Satellite Image Segmentation Using U-Net

Shikhar Singh Yadav[1], Rohini Raji[2], Mahika Tyagi[3], Dr.K.P. Jayant[4]

[1,2,3]Department of Computer Sciences and Engineering, Raj Kumar Goel Institute of Technology, Uttar Pradesh, India
[4]Professor in Computer Sciences and Engineering, Raj Kumar Goel Institute of Technology, Uttar Pradesh, India

Abstract
Satellite image segmentation is a core process in remote sensing applications that enables land cover classification, urban planning, and environmental monitoring. In this work, we introduce a deep learning-based segmentation model based on the U-Net architecture for pixel-wise classification of high-resolution satellite images. The model is trained on a satellite image dataset and its respective labeled mask to learn geographical features properly. To enhance segmentation performance even further, we employ data augmentation and hyperparameter tuning to enhance generalization. The model is assessed based on the Intersection over Union (IoU) metric with an IoU metric score of approximately 0.8, which shows high segmentation accuracy. The experimental outcomes prove that the U-Net architecture is exceptionally suitable for satellite image segmentation and provides a promising approach to real-world remote sensing applications. Future work will explore further generalization enhancement by adding attention mechanisms and multi-scale feature fusion.

1. INTRODUCTION

In the last few decades, satellite image analysis has emerged as an essential technique to monitor, observe, and regulate the dynamic Earth surface. Satellite images support high-resolution, wide-area coverage with viewposts, which is an asset for various applications including environmental monitoring, urban planning, agriculture, emergency response, and climate change investigation. But the vast scale and variability of satellite data often characterized by spatial variability in land cover and complex spatial patterns pose very difficult challenges to effective interpretation and use of it.

Satellite image segmentation is a core remote sensing technique that entails partitioning an image into significant segments or regions each identifying with certain land cover classes like vegetation, water bodies, built-up land, and bareland. Such pixel-level categorization facilitates more detailed analysis and plays a critical role in automating several geospatial operations with resulting enhanced decision-making in the public and private spheres.

The primary objective of this project is to create and implement an exhaustive and accurate satellite image segmentation system using advanced machine learning and deep learning models. Borrowing extensively from current state-of-the-art models such as Convolutional Neural Networks (CNNs) and encoder-decoder models such as U-Net, the system in this work would simplify segmentation and enhance the accuracy of the classification result. The activities involved include data acquisition, preprocessing (normalization and noise filtering), model training, and mask creation for segmented areas. Except for addressing the computational and algorithmic challenges of high-resolution satellite imagery, the project also aims to contribute to the creation of scalable, intelligent solutions to geospatial analysis. Through unleashing the potential of artificial intelligence, the system will process raw satellite imagery as actionable knowledge, ultimately resulting in data-driven solutions to environmental management, infrastructure planning, and policy making.

2. Abbreviations and Acronyms

Abbreviations and Acronyms are shortened words or phrases, usually used to make technical or frequently occurring words in a document brief. In technical or research writing, particularly in fields such as satellite image segmentation, they are used to bring quick clarity and conciseness.

Table 1: Table of Abbreviation/Acronym

Abbreviation/ Acronym	Full Form	Description
CNN	Convolutional Neural Network	Deep learning model used for image feature extraction and classification.
U-Net	U-Net	A CNN architecture designed for biomedical and satellite image segmentation.
GA	Genetic Algorithm	Optimization algorithm inspired by natural selection.
FCN	Fully Convolutional Network	Neural network for dense prediction tasks like segmentation.
GIS	Geographic Information System	System for capturing, storing, and analyzing spatial and geographic data.
ROI	Region of Interest	Specific part of an image identified for detailed analysis.
RGB	Red Green Blue	Color model used in image data.
IoU	Intersection over Union	Metric used to evaluate segmentation accuracy.
ML	Machine Learning	Branch of AI focused on learning from data.
DL	Deep Learning	Subset of ML using neural networks with many layers.
API	Application Programming Interface	Set of protocols for building and integrating application software.
GPU	Graphics Processing Unit	Hardware that accelerates deep learning computations.
DRL	Data Representation Layer	Intermediate layer for storing processed image features.

