International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-8

Deep Learning - Based Satellite Image Segmentation

B.Naresh[1], D. Swasthitha[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana, India.

Abstract* – Semantic segmentation of satellite imagery has become a fundamental task in remote sensing because it enables accurate identification of land-use patterns for applications such as urban planning, environmental monitoring, disaster management, and resource assessment. However, the complex spatial characteristics of satellite images, including varying resolutions, atmospheric interference, and class imbalance, make precise pixel-level classification a challenging problem. This study presents a modified U-Net-based deep learning framework for multi-class semantic segmentation of satellite images using the Dubai Satellite Imagery Dataset. The proposed model classifies six distinct land-use categories, namely buildings, roads, vegetation, water bodies, barren land, and miscellaneous regions. The preprocessing pipeline incorporates image patch generation, MinMax normalization, one-hot encoding, and extensive data augmentation to improve training efficiency and enhance model generalization. To address the imbalance among land-cover classes, the framework combines Dice Loss with Categorical Focal Loss, enabling improved segmentation performance for both majority and minority classes. Model optimization is performed using the Adam optimizer, while segmentation quality is evaluated through Accuracy, Intersection over Union (IoU), and Jaccard Index. Experimental results demonstrate that the proposed approach achieves reliable segmentation performance across diverse satellite scenes, making it suitable for practical remote sensing applications. The study further discusses strategies for handling spatial resolution challenges and improving segmentation robustness in complex urban environments.

***Keywords* – Deep Learning, Satellite Image Segmentation, Remote Sensing Semantic Segmentation, Convolutional Neural Networks (CNN), Image Processing, Geospatial Analysis.**

I. Introduction

Recent advancements in satellite imaging technologies have significantly increased the availability of high-resolution Earth observation data, creating new opportunities for large-scale monitoring of natural and urban environments. These satellite images provide valuable information for numerous applications, including urban development, environmental conservation, agricultural monitoring, disaster response, land-use planning, and resource management. Extracting meaningful information from such large volumes of remotely sensed imagery, however, remains a challenging task because of variations in terrain, illumination conditions, atmospheric disturbances, and the complex spatial distribution of land-cover classes. Consequently, intelligent image segmentation techniques have become essential for transforming raw satellite imagery into actionable geographic information.

Semantic segmentation aims to assign a meaningful class label to every pixel within an image, enabling precise identification of different land-use categories. Traditional image segmentation techniques, including thresholding, region growing, edge detection, and handcrafted feature extraction, often struggle to handle the complexity of high-resolution satellite imagery. These conventional methods generally lack the capability to model intricate spatial relationships and frequently produce unsatisfactory results when confronted with diverse environmental conditions, seasonal variations, occlusions, and highly imbalanced land-cover distributions. Such limitations have encouraged researchers to investigate deep learning approaches capable of automatically learning hierarchical spatial representations directly from image data.

Among modern deep learning architectures, Convolutional Neural Networks (CNNs) have demonstrated remarkable success in various computer vision tasks due to their ability to extract complex visual features automatically. In particular, the U-Net architecture, originally developed for biomedical image segmentation, has proven highly effective for pixel-level classification because of its encoder-decoder structure combined with skip connections that preserve both contextual information and fine spatial details. These characteristics make U-Net particularly suitable for satellite image segmentation, where accurate boundary localization and preservation of small geographical structures such as roads and water channels are essential.

This research presents a modified U-Net-based semantic segmentation framework for analyzing satellite imagery collected from Dubai, United Arab Emirates. The dataset contains six land-use categories: buildings, roads, vegetation, water bodies, barren land, and miscellaneous regions. To improve segmentation performance, the framework incorporates multiple preprocessing operations, including image patch generation, MinMax normalization, one-hot encoding, and extensive data augmentation. Furthermore, the combination of Dice Loss and Categorical Focal Loss effectively addresses the class imbalance commonly encountered in satellite datasets by improving learning for underrepresented classes.

