AI Models for Automatic Objects Classification in Satellite Images

Victoria Vysotska[1], Kirill Smelyakov[2], Serhii Osiievskyi[3] and Volodymyr Yartsev[2]

[1] Lviv Polytechnic National University, Stepan Bandera Street, 12, Lviv, 79013, Ukraine
[2] Kharkiv National University of Radio Electronics, 14 Nauky Ave., Kharkiv, 61166, Ukraine
[3] Kharkiv National University of Air Force, 77/79 Sumska St., Kharkiv, 61023, Ukraine

Abstract

This study investigates the application of artificial intelligence techniques for object segmentation in high-resolution satellite imagery, with a focus on the automatic classification of land cover types such as rivers, forests, and buildings. It includes a comparative analysis of traditional image processing methods and modern deep learning architectures — specifically convolutional neural networks (U-Net, DeepLabV3+, Mask R-CNN) and transformer-based models. The study outlines practical considerations for model deployment and highlights future directions, including the use of self-supervised learning, lightweight models for edge devices, and multi-modal data integration. The findings highlight the advantages of AI-driven segmentation over traditional methods, improving precision and scalability for applications in environmental monitoring, urban planning, and disaster management.

Keywords

Satellite image segmentation, artificial intelligence, deep learning, convolutional neural networks, transformers, remote sensing, land cover classification, semantic segmentation, total variation regularization, loss functions, Swin Transformer, U-Net, DeepLabV3+, Mask R-CNN, environmental monitoring, urban planning, disaster management, geospatial analysis, image processing, self-supervised learning.

1. Introduction

Satellite imagery plays a critical role in numerous domains, ranging from environmental monitoring and urban planning to disaster management and agricultural analysis. These images provide a comprehensive and up-to-date overview of the Earth's surface, enabling researchers, policymakers, and industry experts to make informed decisions. The advent of high-resolution satellite imaging has revolutionized the ability to observe, analyze, and respond to changes in the environment. For instance, satellite images can be used to track deforestation, monitor water levels in rivers, or assess the impact of urbanization. One of the key challenges in leveraging satellite imagery is the vast amount of data generated daily, making manual analysis infeasible. It necessitates the development of automated systems that can efficiently process, analyze, and extract meaningful information from satellite images. Among these tasks, object segmentation stands out as a fundamental step that underpins various applications.

Object segmentation refers to the process of identifying and delineating objects within an image, such as rivers, forests, or buildings. In the context of satellite imagery, segmentation allows for the classification and spatial mapping of different land cover types, which is essential for numerous practical applications:

- Environmental Monitoring is the process of identifying deforestation patterns, monitoring water bodies, and assessing changes in vegetation over time;
- Urban Development is the process of mapping urban growth, analyzing infrastructure distribution, and planning new developments;
- Disaster Response is the act of rapidly assessing affected areas during floods, earthquakes, or wildfires to guide relief efforts.

[1]CMIS-2025: Eighth International Workshop on Computer Modeling and Intelligent Systems, May 5, 2025, Zaporizhzhia, Ukraine

victoria.a.vysotska@lpnu.ua (V. Vysotska); kyrylo.smelyakov@nure.ua (K. Smelyakov); stiv161272@gmail.com (S. Osiievskyi); volodymyr.iartsev@nure.ua (V. Yartsev)

ORCID: 0000-0001-6417-3689 (V. Vysotska); 0000-0001-9938-5489 (K. Smelyakov); 0000-0003-0861-9417 (S. Osiievskyi); 0009-0000-5158-6679 (V. Yartsev)

Manual segmentation is not only time-consuming but also prone to errors due to the complexity of satellite images, which often include overlapping features, varying lighting conditions, and differences in resolution. It underscores the importance of employing advanced technologies, particularly Artificial Intelligence (AI), to achieve accurate and efficient segmentation.