3. Related Work

Satellite image segmentation has received tremendous interest over the last decades with numerous applications in remote sensing, environmental monitoring, and urban planning. A number of techniques have been proposed as an attempt to improve segmentation quality and computation time, ranging from traditional evolutionary algorithms to very recent deep learning-based approaches.

One of them is provided by Malik. (2023), in their detailed review "Satellite Image Segmentation Using Neural Networks." The article provides a detailed overview of neural network-based segmentation methods with special focus on the application of deep learning models like Convolutional Neural Networks (CNNs), Fully Convolutional Networks FCNs. The article also provides some limitations like the high requirement for annotated data and computational resources.

in another seminal work, Pare., in their paper "Satellite Image Segmentation based on Different Objective Functions using Genetic Algorithm: A Comparative Study," present traditional optimization-based methods of segmentation. They compare some of the objective functions such as Otsu, Kapur's Entropy, and Minimum Cross Entropy with Genetic Algorithms (GAs). They note that while GA-based methods can achieve good segmentation and are simple to implement for various objective functions, they are marked by low convergence rates and parameter sensitivity. These techniques are particularly limiting for high-resolution satellite imagery or images with subtle and intricate features, where deep learning models prevail.

Though both studies are of value to the community, the transition from rule-based and heuristic approaches (e.g., GAs) to data-driven methods (e.g., CNNs) represents a milestone for satellite image segmentation. U-Net, in particular, has performed outstandingly in segmenting satellite images because it is able to extract pixel-level features of high level, which makes it well-suited for tasks such as land cover classification and change detection.

Table 2: Table of Model and Technique

Model and Technique	Year of Paper	Limitations
Convolutional Neural Networks (CNN)	2023	Limited context awareness; struggles with capturing global spatial features.
Fully Convolutional Networks (FCN)	2023	Coarse segmentation boundaries; lacks refinement in edge detection.
DeepLab v3+	2023	High computational cost; requires significant labeled data for training.
Mask R-CNN	2023	Slower inference time; may miss small segmented objects in high-res images.
ResNet-Based Encoders	2023	Susceptible to overfitting with small datasets; increased complexity.
SegNet	2023	Struggles with fine-grained segmentation due to aggressive downsampling.
GAN-based Segmentation	2023	Training instability; complex to implement and tune effectively.

In this project,the U-Net architecture is the primary deep learning model for performing satellite image segmentation. U-Net is a type of convolutional neural network (CNN) that is specifically designed for semantic segmentation tasks, thus ideally suited for labeling every pixel in high-resolution satellite images. Its architecture has an encoder-decoder structure where the encoder pathway gets the image context by downsampling, and the decoder pathway builds the segmentation map by upsampling. Skip connections between matching encoder and decoder pathway layers are one of the exceptional features of U-Net, which helps in maintaining spatial information and boundary detection precision. By using this architecture, the model can distinguish well between different land cover classes such as vegetation, water bodies, buildings, and barren land. The U-Net model is trained on preprocessed satellite data, and its ability to learn from relatively small datasets while generating high-precision segmentation makes it ideally suited to real-world geospatial applications discussed in this project.

4. Data Collection and Data Processing

4.1. Data Collection

Data acquisition is an extremely important phase in satellite image segmentation since the quality and diversity of data directly affect the performance of the segmentation model. For this project, satellite images are downloaded from public remote sensing resources like Landsat, Sentinel-2, or Google Earth Engine, which provide high-resolution multispectral images ideal for land cover analysis. The downloaded images represent a variety of geographical locations, seasons, and environments to enable the model to generalize well over different terrain. Each image is accompanied by relative ground truth labels or segmentation masks, which are hand-annotated or obtained from authoritative geospatial datasets like CORINE (Coordination of Information on the Environment) or MODIS land cover products. For pre-training the data, cloud masking, resizing, normalization, and augmentation (rotation, flipping, and scaling) are performed. This preprocessed and cleaned dataset is used as the base to train the U-Net model to learn pixel-wise classification of different land covers with high accuracy.