The primary objective of this study is to develop an efficient deep learning framework capable of performing accurate multi-class semantic segmentation while maintaining computational efficiency and robust generalization across diverse satellite scenes. The experimental findings demonstrate that integrating a modified U-Net architecture with advanced preprocessing and hybrid loss functions significantly improves segmentation accuracy, making the proposed framework suitable for real-world applications involving urban planning, environmental monitoring, and remote sensing analysis.

II. LITERATURE SURVEY

Semantic segmentation of satellite imagery has evolved considerably with the advancement of machine learning and deep learning technologies. Researchers have proposed numerous approaches for improving pixel-level land-cover classification while addressing challenges associated with image complexity, varying spatial resolutions, atmospheric disturbances, and class imbalance. Earlier studies primarily relied on conventional image processing and machine learning techniques, whereas recent investigations increasingly employ deep neural networks capable of learning rich spatial representations directly from satellite images.

Early satellite image segmentation techniques mainly utilized thresholding, region growing, watershed segmentation, and manually engineered spectral features for identifying different land-cover regions. Although these approaches offered relatively low computational complexity, they often struggled to represent the highly heterogeneous characteristics of satellite imagery. Rule-based segmentation methods further depended on manually designed feature extraction procedures, limiting their adaptability to diverse geographical regions and high-resolution remote sensing datasets. As a result, their performance frequently deteriorated when applied to large-scale real-world satellite images.

Several machine learning techniques have subsequently been introduced to improve segmentation accuracy. Awad proposed an unsupervised segmentation framework based on Self-Organizing Maps (SOM), demonstrating the capability of neural clustering for organizing satellite image pixels according to feature similarity. Similarly, Nurwauziyah et al. compared supervised classification algorithms including Decision Trees, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) for satellite image classification, reporting superior performance for Support Vector Machines over conventional classifiers. While these machine learning techniques improved segmentation accuracy, they still relied heavily on handcrafted feature engineering and exhibited limited capability for learning complex hierarchical image representations.

The introduction of deep learning significantly transformed satellite image segmentation. Long et al. proposed Fully Convolutional Networks (FCNs), establishing one of the earliest end-to-end deep learning frameworks for semantic segmentation. Building upon this concept, Badrinarayanan et al. introduced SegNet, an encoder-decoder architecture utilizing pooling indices for improved feature reconstruction. Subsequently, Ronneberger et al. developed the U-Net architecture, whose skip connections effectively preserve fine spatial information while combining local and global contextual features. Because of its ability to perform accurate pixel-level segmentation with relatively limited annotated data, U-Net has become one of the most widely adopted architectures in remote sensing applications.

More recent studies have focused on enhancing U-Net through architectural improvements and attention mechanisms. Zhang et al. proposed a Deep Residual U-Net capable of effectively learning multi-scale spatial representations from high-resolution satellite images. Wang et al. investigated techniques for addressing class imbalance in urban land-cover mapping, significantly improving segmentation performance for underrepresented categories. Furthermore, Liu et al. and Chen et al. incorporated attention mechanisms and atrous convolution into segmentation networks to improve contextual feature extraction and boundary refinement. Additional architectures such as AD-LinkNet have combined attention mechanisms with dilated convolutions to further enhance road and building segmentation performance in complex urban environments.

Despite these advancements, challenges related to class imbalance, spatial variability, and computational efficiency continue to affect satellite image segmentation. Motivated by these limitations, the present study develops a modified U-Net framework incorporating patch-based preprocessing, MinMax normalization, one-hot encoding, data augmentation, and a hybrid loss function combining Dice Loss with Categorical Focal Loss. This integrated approach preserves the original methodology while improving segmentation robustness and pixel-level classification performance across multiple land-use categories within the Dubai Satellite Imagery Dataset.