In recent years, the integration of AI, particularly deep learning techniques, has significantly advanced the field of image segmentation. Traditional image processing methods relied on handcrafted features and domain-specific algorithms, which were limited in their ability to generalize across diverse datasets. AI-based methods, on the other hand, utilize neural networks that can learn complex patterns from large datasets. Notable advancements include:

- Convolutional Neural Networks (CNNs) are widely used for feature extraction and classification in images. Architectures like U-Net and Mask R-CNN have been specifically designed for image segmentation tasks.
- Semantic segmentation is assigning a class label to each pixel in the image, enabling detailed object identification.
- Instance segmentation is distinguishing between different objects of the same class, such as multiple buildings in a cityscape.

AI-driven segmentation not only enhances accuracy but also drastically reduces the time required for analysis. It has made it feasible to process large-scale satellite datasets in near real-time.

This study aims to explore the application of AI techniques for the segmentation of objects in satellite imagery. The primary objectives include developing a robust framework for the automatic classification of land cover types such as rivers, forests, and buildings, evaluating the performance of state-of-the-art segmentation models on satellite datasets, and identifying the challenges and limitations associated with AI-driven segmentation methods while proposing potential solutions.

2. Related works

Analyzing recent studies [1-3], it is evident that the field of automatic segmentation and classification of satellite images has significantly advanced in recent years. The application of deep learning and computer vision techniques has led to improved accuracy in land cover classification, urban planning, and environmental monitoring. Modern AI-driven methods enable precise recognition of objects such as rivers, forests, and buildings, supporting large-scale geospatial analysis.

In this regard, convolutional neural networks (CNNs) remain the dominant approach for image classification. The work [4] introduces a deep learning model that utilizes multiscale feature extraction to enhance segmentation accuracy in high-resolution satellite images. Similarly, [5] explores the use of fully convolutional networks (FCNs) for pixel-wise classification, demonstrating superior performance in detecting land cover changes. A study in [6] proposes an attention-based UNet model to improve feature localization and boundary detection, reducing misclassification errors in heterogeneous landscapes.

Recent research has also investigated hybrid models that integrate traditional machine learning with deep learning approaches. For example, in [7], a combination of random forest classifiers with deep CNNs is proposed to enhance feature selection and improve classification robustness. The paper [8] presents an ensemble learning approach that combines CNNs with support vector machines (SVMs) to refine urban area detection. Additionally, [9] explores self-supervised learning techniques to overcome the challenge of limited labelled datasets, demonstrating their effectiveness in land-use classification.

Another growing trend is the use of transformer-based architectures for satellite image analysis. In [10], a Vision Transformer (ViT) model is applied to large-scale remote sensing datasets, outperforming CNN-based methods in classification accuracy. Similarly, [11] introduces a hybrid Swin Transformer model that captures long-range dependencies in high-resolution imagery, improving segmentation results for complex terrain. Furthermore, [12] proposes a spatio-temporal transformer model for monitoring land cover changes over time, enabling more efficient change detection analysis.

Beyond supervised learning, researchers are exploring semi-supervised and unsupervised techniques for classification. The study [13] utilizes generative adversarial networks (GANs) to generate synthetic training samples, reducing dependency on manually labelled datasets. In [14], self-organizing maps (SOMs) are used for clustering satellite images, effectively identifying regions with similar land cover characteristics. The work [15] proposes a contrastive learning framework that

leverages large unlabeled datasets to improve classification accuracy with minimal human annotation.

Several studies focus on domain adaptation and transfer learning to improve model generalization across different satellite datasets. In [16], a domain adaptation framework is introduced to fine-tune pre-trained models on diverse geospatial datasets, achieving higher accuracy in cross-region classification tasks. The research in [17] explores few-shot learning techniques to classify rare land cover types with limited training samples. Meanwhile, [18] presents a meta-learning approach that adapts AI models to new satellite images with minimal re-training, significantly reducing computational costs.

Additionally, cloud computing and edge AI are being leveraged to accelerate the processing of satellite images in real-time. In [19], a cloud-based deep learning framework is developed for large-scale geospatial analysis, allowing efficient processing of massive satellite datasets. The study [20] investigates the use of edge AI devices for real-time segmentation, enabling fast decision-making in environmental monitoring applications.