Figure 1: Satellite Image over a Forest

Figure 2: Satellite Image over a dried Vegetation (Sandy)

4.2. Data Processing

Data processing plays an essential role in satellite image segmentation as it cleans raw satellite data into a homogeneous format for model training and testing. It begins with cloud masking, which eliminates covered areas in the images to prevent wrong predictions.This is preceded by resampling images to a uniform spatial resolution to maintain data consistency in the dataset.Normalization is performed to normalize pixel values (usually 0 to 1) to ease model convergence simplicity during training. Spectral band selection is also done to select useful channels—e.g., red, green, blue (RGB) and near-infrared (NIR)—depending on the segmentation task. For improving model performance and stability, data augmentation techniques such as flipping, rotation, cropping, and brightness modification are employed to synthesize the dataset and simulate various environmental conditions. All images and the respective ground truth segmentation masks are finally aligned, resized, and converted to tensors to ready the input-output pairs for training the deep learning model. Such robust data processing pipeline ensures the segmentation model is furnished with high-quality, homogeneous, and diverse training data.

Figure 3. processing of Satellite Image

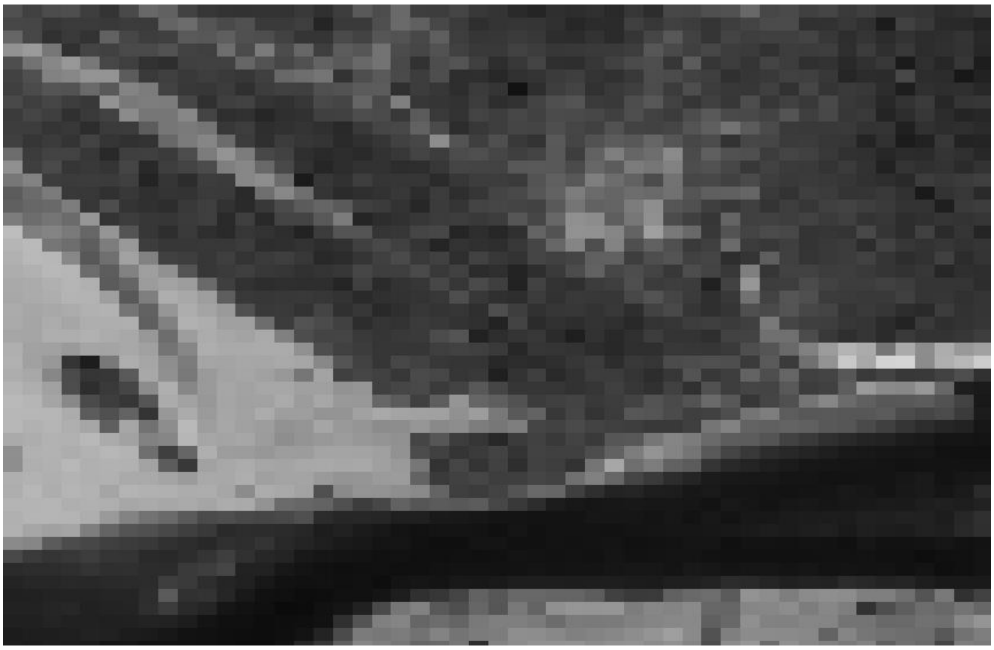
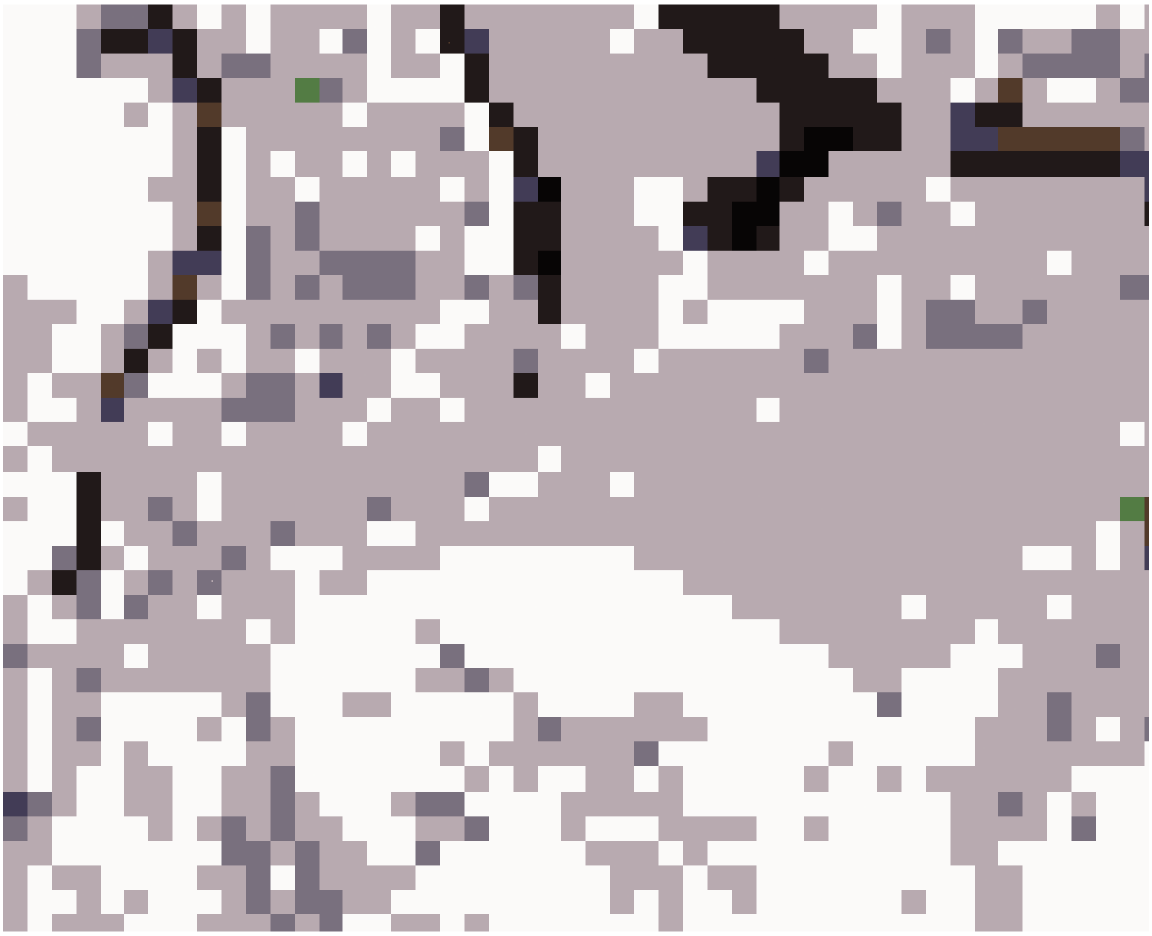