III. PROPOSED SYSTEM

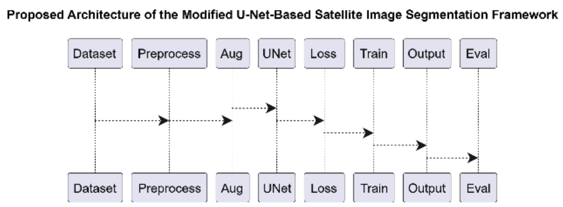

Fig. 1. Proposed Architecture of the Modified U-Net-Based Satellite Image Segmentation Framework

The proposed system presents a deep learning-based framework for multi-class semantic segmentation of satellite imagery using a modified U-Net architecture. The framework is specifically designed to accurately classify every pixel in a satellite image into one of six predefined land-use categories, namely Buildings, Roads, Vegetation, Water Bodies, Barren Land, and Miscellaneous. Unlike conventional image processing techniques that rely on handcrafted features, the proposed model automatically learns hierarchical spatial representations from satellite images through an encoder-decoder architecture. This enables the system to preserve both global contextual information and fine-grained spatial details, making it highly effective for remote sensing applications such as urban planning, environmental monitoring, land-cover analysis, and disaster management.

The segmentation pipeline begins by acquiring high-resolution satellite images from the Dubai Satellite Imagery Dataset. Since the original satellite images are considerably large, they are divided into smaller image patches to facilitate efficient processing and reduce memory consumption during training. Corresponding segmentation masks are partitioned in the same manner to maintain pixel-level correspondence between input images and ground-truth labels. This patch-based strategy enables the U-Net model to learn localized spatial patterns while increasing the number of training samples available for model optimization.

Before training, several preprocessing operations are applied to improve data quality and model convergence. The RGB segmentation masks are converted into numerical class labels representing the six land-use categories. Image pixel values are normalized using MinMaxScaler, scaling the intensity values between 0 and 1 to provide consistent input distributions. Subsequently, the class labels are transformed using One-Hot Encoding, enabling compatibility with the multi-class Softmax output layer of the segmentation network. To improve model robustness and reduce overfitting, extensive data augmentation techniques including horizontal and vertical flipping, random rotation, scaling, brightness adjustment, contrast modification, and image translation are performed while preserving the alignment between images and their corresponding masks.

The modified U-Net architecture consists of four major components: the input layer, encoder, bottleneck, and decoder. The encoder progressively extracts hierarchical image features using multiple convolutional layers followed by Max-Pooling operations, enabling the network to learn increasingly complex visual representations. At the deepest level, the bottleneck captures high-level semantic information through convolutional layers with an increased number of feature maps while dropout layers minimize overfitting by randomly deactivating neurons during training. The decoder subsequently reconstructs the segmentation map through transposed convolution (upsampling) operations. Skip connections between corresponding encoder and decoder stages preserve fine spatial information that would otherwise be lost during pooling, resulting in accurate boundary reconstruction for roads, buildings, vegetation, and other land-cover classes.

To address the severe class imbalance commonly observed in satellite imagery, the proposed framework employs a hybrid loss function combining Dice Loss and Categorical Focal Loss. Dice Loss maximizes the overlap between predicted segmentation masks and the corresponding ground-truth labels, whereas Categorical Focal Loss assigns greater importance to underrepresented classes by emphasizing difficult-to-classify samples. This hybrid optimization strategy improves segmentation accuracy across all land-use categories, including minority classes that typically occupy smaller regions within satellite images. Network optimization is performed using the Adam optimizer, which adaptively updates model parameters to achieve faster convergence during training. The model is trained for 50 epochs using a batch size of 8, preserving the original experimental configuration. Performance is evaluated using Accuracy, Intersection over Union (IoU), and the Jaccard Index, providing a comprehensive assessment of segmentation quality.

Overall, the proposed system integrates comprehensive preprocessing, extensive data augmentation, a modified U-Net architecture, and hybrid loss optimization into a unified semantic segmentation framework. The combination of these components enables accurate pixel-level classification across diverse satellite scenes while maintaining computational efficiency and strong generalization capability, making the framework suitable for practical remote sensing applications involving urban development, environmental monitoring, land-use analysis, and disaster response.