2.1. Traditional Approaches to Object Segmentation in Satellite Imagery

Before the advent of AI and deep learning, object segmentation in satellite imagery relied primarily on conventional image processing and computer vision techniques. These methods often utilized handcrafted features, statistical models, and rule-based systems to identify and classify objects. One of the earliest and most commonly used approaches was thresholding, where pixel values were categorized based on predefined intensity levels. This method [21] proved to be particularly effective for binary segmentation tasks, such as differentiating water bodies from land. However, it was not capable of handling complex landscapes with multiple land cover types.

Another widely adopted technique was edge detection [22], which involved detecting boundaries between objects using operators such as Sobel, Canny, and Laplacian filters. While effective in delineating distinct objects, edge detection often struggled in cases where boundaries were unclear due to noise, shadows, or similar textures.

Region-based segmentation methods [23], such as Watershed and Mean-Shift, sought to improve edge detection by clustering pixels based on similarities in colour, texture, or spatial proximity. These methods worked well for specific applications but required extensive tuning and often failed when dealing with highly heterogeneous satellite images.

A more advanced approach was object-based image analysis (OBIA), which segmented images into meaningful objects rather than individual pixels. OBIA utilized techniques such as hierarchical clustering and region-growing algorithms, making it more effective for land-use classification. However, it still required human intervention for parameter selection and lacked adaptability to varying datasets.

Despite their utility, traditional segmentation methods had several limitations, including:

- Poor generalization across different geographic regions and image conditions;
- High sensitivity to noise and lighting variations, leading to inconsistent results;
- There is a lack of contextual understanding, as these methods relied solely on pixel values rather than learning from large datasets.

2.2. The Emergence of Machine Learning for Image Segmentation

To address the limitations of traditional methods [24], machine learning (ML) techniques were introduced, leveraging statistical models to improve segmentation accuracy. Supervised learning approaches, such as decision trees, support vector machines (SVM), and random forests, became popular for classifying satellite images. These models were trained on labelled datasets, enabling them to recognize patterns more effectively than rule-based systems.

One of the significant breakthroughs in ML-based segmentation was the adoption of k-means clustering and Gaussian mixture models (GMMs) for unsupervised classification. These methods grouped pixels based on statistical similarities, allowing for automatic identification of land cover categories. However, they still required feature engineering and struggled with complex object boundaries. A key advancement came with the introduction of deep learning [25], which eliminated the need for manual feature extraction by allowing models to learn hierarchical representations

directly from data. It marked a paradigm shift in satellite image segmentation, as deep learning models significantly outperformed traditional machine learning methods.

2.3. Deep Learning for Satellite Image Segmentation

Deep learning, particularly convolutional neural networks (CNNs), revolutionized the field of image analysis by enabling end-to-end learning of spatial features. Several architectures [26] have been developed to tackle the specific challenges of satellite image segmentation:

- Fully Convolutional Networks (FCNs) are one of the first deep-learning approaches for segmentation. FCNs replaced traditional fully connected layers with convolutional layers, allowing for pixel-wise classification.
- U-Net is an architecture designed specifically for biomedical and remote sensing applications, featuring an encoder-decoder structure that enhances segmentation accuracy.
- Mask R-CNN is an extension of Faster R-CNN that enables instance segmentation by distinguishing between different objects of the same category.
- DeepLabV3+ is a model that utilizes atrous spatial pyramid pooling to capture multiscale information, making it practical for segmenting objects of varying sizes.

These models have significantly improved segmentation accuracy in satellite imagery by learning complex spatial relationships and handling diverse environments. However, they also introduce new challenges, such as high computational costs and the need for large labelled datasets.

2.4. Comparison of Traditional and AI-Based Methods

A comparison of traditional and AI-based segmentation methods highlights the advantages of deep learning in terms of accuracy, adaptability, and scalability. The list is shown in Table 1.