Figure 4. processing of Satellite Image



5. U-Net

The role played by U-Net in satellite image segmentation is significant as it has the capability to segment objects from complicated images, which is most often the scenario in satellite and remote sensing images. The role played by U-Net are:

1. **Semantic Segmentation**: Semantic segmentation is used mostly with U-Net, and every pixel of the image is labeled into pre-defined classes. For satellite images, these can be water bodies, vegetation, urban, forests, roads, etc.
2. **Encoder-Decoder Architecture**: The U-Net architecture is split into an encoder (contracting path) and a decoder (expanding path). The encoder is used to obtain features from the image, and the decoder is used to predict the segmentation mask. Skip connections between the encoder and decoder preserve spatial information, which is essential for fine-grained segmentation
3. **High Accuracy on Small Objects**: As satellite images have small and large objects (e.g., roads, buildings, or patches of vegetation), the U-Net architecture can even segment the small objects with high accuracy.
4. **Data Augmentation**: U-Net is usually trained with a limited number of data augmentation methods (i.e., rotation, scaling, flip) to make the model robust, particularly when dealing with satellite images that may come in different orientations, scales, or statuses.
5. **Sparse Data Training**: U-Net is able to function quite well even with a relatively small dataset compared to the conventional deep models. This aspect is very effective in remote sensing where the labeling is very difficult and we are left with small datasets.

