www.ijarp.com ISSN 2456-9992 Page: 01-07

International Journal Advanced Research Publications

Review Article

Volume: 01

Issue: 03

UNCERTAINTY CHARACTERIZATION IN DEEP LEARNING MODELS FOR URBAN OBJECT DETECTION FROM SATELLITE IMAGERY

Reedhi Shukla*, Sameth Kumar P, Suresh babu, Kamini J.

National Remote Sensing Centre (NRSC), Indian Space Research Organisation (ISRO), Hyderabad, India.

Article Received: 16 November 2025, Article Revised: 06 December 2025, Published on: 26 December 2025

***Corresponding Author: Reedhi Shukla**

National Remote Sensing Centre (NRSC), Indian Space Research Organisation (ISRO), Hyderabad, India.

DOI: https://doi-doi.org/101555/ijarp.7127

ABSTRACT

Satellite image segmentation plays a vital role in various applications such as urban planning, infrastructure monitoring, and disaster management. However, the performance of deep learning architectures like U-Net often deteriorates due to environmental noise, sensor-related distortions, and the inherent complexity of urban terrain. To address these challenges, this study integrates attention mechanisms into the conventional U-Net framework, developing an Attention U-Net model designed to enhance feature learning and spatial focus. The robustness of the model is further evaluated through uncertainty characterization in the context of building footprint extraction. Experimental results demonstrate that the proposed Attention U-Net consistently outperforms the baseline U-Net across both training and validation datasets, achieving an accuracy of 81.15% on the validation set, compared to 79.75% achieved by the standard U-Net. The inclusion of the attention mechanism enables the network to selectively emphasize relevant spatial regions, resulting in improved segmentation precision, enhanced visual delineation of buildings, and reduced predictive uncertainty, particularly in densely built and heterogeneous urban environments. Overall, the findings highlight that Attention U-Net represents a more robust and reliable solution for complex satellite image segmentation tasks where high accuracy and interpretability are essential.

INTRODUCTION

Satellite imagery plays an important role for the scientific study of diverse applications such as environmental monitoring, urban planning, disaster management, and agricultural assessment (Li et al., 2023; Zhang et al., 2022). It serves as a scientific tool for accurate land use classification, feature extraction, and change detection. It also enables informed decision-making, particularly in the face of rapid urbanization, climate change, and growing resource demands.

For instance, in urban planning, extracting building footprints from satellite images supports monitoring urban sprawl, infrastructure development, and environmental impact assessment (Zhu et al., 2017). Similarly, in disaster management, rapid segmentation of affected areas facilitates timely resource allocation, improving emergency response and potentially saving lives (Sharma & Ghosh, 2021). Environmental monitoring also benefits significantly from satellite-based feature mapping, particularly in tracking deforestation, water dynamics, and other ecological indicators (Roy et al., 2021).Thse monitoring are feasible only due to availability of very high-resolution resolution data on timely bases with bigger coverfage. Though the availability of satellite data has expanded opportunities for Earth better scientific monitoring approaches but it has also introduced significant challenges in data processing and analysis. Manual interpretation is time-consuming, making it unsuitable for large-scale applications. To address this, recent developments in deep learning (DL) have revolutionized remote sensing through automated, accurate, and scalable image interpretation techniques.

Architectures such as **Convolutional Neural Networks (CNNs)** and **U-Net** have shown impressive performance in semantic segmentation tasks, which are essential for extracting spatial features like roads, vegetation, and buildings from high-resolution satellite imagery (Ronneberger et al., 2015; Ma et al., 2019) but, the accurate assessment of model uncertainty remains an underexplored area, particularly in the context of object detection from VHRS imagery. This gap highlights the necessity of transferring and adapting uncertainty modeling techniques to the remote sensing domain, where environmental complexity, data heterogeneity, and sensor variability introduce additional challenges.

The central idea of this study is that a model may occasionally produce incorrect predictions, but a well-calibrated model should be able to **recognize when it is likely to be wrong**. Developing such models can lead to more robust, interpretable, and trustworthy AI systems for geospatial analysis. In order to deploy deep learning models confidently in real-world remote sensing applications, it is essential to obtain better estimates of model confidence and predictive uncertainty. A comprehensive understanding of these uncertainties can substantially improve both the reliability and performance of DL models, ultimately advancing the field of automated remote sensing interpretation.