Table 1
Comparison of traditional and AI-based methods

Method	Strengths	Weaknesses
Thresholding	Simple and computationally efficient	Limited to binary segmentation, sensitive to noise
Edge Detection	Effective for boundary delineation	Struggles with complex landscapes and occlusions
Region-Based Methods	Captures spatial relationships	Requires fine-tuned parameters, not scalable
Machine Learning (SVM, Random Forests)	More robust than traditional methods	Requires handcrafted features, limited contextual understanding
Deep Learning (CNNs, U-Net, Mask R-CNN)	High accuracy, automatic feature extraction	Requires large datasets, computationally expensive

2.5. Gaps in Existing Research and Future Directions

Despite significant progress in AI-driven segmentation, several challenges remain. One of the main issues is data scarcity, as high-quality labelled satellite datasets are often limited, making it difficult to scale supervised learning approaches. Another challenge lies in computational constraints, since training deep learning models requires substantial resources that may not be accessible in all research settings. Additionally, there is the problem of generalization across regions — models trained on specific geographic areas often struggle to perform accurately in different environments due to variations in landscape features.

To address these challenges, future research should focus on developing self-supervised and semi-supervised learning approaches that reduce dependence on labelled data. There is also a growing need to optimize lightweight AI models capable of real-time processing on edge devices and satellites. Furthermore, integrating multi-modal data sources, such as LiDAR and hyperspectral imagery, can significantly enhance segmentation accuracy and model robustness.

3. Methodology

3.1. Overview of the Methodology

The proposed study [27] employs AI techniques to perform object segmentation on satellite images, focusing on classifying land cover types such as rivers, forests, and buildings. The methodology consists of several key stages, including data collection, preprocessing, model selection, training, and evaluation. This structured approach ensures the development of an efficient and accurate segmentation system tailored for satellite imagery analysis. The workflow begins with the identification of suitable high-resolution satellite imagery datasets for training and evaluation. This is followed by preprocessing steps aimed at enhancing image quality, normalizing data, and preparing segmentation masks. Next, appropriate deep learning architectures optimized for segmentation tasks are selected. The training and optimization phase involves using annotated datasets and fine-tuning model hyperparameters. Model performance is then assessed using standard segmentation metrics to ensure effectiveness. Finally, deployment considerations are addressed, focusing on the real-world applicability of the system and its computational requirements. Each of these stages plays a critical role in ensuring the accuracy and robustness of the segmentation model.

3.2. Dataset Selection

Selecting an appropriate dataset is essential for training an AI-based segmentation model. This study considers publicly available satellite datasets that provide high-resolution images and corresponding segmentation masks. Some of the most commonly used datasets include:

- Sentinel-2 Dataset is a multispectral satellite dataset provided by the European Space Agency (ESA), which is widely used for land cover classification;
- LandCover.ai is a dataset specifically designed for semantic segmentation of aerial and satellite imagery featuring manually annotated masks;
- DeepGlobe Land Cover Classification Dataset is a benchmark dataset that provides annotated satellite images covering urban, agricultural, and forested areas;
- SpaceNet is a dataset containing high-resolution satellite imagery and building footprint annotations that is functional for urban planning applications.

3.3. Data Preprocessing

Before training deep learning models, raw satellite images must undergo preprocessing to enhance their quality and suitability for analysis [28]. This preprocessing pipeline involves several essential steps. First, image resizing is performed to standardize image dimensions and ensure consistency across the dataset. Next, normalization scales pixel values to a uniform range, such as [0,1] or [-1,1] — which helps facilitate stable and efficient neural network training. To improve model generalization and reduce overfitting, data augmentation techniques such as rotation, flipping, and brightness adjustments are applied to increase dataset diversity. Finally, mask generation is carried out to create binary or multiclass segmentation masks that correspond to different land cover types, providing the necessary ground truth for supervised learning.

3.4. Model Selection and Implementation

This study explores several state-of-the-art deep learning architectures for semantic segmentation, focusing on convolutional neural networks (CNNs) and transformer-based models. The selected models include:

- U-Net is a widely used segmentation model with an encoder-decoder architecture designed for biomedical and remote sensing applications;
- DeepLabV3+ is a model incorporating atrous spatial pyramid pooling, enabling multiscale feature extraction for improved segmentation accuracy;
- Mask R-CNN is a region-based convolutional neural network capable of performing both instance segmentation and object detection;

- Swin Transformer is a transformer-based model that leverages self-attention mechanisms for efficient image segmentation.