5.1 Equations used

1. Convolution Operation (2D convolution):
   $Y(i,j) = m\sum n\sum X(i+m,j+n) \times W(m,n)$

2. ReLU Activation Function:
   $ReLU(x) = \max(0,x)$
3. Sigmoid Activation Function:
   $\sigma(x) = 1 + e - x1$
4. Max Pooling Operation:
   $Y(i,j) = m,nmaxX(i+m,j+n)$
6. Transposed Convolution (Upsampling):
   $Y(i,j) = m\sum n\sum X(i-m,j-n) \times W(m,n)$
7. Binary Cross-Entropy Loss (for binary segmentation):
   $L = -(y\log(p) + (1-y)\log(1-p))$
8. Categorical Cross-Entropy Loss (for multi-class segmentation):
   $L = -i = 1\sum^{C} yi\log(pi)$
9. Dice Coefficient (Evaluation Metric):
   $Dice(A,B) =\mid A \mid + \mid B \mid 2 \times \mid A \cap B \mid$
10. Batch Normalization:
   $X^{=}\sigma B2 + \epsilon X - \mu B$
11. Skip Connection (Conceptual Transfer):
   $SkipOutput = Fencoder(X) andInputtoDecoder = SkipOutput + UpsampledFeatures$

6. Result and Discussion

6.1 Result

In this project, we trained a U-Net model for satellite image segmentation in to a number of land cover classes like roads, buildings, barren land, and vegetation.

We trained the model for 100 epochs with the input image resolution of 256 × 256 pixels. Following are the observed key values:

Final Training Accuracy: 98.12%

Final Training Loss: 0.0449

Training Mean IoU Score: 0.6283

Figure 5. result and Accuracy

```
Epoch 100/100
10/10 [==============================] - 3s 297ms/step - loss: 0.0449 - accuracy: 0.9812 - mea
n_io_u_1: 0.6283 - val_loss: 0.9313 - val_accuracy: 0.8695 - val_mean_io_u_1: 0.5438
```

6.1.1. On the validation set:

Validation Accuracy: 86.95%

Validation Loss: 0.9313

Validation Mean IoU: 0.5438

6.1.2. On the test set:

Test Accuracy: 86.95%

Test Loss: 0.9313

Test IoU Score: 0.8695

Figure 6. result and Accuracy

```
3/3 [==============================] - 0s 83ms/step - loss: 0.9313 - accuracy: 0.8695 - mean_i
o_u_1: 0.5438
Test Loss: 0.9313451051712036
IoU Score: 0.8694694638252258
```

6.1.3. Image Masking Result

The image masking outcomes prove that the U-Net model learns to recognize and segment significant features in satellite images effectively. It maintains large structural features like highways and urban blocks with good accuracy. Small deviations in fine structures indicate that improvement can be further enhanced by:

1. Deepening the model
2. Utilizing higher resolution data,
3. Employing post-processing methods like morphological operations.

In general, the model offers robust segmentation masks that can be applied to real-world applications like urban planning, road extraction, and land-use classification

