerial images using incomplete annotations. The aim is to address the labor- intensive and time-consuming process of pixel-level annotation by leveraging sparse annotations in the form of easy-to-draw scribbles. The numerical and visual results show an improvement in semantic segmentation when using different types of sparse annotations.

Pan, S., and Tao, Y. [6] want to propose a Progressive Edge Guidance Network (PEGNet) for semantic segmentation of remote sensing pictures. The network is specifically designed to integrate edge detection and semantic segmentation in order to enhance the discriminative power of the model. The proposed PEGNet consists of an edge branch and a segmentation branch. While the edge branch forecasts an edge-region map utilizing dilated edge information and deep semantic knowledge, the segmentation branch uses a guidance module to gradually retrain error- prone pixels. The proposed PEGNet uses a multipath atrous module to augment deep semantic information and combine it with dilated edge information to obtain edge-region maps. The proposed Progressive Edge Guidance Network (PEGNet) achieved a new benchmark with an overall performance on the ISPRS Vaihingen test set.

The research article by Lilay, M. Y., & Taye, G. D. [7]] aimed to create a semantic segmentation model for classifying land cover from satellite pictures, with a particular emphasis on resolving the shortcomings of earlier research and offering precise land cover classification in the Ethiopian region of Gambella National Park (GNP). The goal was to evaluate the differences in land cover categorization performance between deep learning and machine learning classifiers. Using high- resolution Sentinel-2 satellite pictures, the researchers proposed a deep learning-based semantic segmentation model and assessed the models' performance using both deep learning and traditional machine learning classifiers. The study's findings demonstrated that the created models for semantic segmentation obtained average F1-Score values of 83%, 82%, and 87.4% for the LinkNet model, random forest with CNN features (CNN-RF), and support vector machine with CNN features (CNN-SVM).with ResNet-34 as encoder (LinkNet-ResNet34) respectively.

Singh, N. J., & Nongmeikapam, K. [8] created and assessed a DeepUNet model with an emphasis on land cover mapping for semantic segmentation of satellite photos using multispectral data. Convolutional models have to be trained and tested in order to map land cover, assess its applicability, and detect changes in land cover. Obtaining ground-level data sets is complex and time-consuming, which makes it difficult to create training datasets for satellite image segmentation. Another restriction is the absence of labeled training data. This work describes the creation and assessment of a DeepUNet model for semantic segmentation of satellite pictures, demonstrating its effectiveness and precision over alternative approaches. According to the study findings, the suggested DeepUNet model performs better than alternative approaches, obtaining a mean Intersection overUnion (mIoU) of 89.51% and a global accuracy of 90.6%.

The aim of the document [9] aims to present a brand-new multi-level adaptive-scale context aggregating network (MACANet) for high-resolution remote sensing (HR2S) image semantic segmentation. It introduces the design of MACANet—a sequential aggregation block (SAB) and an adaptive-scale context extraction block (AS-CEB)—and demonstrates how well it works to get superior segmentation results on benchmark HR2S datasets. The paper addresses the difficulties presented by HR2S images, including complex features and land objects with wide scale variances, and highlights the significance of multiscale information extraction for efficient semantic segmentation. A sequential aggregate block (SAB) and an adaptive-scale context extraction block (AS-CEB) make up the MACANet architecture. According to the performance comparison, MACANet performs best on the Vaihingen and Potsdam, achieving the highest mean F1 (mF1) and mean pixel accuracy (mPA).datasets, outperforming recent multiscale feature aggregation methods designed for natural images and HR2S images.

Wujie Zhou, Jianhui Jin, Jingsheng Lei, and Lu Yu [10] use CIMFNet model, which increases the segmentation accuracy. The main advatages are Cross-layer intersection modules, multi scale feature fusion and attention mechanism. The network architecture incorporates cross-layer connections, enabling effective information flow between different layers, promoting enhanced contextual understanding. The main advantages are Cross-layer intersection

modules, multi scale feature fusion and attention mechanism. It depends on some specific datasets and long-term performance. The metrics used in this paper are accuracy, IOU and MIOU.