Each model is implemented using the TensorFlow and PyTorch deep learning frameworks, leveraging pre-trained weights to accelerate training and improve performance.

3.5. Training and Optimization

The training process involves feeding annotated satellite images into the selected models and optimizing their parameters using backpropagation. A key aspect of this procedure is selecting an appropriate loss function, such as cross-entropy loss for multi-class segmentation or Dice loss for imbalanced datasets. Optimization is performed using adaptive techniques, such as Adam or SGD with momentum, to adjust model parameters effectively. Learning rate scheduling is employed to dynamically adjust the learning rate during training, improving convergence. Additionally, hyperparameters like batch size and epochs are tuned to balance training efficiency with model performance. To prevent overfitting, regularization techniques such as dropout and batch normalization are also applied throughout the training process.

3.6. Model Evaluation

To evaluate the performance of segmentation models, a range of quantitative metrics is applied [29], each capturing different aspects of model accuracy. One of the most widely used metrics is Intersection over Union (IoU), which quantifies how well the predicted segmentation overlaps with the ground truth. Complementing this, the Dice Coefficient provides a measure of similarity between predicted and actual regions, making it especially effective for datasets with class imbalance. Pixel Accuracy offers a straightforward metric by calculating the proportion of correctly classified pixels in an image. In the context of instance segmentation, Mean Average Precision (mAP) is utilized to assess how accurately individual objects are detected and segmented. Collectively, these metrics enable a thorough and multi-faceted evaluation of model performance across diverse land cover categories.

3.7. Deployment Considerations

Beyond model training, practical deployment considerations are addressed, including:

- Computational Requirements is evaluating hardware demands for real-time segmentation;
- Scalability is ensuring the model can process large-scale satellite datasets efficiently.
- Edge Deployment is exploring lightweight models for satellite or UAV-based applications.

3.8. Visualization of segmentation results

Semantic segmentation is used to identify land surface types from satellite images. The most basic use of the technology is to determine water body contours to provide more accurate cartographic information. Advanced algorithms are used to map roads, identify crop types, and so on.

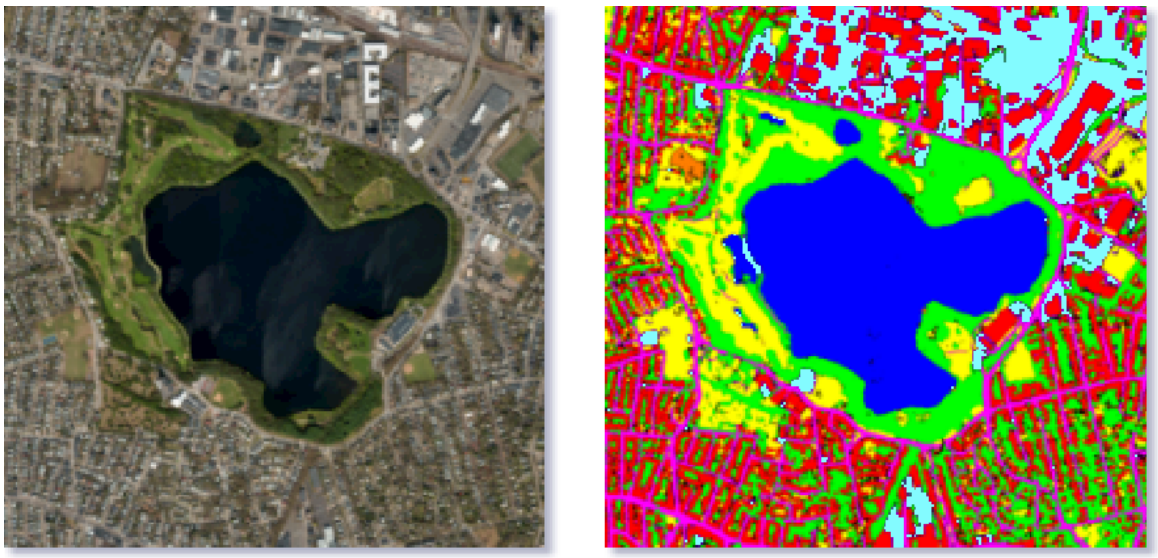

Figure 1: Semantic segmentation of satellite/aerial images [30]