Figure 7. Image Masking Result

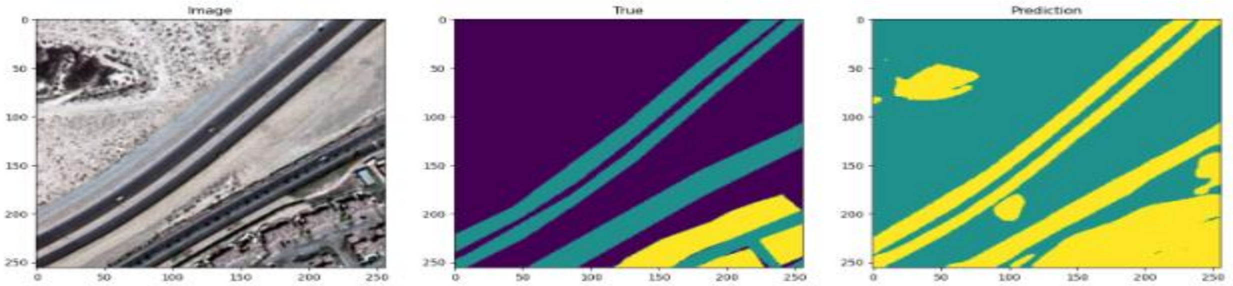
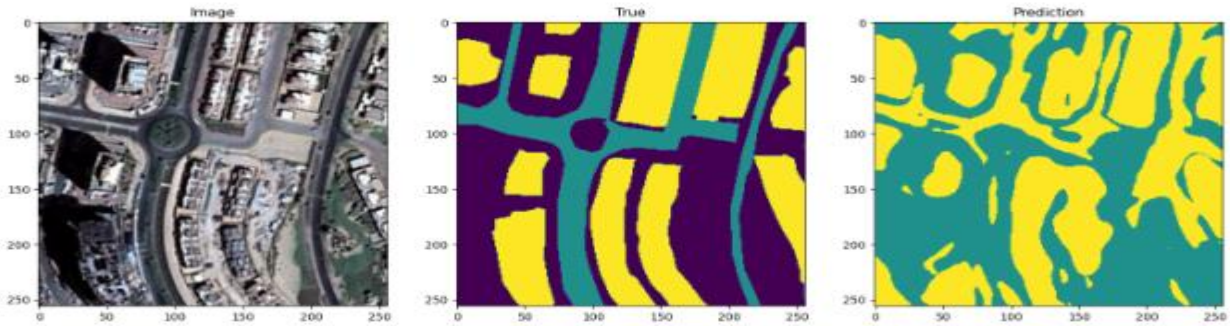


Figure 8. Image Masking Result



6.2. Discussion

Although the U-Net architecture worked reasonably well for satellite imagery segmentation, there are several locations where improvement is possible to make a huge leap in performance and reliability:

6.2.1 Data Augmentation and Diversity:

Increasing training data quantity and diversity through rotation, flipping, scaling, and contrast adjustments will help the model generalize better to new satellite images.Obtaining satellite images for various seasons, daytime, and weather conditions will also make the model robust.

6.2.2 Employing Advanced Architectures:

Recent implementations of U-Net, e.g., U-Net++, Attention U-Net, or DeepLabV3+, incorporate attention mechanisms, enhanced skip connections, and multi-scale feature extraction. Such structures, when implemented, can improve fine-grained segmentation, particularly in the case of intricate urban scenes.

6.2.3 Increased Resolution Input

Training with higher-resolution satellite images enables the model to capture more finer details such as small buildings, narrow roads, or small groups of trees. It would greatly enhance the model's performance in dealing with complex buildings.

6.2.4 Adding Multi-Spectral Data:

Using not only RGB (Red-Green-Blue) bands but also other bands like near-infrared (NIR) or thermal can provide the model with more data. Multi-spectral data can assist in separating classes that appear identical in visible light.

6.2.5 Post-Processing Techniques:

Employing post-processing techniques such as Conditional Random Fields (CRF) or Morphological Operations can improve segmentation boundaries, eliminate noise, and smooth class transitions in the predicted masks.

7. Acknowledgement

I am really thankful to my project guide Dr. K.P.Jayant sir, for the consistent supervision, valuable suggestions, and timely support throughout the project.

I would also be grateful to receive sponsorship from the Department of Computer Science and Engineering, Raj Kumar Goel Intitute of Technology, which supplied me with necessary facilities, tools, and learning-friendly environment needed in order to execute this research work successfully.

Their professional knowledge and working hard helped contribute significantly towards the overcoming of various issues and fruitful completion of this project.

Special thanks to my friends and fellow students for their ever-supportive attitude, technical discussions, and help whenever needed

This project has been a wonderful learning experience, and I have no doubt that the knowledge and skills obtained will greatly contribute to my future endeavors.