The Multiscale Progressive Segmentation Network (MPS- Net) emerges as an innovative solution tailored for the intricate domain of high-resolution remote sensing imagery. Jianhui Jin, Wujie Zhou [11] used Edge Detection Guide Network (EDGNet). The aim of the paper is to explore the spatial position information, modular and explainable design. In this paper the overfitting to edge information, lack of explorations on light-weight versions. It has limited hyper- parameters, ignoring temporal information, it focus only bench mark datasets. In this paper the metrics used are miou.

Haifeng Li [12] proposes a novel self-supervised learning approach (GLSSL), it also utilizes a backbone network for feature extraction. By ingeniously combining global and local contrastive learning strategies, the framework autonomously extracts meaningful features from the intricate details present in high-resolution imagery. This paper aims to focus on high- resolution images, self-supervised learning. In this work there is limited handling of class imbalance. The metrics used are accuracy and kappa.

Rui Li [13] propses a MANet (multi attention network)model.this is the proposed novel architecture for incorporating multi-scale and local-global contextual information into the segmentation process. This paper aims to focus on fine resolution images and Multi-Scale and Local- Global Attention The multiattention approach enables the network to simultaneously focus on multiple salient regions within the high-resolution images, addressing the intricate details and subtle features crucial for accurate segmentation.

The multi attention approach enables the network to simultaneously focus on multiple salient regions within the high-resolution images, addressing the intricate details and subtle features crucial for accurate segmentation. This collaborative effort signifies a notable stride forward in the field, offering a promising solution for the intricate task of semantic segmentation in fine-resolution remote sensing applications. Renlong Hang[14] uses Multiscale Progressive Segmentation Network (MPSNet) is one of the major advantage. This work has lack of credibility , bias, and incoherence. The limitations of this work are sample size, data collection and time frame. The accuracy and mlou are the metrics used in this paper.

Gaston Lenczner , Adrien Chan-Hon-Tong[15] addresses the unique challenges posed by high-resolution remote sensing, offering a holistic approach to segmentation. This innovative design enhances the network's capability to discern complex features within the imagery, promising a more nuanced and accurate segmentation. The collaborative endeavor underscores the technical robustness of MPS-Net, positioning it as a promising advancement in automated image analysis for remote sensing applications. This paper mainly focuses on experimental validation, active learning strategy, comparison of acquisition functions. Lack of specific model details, user expertise requirement, dataset bias and scope are the main disadvantages of this paper.

II. METHODOLOGY

The figure-1 depicts the workflow of the model from step one. There are two main stages in the above flow chart one is data preprocessing and the second one is model training.

Data Preprocessing:

In the data preprocessing step the image is converted into an array and normalization is done. For mask images, they are encoded by replacing the pixel values in RGB format with corresponding class labels. A total of 7128 images are used for training and 891 images for validation.

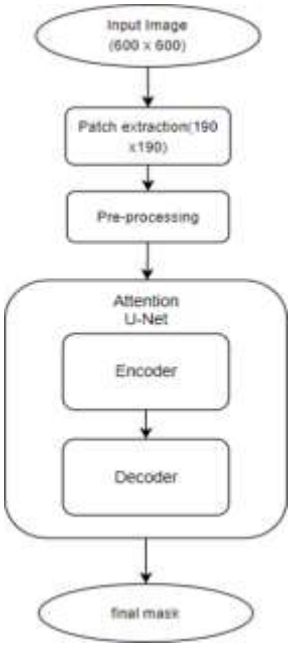

Fig-1 Flow Chart

Attention U-Net: By including an attention component, the Attention U-Net architecture expands upon the conventional U-Net concept. With this update, the model is better able to concentrate on pertinent characteristics throughout the processing of input photos, which improves segmentation performance, especially in situations where collecting minute details is essential, like medical image segmentation.

Structure of Encoder-Decoder: The Attention U-Net has an encoder-decoder structure, just as the U-Net design. The input image is gradually downsampled by the encoder to extract high-level features, which are then upsampled by the decoder to produce the final segmentation map.

Skip Connections: The integration of low-level and high- level features is facilitated by skip connections between relevant encoder and decoder layers, which allows the model to capture contextual and detailed information.

Attention Gates: The Attention U-Net's unique characteristic is the incorporation of attention gates within the skip connections. These attention gates dynamically modulate the flow of information across the network by assigning importance weights to features at different spatial locations.