The first example shows a comparison of the original satellite image and its segmented version, where different objects are marked in colours. Automatic segmentation allows you to highlight water bodies, vegetation, buildings, and roads, which is helpful for environmental monitoring and urban planning. Deep learning methods such as U-Net were used. Possible segmentation errors may be due to shadows, low resolution, or insufficient training data. This approach is practical for analyzing landscape changes and mapping territories.

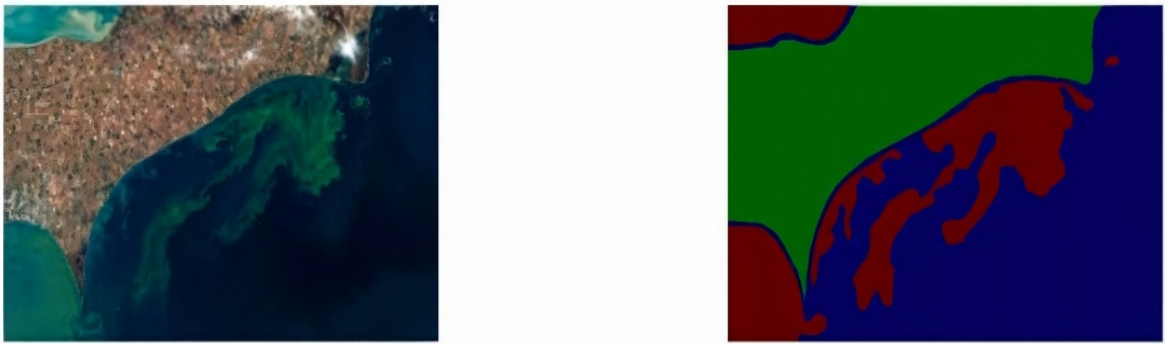

Figure 2: Semantic segmentation of coastal ecosystems [31]

The second example demonstrates the process of segmentation of a satellite image for the analysis of coastal ecosystems.

This method of analysis allows for automatic classification of areas based on spectral characteristics, which is helpful for monitoring the state of water bodies, identifying environmental problems, and planning ecological protection measures.

The third figure is a good example, which combines AI with satellite data to assess real-time disaster impacts like floods, wildfires, and hurricanes. This approach enables rapid situational awareness by visually differentiating damage severity, allowing emergency response teams to prioritize critical areas.

Figure 3: An example of post-disaster images that show damaged areas with colours: green for minor damage, orange for significant damage, and red for destroyed [32]

3.9. Mathematical Formulation

To formalize the segmentation process, let I represent a high-resolution satellite image, where a feature vector x_p characterizes each pixel. The goal of segmentation is to assign a label y_p to each pixel such that the function $f: x_p \rightarrow y_p$ maps input features to semantic categories (e.g., water, vegetation, urban areas).

A typical deep learning-based segmentation model optimizes a loss function $\mathcal{L}$ to minimize the difference between predicted and ground truth labels. One commonly used function is the cross-entropy loss, defined as:

$$L_{ce} = -\sum_{P}\sum_{c} y_p^c lg\left(\hat{y}_p^c\right),$$

where y_p^c is the ground truth probability for class c at pixel p, and $\hat{y}_p^c$ is the predicted probability.

For imbalanced datasets, Dice loss is often used to improve segmentation performance:

$$L_{Dice} = 1 - \frac{2\sum_{p} y_P \hat{y}_p}{\sum_{P} y_p + \sum_{p} \hat{y}_p},$$

where y_p and $\hat{y}_p$ are the ground truth and predicted segmentation masks.