IV. DATASET

The proposed study utilizes the Dubai Satellite Imagery Dataset, obtained from Kaggle, to develop and evaluate the semantic segmentation framework. The dataset consists of high-resolution aerial satellite images captured by the Mohammed Bin Rashid Space Centre (MBRSC), representing both urban and natural environments within Dubai, United Arab Emirates. It contains 72 satellite images organized into six larger image tiles, each accompanied by corresponding pixel-level segmentation masks. These annotated masks assign every image pixel to one of six predefined land-use categories, namely Buildings, Roads, Vegetation, Water Bodies, Barren Land, and Miscellaneous, making the dataset highly suitable for supervised multi-class semantic segmentation tasks.

The dataset is organized into two primary components:

- Satellite Images: High-resolution aerial images stored in JPG format.
- Segmentation Masks: Corresponding pixel-wise annotated masks stored in PNG format.

Each segmentation mask provides detailed class annotations that enable the model to learn accurate pixel-level classification across multiple land-cover categories. The availability of both urban infrastructure and natural landscapes allows the proposed framework to evaluate segmentation performance under diverse environmental conditions while supporting practical applications such as urban planning, land-cover mapping, environmental monitoring, disaster management, and remote sensing analysis.

To ensure reliable model evaluation, the dataset is divided into separate training, validation, and testing subsets. The experimental configuration follows an 85:15 training-to-testing ratio, preserving the original research methodology. During preprocessing, large satellite images are partitioned into smaller image patches to facilitate efficient model training and reduce memory consumption. Corresponding segmentation masks are partitioned identically to preserve pixel-level alignment between input images and ground-truth annotations.

Although the dataset provides representative samples from all six land-use classes, the spatial distribution of these categories is naturally imbalanced because certain classes occupy significantly smaller image regions. To address this challenge, the proposed framework incorporates comprehensive preprocessing and data augmentation strategies that improve class representation and enhance model generalization. These preprocessing operations ensure that the dataset provides a robust foundation for training the modified U-Net architecture while maintaining accurate semantic segmentation across diverse satellite imagery.

V. DATA PREPROCESSING

Effective preprocessing is an essential stage in semantic satellite image segmentation because the quality of the input data directly influences the learning capability and prediction accuracy of deep learning models. Satellite images generally contain variations in spatial resolution, illumination, and object distribution that may negatively affect model convergence if appropriate preprocessing is not performed. Therefore, the proposed framework applies a sequence of preprocessing operations designed to standardize the dataset, improve computational efficiency, and generate high-quality training samples for the modified U-Net architecture.

Initially, the large satellite images contained in the Dubai Satellite Imagery Dataset are divided into smaller fixed-size image patches. This patch generation process enables efficient utilization of computational resources while reducing memory requirements during training. Smaller image patches also allow the segmentation network to learn localized spatial features more effectively without losing important structural information present within the original high-resolution satellite images. Corresponding segmentation masks are partitioned using the same strategy to preserve pixel-level alignment between input images and their associated ground-truth annotations.

Following patch generation, the segmentation masks undergo RGB-to-class label conversion. Each pixel within the RGB mask is mapped to a unique numerical class index corresponding to one of the six predefined land-use categories: Buildings, Roads, Vegetation, Water Bodies, Barren Land, and Miscellaneous. This transformation converts the colored segmentation masks into a numerical representation suitable for supervised semantic segmentation. Accurate class mapping ensures that every training image maintains precise pixel-wise correspondence with its ground-truth segmentation mask.

To improve numerical stability during model optimization, all satellite image pixel values are normalized using MinMaxScaler, which scales the pixel intensities to the range 0–1. Normalization minimizes differences in feature magnitude, accelerates model convergence, and enables more stable gradient updates during training. Subsequently, the converted class labels are transformed using One-Hot Encoding, allowing the segmentation masks to become fully compatible with the Softmax output layer employed by the modified U-Net architecture. This encoding strategy enables efficient multi-class pixel classification across all land-use categories.

For reliable performance evaluation, the complete dataset is divided into separate training, validation, and testing subsets while maintaining the original 85:15 training-to-testing ratio. This partitioning strategy ensures that the trained model is evaluated using previously unseen satellite images, thereby providing an unbiased estimate of segmentation performance. Prior to training, representative image-mask pairs are visually inspected to verify correct preprocessing, accurate class assignments, and proper alignment between satellite images and their corresponding segmentation masks.