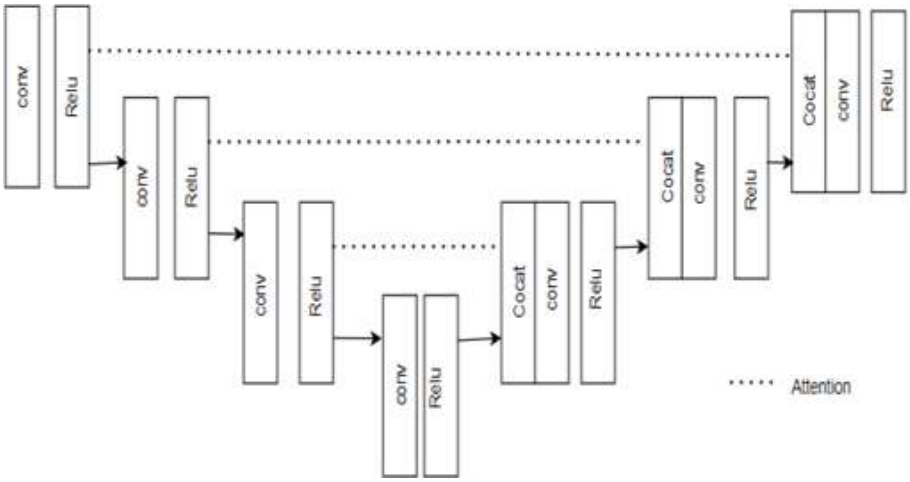

Fig-2 Attention U-Net

Attention Mechanism:

Channel-Wise Attention: The attention mechanism in Attention U-Net operates at both the spatial and channel-wise levels. It computes attention maps for each feature map channel, allowing the model to selectively attend to relevant channels.

Spatial Attention: Additionally, spatial attention mechanisms are employed to focus on informative regions within feature maps. This ensures that the model allocates more resources to regions of interest, enhancing segmentation accuracy.

III. RESULTS AND DISCUSSION

Unet

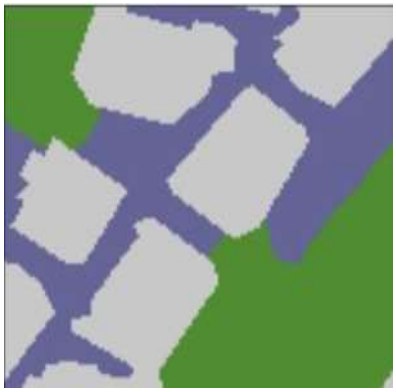

Fig-3 Original Mask

Fig-4 Predicted Mask

Attention Unet

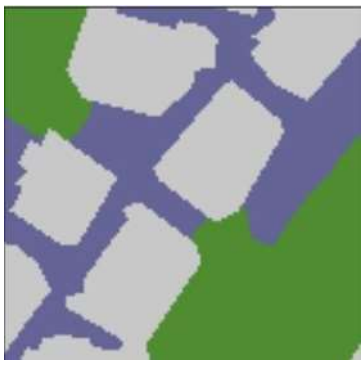

Fig-5 Original Mask

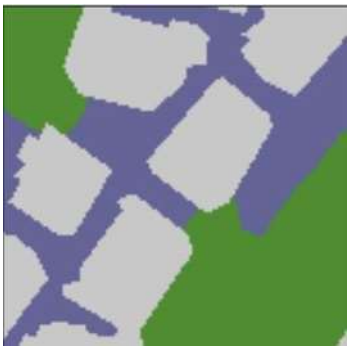

Fig-6 Predicted Mask

Model	MIOU
U-Net	0.89
Attention U-Net	0.94

Table-1 Results

Initially, we employed the U-Net architecture for remote sensing image segmentation. Despite its widely recognized effectiveness in semantic segmentation tasks, we encountered certain limitations, particularly in accurately delineating complex structures within the images. The U-Net model exhibited decent performance, as evidenced by the achieved IoU scores. However, there were noticeable inaccuracies in segmenting fine details and handling class imbalances within the dataset. To address the aforementioned challenges and enhance the segmentation accuracy, we transitioned to the Attention U-Net architecture. By incorporating attention mechanisms within the network, we aimed to improve the model's ability to focus on relevant features and alleviate the limitations observed with the traditional U-Net.