To overcome these limitations, the present study focuses on enhancing deep learning-based satellite image segmentation through the integration of **attention mechanisms** within the U-Net architecture, forming an **Attention U-Net**. The attention module enables the network to selectively focus on the most relevant spatial features while suppressing background noise, improving performance in challenging and noisy environments (Oktay et al., 2018). Furthermore, the research emphasizes **uncertainty characterization**, an essential yet often overlooked aspect of deep learning models in remote sensing. Uncertainty quantification helps identify the confidence level of predictions, thereby improving the interpretability and reliability of model outputs (Kendall & Gal, 2017). Understanding these uncertainties is crucial for real-world applications, where model predictions directly influence high-stakes decisions in urban management and disaster mitigation.

The main objectives of this study are threefold. First, to enhance the baseline U-Net model by integrating attention mechanisms that enable the network to focus more effectively on critical regions within satellite imagery. Second, to optimize the performance of the Attention U-Net through hyperparameter tuning and robust training strategies to ensure better generalization and resistance to overfitting. Third, to quantify and analyze uncertainty in model predictions, assessing how environmental, architectural, and data-induced factors influence segmentation reliability. These objectives collectively aim to improve not only the accuracy but also the **trustworthiness** of satellite-based deep learning systems.

The **scope** of this study encompasses the development and evaluation of a deep learning framework for **building footprint extraction** from very high-resolution (VHR) satellite imagery. It includes data preprocessing, model development, hyperparameter optimization, and uncertainty quantification. The performance of the proposed Attention U-Net is compared with existing baseline models using standard segmentation metrics such as Intersection over Union (IoU), Precision, Recall, and F1-score, along with uncertainty measures to assess predictive confidence and robustness.

The **significance** of this research lies in its contribution to both practical applications and methodological advancements in geospatial analysis. From a practical perspective, the improved segmentation accuracy and uncertainty awareness of the proposed model will benefit urban planners, disaster management authorities, and environmental agencies by providing more reliable insights for decision-making. From a methodological perspective, this study contributes to the growing body of research that integrates explainability and uncertainty quantification into deep learning workflows, thereby advancing the interpretability and reliability of artificial intelligence (AI) models in remote sensing (Abdar et al., 2021). As the well-known maxim suggests, "All models are wrong, but models that know when they are wrong are useful" (Box, 1976). By quantifying uncertainty, this work aims to create models that are not only accurate but also self-aware—capable of signaling when their predictions may be unreliable.

This study also explores the **types and sources of uncertainty** that affect deep learning models in satellite image segmentation. These can be broadly classified as **aleatoric uncertainty**, arising from inherent noise or randomness in the input data; **epistemic uncertainty**, stemming from limited data or model knowledge; **model uncertainty**, resulting from architectural or parameter variability; and **data uncertainty**, related to inconsistencies or errors in training and input datasets (Kiureghian & Ditlevsen, 2009; Gal, 2016). In the context of urban object detection from very high-resolution satellite data, these uncertainties may originate from environmental factors, data quality, and model configuration. Understanding and mitigating these uncertainties is vital to improving model reliability and ensuring that the model's predictions can be confidently utilized in operational scenarios.

In summary, this research aims to develop an enhanced Attention U-Net model that not only improves segmentation accuracy for building footprint extraction but also characterizes and quantifies uncertainty to enhance interpretability and reliability. By integrating attention mechanisms with associated uncertainty in prediction this study contributes to building more transparent, dependable, and practically applicable deep learning systems for satellite image analysis. The outcomes of this work are expected to bridge the gap between theoretical advancements in deep learning and their operational deployment for better urban planning, efficient disaster response, and improved environmental monitoring.

METHODOLOGY

The system architecture for this project uncertainty associated with satellite image segmentation tasks, with a primary focus on building footprint extraction. The includes several component like data preprocessing, model architecture, training and evaluation and uncertainty estimate as shown in Figure 1.

1. Dataset Description

The dataset used in this study comprises **Very High-Resolution Satellite (VHRS)** imagery from the Hyderabad region. The imagery was sourced from **SAS Planet** and the **Open Buildings Dataset**, consisting of **TIFF files** that include both RGB images and corresponding binary masks representing building footprints. This dataset provides the spatial and spectral detail necessary for accurate extraction of urban structures, particularly building footprints.