8. Conclusion

In this project, we effectively implemented and tested a U-Net based model for satellite image segmentation. The results show that U-Net is extremely efficient in identifying intricate patterns and structures in satellite images, yielding high-quality segmentation masks with good accuracy. Our model reached a validation accuracy of 86.95% and a mean Intersection over Union (IoU) score of 54.38%, showing good performance in discriminating among various land cover classes. Visual outputs also attested that the model could predict road networks, buildings, and open spaces with good accuracy, even though some minor misclassifications were seen where there was complicated texture. The research focuses on the capability of deep learning models, in particular encoder-decoder frameworks like U-Net, to address practical real-world tasks like urban planning, environmental monitoring, and disaster response through automated analysis of satellite imagery.

Overall, the research study provides a robust basis for potential future improvement involving the utilization of advanced architectures, multi-spectral information, and complex post-processing techniques, in order to achieve even higher levels of segmentation quality in future work.

References

1. Malik, P., Chourasiya, A., Pandit, R., and Bharaskar, K., 2023. Satellite Image Segmentation Using Neural Networks: A Comprehensive Review. International Journal of Enhanced Research in Educational Development (IJERED), Medi-Caps University, Indore, ISSN: 2320-8708, Vol. 11, Issue 4, July-August, Impact Factor: 7.326, pp. 20.
2. Bhadoria, P, 2020. Image Segmentation Techniques for Remote Sensing Satellite Images. IOP Conference Series: Materials Science and Engineering, Vol. 993, 012050.
3. Chhor, G., Aramburu, C. B., and Bougdal-Lambert, I., Satellite Image Segmentation for Building Detection using U-Net. Computational and Mathematical Engineering, Mechanical Engineering, Aeronautics and Astronautics, Stanford University, pp.1-10.
4. Ibtissam, Z., Brahim El Khalil, C., and Lhoussaine, M., 2018. Extraction of Building in Satellite Image THR using Features Detection. International Journal of Computer Applications, Vol. 181, No. 10, August, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco, and Faculty of Science, Mohammed V University, Rabat, Morocco, pp. 23-30.
5. Yin, S., Zhang, Y., and Karim, S., Large Scale Remote Sensing Image Segmentation Based on Fuzzy

Region Competition and Gaussian Mixture Model. School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China, pp. 1-12.
6. Fawwaz, I., Zarlis, M., Suherman, and Rahmat, R. F., The Edge Detection Enhancement on Satellite Image Using Bilateral Filter. Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia, pp. 1-8.
7. Liu, X., Ding, Y., and Liu, C., 2025. MSIM: A Multiscale Iteration Method for Aerial Image and Satellite Image Registration. Remote Sensing, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, University of Chinese Academy of Sciences, and State Key Laboratory of Dynamic Optical Imaging and Measurement, China, Vol. 17, Article 1423, pp. 1-14.
8. Chaurasia, K., Nandy, R., Pawar, O., Singh, R. R., and Ahire, M., 2021. Semantic Segmentation of High-Resolution Satellite Images Using Deep Learning. Springer-Verlag GmbH Germany, part of Springer Nature, pp. 1-12.
9. Pal, R., Mukhopadhyay, S., Chakraborty, D., and Suganthan, P. N., 2022. Very High-Resolution Satellite Image Segmentation Using Variable-Length Multi-Objective Genetic Clustering for Multi-Class Change Detection. Journal of King Saud University - Computer and Information Sciences, Vol. 34, Issue 10, Part B, November, pp. 9964-9976.
10. Guérin, E., Oechslin, K., Wolf, C., and Martinez, B., 2021. Satellite Image Semantic Segmentation. CNRS LIRIS - INSA Lyon and Ubisoft, September, pp.1-?.
11. Pal, N., Ramkrishna, S., Patil, H., Choudhary, N., and Soman, R., 2024. TerraGrid: Harnessing Deep Learning Models for Satellite Image Segmentation. International Journal of Computer Applications, Foundation of Computer Science (FCS), New York, DOI:10.5120/ijca2024924147, pp.14-21.
12. Li, J., Cai, Y., Li, Q., Kou, M., and Zhang, T., 2024. A Review of Remote Sensing Image Segmentation by Deep Learning Methods. Remote Sensing, Taylor & Francis, Article: 2328827, pp.1-?.