Overall, the preprocessing pipeline integrates image patch generation, RGB-to-class conversion, MinMax normalization, One-Hot Encoding, dataset partitioning, and preprocessing validation into a unified workflow. These operations provide standardized, high-quality input data that enable the modified U-Net architecture to perform accurate and efficient semantic segmentation across complex satellite imagery.

VI. DATA AUGMENTATION

Data augmentation plays a significant role in improving the robustness and generalization capability of deep learning models, particularly when training datasets contain limited samples or exhibit class imbalance. Satellite imagery often contains considerable variations in orientation, scale, illumination, and object positioning, making augmentation essential for exposing the segmentation network to diverse environmental conditions. The proposed framework therefore incorporates multiple augmentation techniques that artificially increase dataset diversity while preserving the semantic consistency between satellite images and their corresponding segmentation masks.

The first augmentation technique involves horizontal and vertical image flipping, enabling the model to recognize land-cover patterns regardless of their orientation. Since roads, buildings, vegetation, and water bodies may appear in multiple directions within aerial imagery, flipping operations improve rotational invariance and enhance the model's ability to generalize across diverse satellite scenes. The framework additionally applies random image rotation of up to ±90 degrees, allowing the segmentation model to learn orientation-independent spatial representations. Because satellite images may be captured from different viewing angles, rotational augmentation enables the U-Net architecture to recognize identical land-cover structures despite changes in image orientation. Similarly, uniform image scaling is performed to simulate variations in spatial resolution, allowing the model to adapt effectively to satellite images acquired at different scales.

To improve robustness under varying environmental conditions, brightness and contrast adjustments are introduced during training. These transformations simulate changes in illumination caused by weather conditions, seasonal differences, atmospheric effects, and sensor characteristics. Consequently, the segmentation network becomes less sensitive to lighting variations while maintaining accurate pixel-level classification across different satellite acquisition conditions.

The augmentation pipeline further incorporates random image translation, where satellite images are shifted horizontally and vertically without altering their semantic content. Translation augmentation improves positional invariance by enabling the network to recognize identical land-cover objects appearing at different image locations. Throughout all augmentation operations, careful attention is given to preserving exact alignment between augmented satellite images and their corresponding segmentation masks, thereby maintaining accurate pixel-wise annotations required for supervised learning.

The augmented samples are generated dynamically during model training, substantially increasing the diversity of training data without modifying the original dataset. This strategy effectively reduces overfitting, particularly for minority land-cover categories that occupy relatively small image regions. According to the experimental observations, data augmentation contributes to a measurable improvement in segmentation performance, increasing Accuracy from 83.3% to 86.4% while improving Intersection over Union (IoU) from 70.2% to 76.9%. Furthermore, evaluation on previously unseen satellite images demonstrates consistent segmentation performance under diverse augmented conditions, confirming the effectiveness of the proposed augmentation strategy in enhancing model generalization and robustness.

VII. EXPERIMENTAL RESULTS

The proposed semantic segmentation framework was experimentally evaluated to assess its capability in accurately classifying satellite imagery into six predefined land-use categories. The modified U-Net architecture was trained using the Adam optimizer for 50 epochs with a batch size of 8, while a hybrid loss function combining Dice Loss and Categorical Focal Loss was employed to address class imbalance and improve segmentation quality. Throughout the training process, model performance was continuously monitored using Accuracy, Intersection over Union (IoU), and the Jaccard Index, providing a comprehensive assessment of segmentation effectiveness.

The experimental results demonstrate that the proposed hybrid loss strategy significantly enhances segmentation performance by simultaneously improving overlap between predicted masks and ground-truth annotations while increasing the contribution of minority land-cover classes. During model optimization, the network exhibited stable convergence across the training epochs, with continuous improvements observed in both segmentation accuracy and IoU values. The combined loss function effectively minimized misclassification errors for underrepresented categories such as roads and water bodies while maintaining high prediction accuracy for dominant land-cover classes.