2. Data Preprocessing

The preprocessing pipeline involves several key steps to ensure consistency and model readiness:

- **Chipping:** Large VHRS images are divided into **512×512 pixel tiles** to handle memory efficiently and capture fine-grained spatial features.
- **Normalization:** Pixel values are rescaled to the **[0, 1]** range using normalization functions to stabilize the training process.
- **Data Augmentation:** Random **rotations, flips, scalings, and translations** are applied to increase variability in the training set, improving the model's robustness and generalization.
- **Sorting and Labeling:** Image–mask pairs are systematically organized to prevent mismatches during training and validation.

These steps enhance both the reliability and performance of the model during the segmentation task.

3. Model Training Configuration

The proposed segmentation model is based on the **Attention U-Net** architecture, an improved version of the traditional U-Net. The attention gates embedded in the network enhance focus on relevant image regions, thereby improving segmentation precision in complex urban areas. The configuration adopted for training includes:

- **Loss Function:** *Categorical Cross-Entropy*, chosen for its effectiveness in pixel-wise binary segmentation.

- **Optimizer:** *Adam Optimizer* with a **learning rate of 0.001**, ensuring adaptive gradient updates for stable convergence.
- **Epochs:** The network is trained for **50 epochs**, balancing computational cost and convergence.
- **Batch Size:** A **batch size of 4** is used to maintain training stability given the large input dimensions.
- **Training–Validation Split:** The dataset is divided into **80% training** and **20% validation**, totaling approximately **10,000 samples**.
- **Callbacks:** *Early stopping* and *model checkpointing* are implemented to prevent overfitting and retain the best-performing model weights.

This configuration ensures that the network learns efficiently while preserving generalization across varied urban landscapes.

4. Uncertainty Estimation

To evaluate the reliability of predictions, **uncertainty quantification** is performed using **Monte Carlo Dropout (MC-Dropout)** during inference. Multiple stochastic forward passes through the model yield slightly varied segmentation outputs. The **mean** and **standard deviation (uncertainty map)** of these predictions provide insight into confidence levels.

Regions with high variance indicate greater model uncertainty — often corresponding to irregular terrain or occluded structures. The analysis demonstrates that the **type of terrain** is a critical determinant of segmentation accuracy and uncertainty. Dense urban environments, irregular elevation, and vegetative occlusion lead to higher uncertainty, while well-structured, flat regions yield the most confident predictions.

REFERENCES

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, *76*, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008
2. Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, 20–32. https://doi.org/10.1016/j.isprsjprs.2017.11.011

3. Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949
4. Gal, Y. (2016). *Uncertainty in deep learning* (Doctoral dissertation, University of Cambridge).
5. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* (pp. 5574–5584).
6. Kiureghian, A. D., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety, 31*(2), 105–112. https://doi.org/10.1016/j.strusafe.2008.06.020
7. Li, W., Li, Z., & Shen, H. (2023). Deep learning for high-resolution remote sensing image analysis: A comprehensive review. *IEEE Geoscience and Remote Sensing Magazine, 11*(2), 8–36. https://doi.org/10.1109/MGRS.2023.3247875
8. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in ote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing, 152*, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.
9. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999.*
10. Roy, D. P., Li, J., Zhang, H. K., & Huang, H. (2021). A general method to normalize Landsat reflectance data to nadir BRDF-adjusted reflectance (NBAR). *Remote Sensing of Environment, 252*, 112129. https://doi.org/10.1016/j.rse.2020.112129.
11. Sharma, A., & Ghosh, S. (2021). Deep learning-based building footprint extraction from high-resolution satellite imagery. *Remote Sensing Applications: Society and Environment, 22*, 100518. https://doi.org/10.1016/j.rsase.2021.100518.
12. Zhang, Y., Zhu, Q., & Liu, Y. (2022). A survey on deep learning applications in remote sensing image analysis. *ISPRS International Journal of Geo-Information, 11*(4), 206. https://doi.org/10.3390/ijgi11040206.
13. Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). learning in remote sensing: A comprehensive review and list of resources. *IEEE oscience and Remote Sensing Magazine, 5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307.