Our experiments with Attention U-Net yielded promising results, showcasing noticeable improvements over the baseline U-Net architecture. The attention mechanism played a crucial role in enhancing feature learning and localization capabilities, thereby leading to more accurate segmentation results. The model demonstrated a higher level of precision in delineating intricate structures within the remote-sensing images, effectively addressing the previously encountered inaccuracies.

Comparing the performance metrics between U-Net and Attention U-Net, we observed a discernible increase in the IoU scores with the latter architecture. The attention mechanism facilitated better feature extraction and attention focusing, resulting in more refined segmentation outputs. Notably, the Attention

U-Net exhibited enhanced performance in scenarios with complex background structures and class imbalances, where traditional U-Net models often struggled.

IV. CONCLUSION

In conclusion, our experiments demonstrate the effectiveness of Attention U-Net in remote sensing image segmentation tasks. By leveraging attention mechanisms, the model exhibits improved feature learning and localization capabilities, leading to more accurate and refined segmentation results compared to traditional U-Net architectures. The integration of attention mechanisms represents a significant advancement in semantic segmentation techniques, offering promising avenues for further research and application in various domains, including remote sensing.

REFERENCES

[1] Zhang, C., Jiang, W., Zhang, Y., Wang, W., Zhao, Q., & Wang, C. (2022). Transformer and CNN hybrid deep neural network for semantic segmentation of very-high- resolution remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-20.

[2] Ragab, M. (2023). Multi-Label Scene Classification on Remote Sensing Imagery using Modified Dingo Optimizer with Deep Learning. IEEE Access

[3] Zhang, L., Lu, W., Zhang, J., & Wang, H. (2022). A Semisupervised Convolution Neural Network for Partial Unlabelled Remote-Sensing Image Segmentation. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

[4] Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS journal of photogrammetry and remote sensing, 150, 59-69Jiang, X., Wang, N., Xin, J., Xia, X., Yang, X., & Gao, X.

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

[6] Pan, S., Tao, Y., Nie, C., & Chong, Y. (2020). PEGNet: Progressive edge guidance network for semantic segmentation of remote sensing images. IEEE Geoscience and Remote Sensing Letters, 18(4), 637-641.

[7] Lilay, M. Y., & Taye, G. D. (2023). Semantic segmentation model for land cover classification from satellite images in Gambella National Park, Ethiopia. SN Applied Sciences, 5(3), 76.

[8] Singh, N. J., & Nongmeikapam, K. (2023). Semantic segmentation of satellite images using deep- UNet. Arabian Journal for Science and Engineering, 48(2), 1193-1205.

[9] Li, X., Lei, L., & Kuang, G. (2021). Multilevel adaptive- scale context aggregating network for semantic segmentation in high-resolution remote sensing images. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

[10] CIMFNet: Cross-Layer Interaction and Multiscale Fusion Network for Semantic Segmentation of High- Resolution Remote Sensing Images Wujie Zhou , Jianhui Jin, Jingsheng Lei, and Lu Yu, Senior Member IEEE

[11] Edge Detection Guide Network for Semantic Segmentation of Remote-Sensing Images Jianhui Jin, Wujie Zhou , Member, IEEE, Rongwang Yang, Lv Ye, and Lu Yu, Senior Member, IEEE

[12] Global and Local Contrastive Self-Supervised Learning for Semantic Segmentation of HR Remote Sensing Images Haifeng Li , Member, IEEE, Yi Li , Guo Zhang , Ruoyun Liu, Haozhe Huang , Qing Zhu , and Chao Tao

[13] Multi attention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images Rui Li , Member, IEEE, Shunyi Zheng , Ce Zhang , Chenxi Duan , Member, IEEE, Jianlin Su, Libo Wang , Graduate Student Member, IEEE, and Peter M. Atkinson

[14] Multiscale Progressive Segmentation Network for High- Resolution Remote Sensing Imagery Renlong Hang , Member, IEEE, Ping Yang, Graduate Student Member, IEEE, Feng Zhou , Member, IEEE, and Qingshan Liu , Senior Member, IEEE.

[15] DIAL: Deep Interactive and Active Learning for Semantic Segmentation in Remote Sensing Gaston Lenczner , Adrien Chan-Hon-Tong, Bertrand Le Saux , Senior Member, IEEE, Nicola Luminari, and Guy Le Besnerais