Performance evaluation indicates that the proposed framework achieved a substantial improvement in segmentation quality. Following the application of comprehensive preprocessing and data augmentation techniques, the overall segmentation accuracy increased from 84.5% to 86.7%, while the Intersection over Union (IoU) improved from 70.9% to 76.3% during training. These improvements demonstrate the effectiveness of integrating hybrid loss functions with the modified U-Net architecture for handling complex satellite imagery characterized by varying object sizes, spatial distributions, and class imbalance.

The learning behavior of the segmentation model was further analyzed using training and validation performance curves. The graphical evaluation of training loss, validation loss, Accuracy, and IoU demonstrates smooth convergence throughout the optimization process without significant fluctuations, indicating stable learning and effective generalization. During final evaluation, the trained model was tested on 142 previously unseen satellite image samples, where the predicted segmentation masks closely matched the corresponding expert-annotated ground-truth masks. Visual comparisons between the original satellite images, annotated segmentation masks, and predicted outputs confirm the model's capability to accurately identify buildings, roads, vegetation, water bodies, barren land, and miscellaneous regions under diverse environmental conditions.

The incorporation of extensive data augmentation, including flipping, rotation, scaling, brightness adjustment, and image translation, further improved model robustness by exposing the network to a wide variety of spatial transformations during training. As a result, the proposed segmentation framework demonstrated strong generalization capability across previously unseen satellite scenes while maintaining consistent pixel-level classification performance. Overall, the experimental findings validate the effectiveness of the modified U-Net architecture combined with hybrid loss optimization as a scalable and reliable solution for semantic satellite image segmentation in remote sensing, environmental monitoring, urban planning, and disaster management applications.

VIII. DISCUSSION

The experimental results demonstrate that the proposed modified U-Net framework effectively performs multi-class semantic segmentation of satellite imagery by accurately classifying six land-use categories. The integration of comprehensive preprocessing techniques, patch-based image generation, data augmentation, and hybrid loss optimization enables the network to learn meaningful spatial representations while preserving fine structural details within high-resolution satellite images. Compared with conventional image segmentation techniques that rely on handcrafted features, the proposed deep learning framework automatically extracts hierarchical image features, resulting in improved segmentation accuracy and enhanced robustness across diverse urban and natural environments.

One of the primary strengths of the proposed framework lies in its ability to effectively address the class imbalance commonly encountered in satellite imagery. Certain land-cover categories, such as roads and water bodies, occupy relatively small portions of satellite images, making them more difficult to segment accurately using conventional loss functions. By combining Dice Loss with Categorical Focal Loss, the proposed model assigns greater importance to difficult and underrepresented classes while simultaneously maximizing the overlap between predicted segmentation masks and ground-truth annotations. This hybrid optimization strategy contributes significantly to the improved segmentation performance observed during experimental evaluation.

The incorporation of extensive data augmentation further enhances model generalization by exposing the segmentation network to diverse spatial transformations, including image flipping, rotation, scaling, brightness adjustment, contrast modification, and translation. These augmentation techniques reduce the risk of overfitting while enabling the model to adapt to variations in satellite image orientation, illumination, and environmental conditions. Consequently, the trained model maintains stable segmentation performance when evaluated on previously unseen satellite imagery, demonstrating its suitability for practical remote sensing applications.

Visual analysis of the predicted segmentation masks indicates that the modified U-Net architecture successfully preserves object boundaries while accurately identifying complex land-cover structures. The encoder-decoder architecture, together with skip connections, enables simultaneous learning of global contextual information and fine spatial details. This capability is particularly important for accurately segmenting narrow structures such as roads and irregular boundaries associated with vegetation and water bodies. Furthermore, the experimental improvements in Accuracy, Intersection over Union (IoU), and Jaccard Index confirm the effectiveness of integrating advanced preprocessing with the proposed segmentation architecture. Despite the encouraging performance, certain challenges remain. Satellite imagery acquired under varying atmospheric conditions, seasonal changes, and extreme illumination differences may still influence segmentation accuracy. In addition, processing ultra-high-resolution satellite images requires considerable computational resources and memory during both training and inference. Future optimization strategies should therefore focus on improving computational efficiency while maintaining segmentation quality across larger and more diverse satellite datasets.

Overall, the proposed framework demonstrates that combining a modified U-Net architecture, patch-based preprocessing, data augmentation, and hybrid Dice–Categorical Focal Loss optimization provides an effective and scalable solution for semantic satellite image segmentation. The developed model offers significant potential for practical deployment in urban planning, environmental monitoring, precision agriculture, land-use analysis, disaster response, and smart city development, where accurate pixel-level classification of satellite imagery plays a critical role.

IX. CONCLUSION

This study presents a modified U-Net-based deep learning framework for multi-class semantic segmentation of satellite imagery using the Dubai Satellite Imagery Dataset. The proposed approach combines comprehensive preprocessing, patch-based image generation, MinMax normalization, One-Hot Encoding, extensive data augmentation, and a hybrid loss function consisting of Dice Loss and Categorical Focal Loss to improve segmentation performance across six land-use categories: Buildings, Roads, Vegetation, Water Bodies, Barren Land, and Miscellaneous. By integrating these components into a unified segmentation pipeline, the framework effectively addresses challenges associated with complex spatial structures and class imbalance commonly encountered in high-resolution satellite imagery.

Experimental evaluation demonstrates that the proposed model achieves reliable semantic segmentation performance while maintaining stable convergence during training. The use of the Adam optimizer, 50 training epochs, and a batch size of 8 enables efficient optimization of the modified U-Net architecture. Furthermore, the incorporation of hybrid loss functions significantly improves segmentation quality by enhancing pixel-level classification for both dominant and minority land-cover classes. Improvements observed in Accuracy, Intersection over Union (IoU), and the Jaccard Index validate the effectiveness of the proposed framework for remote sensing applications.

The experimental findings further indicate that extensive data augmentation substantially enhances model generalization, enabling the segmentation network to perform consistently across diverse satellite scenes characterized by varying object orientations, scales, and illumination conditions. The encoder-decoder architecture with skip connections successfully preserves fine spatial information while accurately reconstructing segmentation masks, making the proposed framework particularly suitable for applications requiring precise land-cover mapping.

Future research may focus on extending the proposed framework through the integration of advanced attention mechanisms, transformer-based segmentation networks, and larger multi-regional satellite datasets to improve segmentation accuracy under more challenging environmental conditions. Additionally, optimizing computational efficiency for real-time processing of ultra-high-resolution satellite imagery would further expand the practical applicability of the proposed system. Overall, the modified U-Net framework provides a robust, scalable, and efficient solution for semantic satellite image segmentation, supporting a wide range of applications in urban planning, environmental monitoring, disaster management, agricultural analysis, and intelligent geospatial information systems.

REFERENCES

1. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 2015, pp. 234–241.
2. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431–3440.
3. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
4. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, Apr. 2018.
5. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1800–1807.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.
7. O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.
8. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2881–2890.
9. F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in Proc. Int. Conf. Learning Representations (ICLR), 2016.
10. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learning Representations (ICLR), 2015.
11. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 2980–2988.
12. C. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 240–248.
13. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
14. G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," Proceedings of the IEEE, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
15. X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8–36, Dec. 2017.
16. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105.
17. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
18. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
19. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
20. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

21. R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed. Cham, Switzerland: Springer, 2022.
22. J. A. Richards, Remote Sensing Digital Image Analysis, 5th ed. Berlin, Germany: Springer, 2013.
23. T. Blaschke, "Object based image analysis for remote sensing," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, no. 1, pp. 2–16, Jan. 2010.
24. M. A. Wulder, J. C. White, T. R. Loveland, C. E. Woodcock, A. S. Belward, W. B. Cohen, E. F. Vermote, J. G. Masek, D. P. Roy, and C. O. Justice, "The global Landsat archive: Status, consolidation, and direction," Remote Sensing of Environment, vol. 185, pp. 271–283, Nov. 2016.
25. Y. Liu, L. Ding, C. Chen, and H. Gao, "Deep learning for high-resolution remote sensing image segmentation: A survey," Remote Sensing, vol. 15, no. 2, Art. no. 412, 2023.