To enhance the spatial coherence of segmentation predictions, a Total Variation (TV) regularization term can be introduced. This regularizer is particularly effective in reducing noise and producing smoother segmentations by discouraging abrupt changes in neighboring pixel classifications. The TV regularization term is defined as follows:

$$L_{tv} = \sum_{p}\left(\vee \hat{y}_{+1} - \hat{y} \vee + \vee \hat{y}_{-1} - \hat{y} \vee\right),$$

where $\hat{y}$ represents the predicted probability or class value at pixel ppp. The expression quantifies the total amount of variation across neighboring pixels, effectively penalizing high-frequency fluctuations in predictions that are not supported by image features. This promotes local smoothness and improves spatial consistency in the segmented output.

However, regularization alone is not sufficient. In practice, the training of segmentation models involves optimizing a composite loss function that balances multiple objectives. For semantic segmentation tasks, commonly used components include the categorical cross-entropy loss L_{ce}, which measures the pixel-wise classification error, the Dice loss L_{Dice}, which is particularly useful in handling class imbalance, and the aforementioned total variation loss L_{tv} .

The final objective function used to train the segmentation network is a weighted combination of these three terms:

$$L_{total} = \alpha L_{ce} + \beta L_{Dice} + \gamma L_{tv},$$

where α, β, γ are hyperparameters controlling the influence of each term. Tuning these coefficients is crucial for achieving optimal performance, as they determine the trade-off between segmentation accuracy, boundary precision, and spatial smoothness.

In most implementations, the choice of these weights depends on the characteristics of the dataset. For instance, datasets with noisy annotations or frequent texture artifacts may benefit from higher γ values to enforce smoother transitions.

4. Experiments and Results

To ensure reliable and reproducible results, the experimental setup is carefully designed, incorporating high-performance computing resources and standardized deep learning frameworks. The key components of the environment include:

Table 2
Computing resources

N	Computing resources	Components	Explanation
1	Hardware Configuration	GPU	NVIDIA RTX 3080 (10GB VRAM) for accelerated model training
		CPU	AMD Ryzen 9 7950X for efficient data preprocessing
		RAM	64GB to handle large satellite image datasets
		Storage	1TB SSD for fast data access and model checkpoints
2	Software and Frameworks	Python 3.8	
		TensorFlow 2.x and PyTorch 1.x	
		OpenCV for image processing	
		GDAL (Geospatial Data Abstraction Library) for handling satellite image formats	
		Albumentations for data augmentation	
3	Dataset	Image resolution	512 × 512 and 1024 × 1024 pixels
		Number of classes	3 (rivers, forests, buildings)
		Training-validation-test split	70%-20%-10%

By using this experimental environment, we ensure that the results are optimized for both accuracy and computational efficiency.

The training process involves fine-tuning hyperparameters to achieve optimal segmentation accuracy. Several aspects of the training procedure are adjusted:

- Experimented with 8, 16, and 32 to balance GPU memory usage and convergence speed;
- Initialized at 0.001 with a step decay to 0.0001 using ReduceLROnPlateau;
- Optimizers Adam and SGD were tested, with Adam providing better stability in the early training phases;
- Set to 100 number of epochs, with early stopping applied when validation loss plateaued;
- Loss Functions are Dice Loss for imbalanced datasets (improves segmentation for small objects like rivers) and Categorical Cross-Entropy for multiclass segmentation.

These hyperparameters were determined through an extensive grid search, ensuring that the models achieved the best possible performance. The trained models were evaluated using standard

segmentation metrics, and the results were compared across different architectures. The performance of each model is summarized in Table 3.

Table 3
The performance of models

Model	IoU (Intersection over Union)	Dice Coefficient	Pixel Accuracy	Training Time (per epoch)
U-Net	85.2%	89.4%	92.3%	12 min
DeepLabV3+	83.7%	88.2%	93.1%	15 min
Mask R-CNN	81.5%	86.8%	91.5%	18 min
Swin Transformer	86.8%	90.1%	92.8%	22 min

From these results, we observe that U-Net performs well across all metrics, making it a strong choice for semantic segmentation tasks. Its encoder-decoder architecture with skip connections allows it to preserve spatial information, which is essential for delineating land cover boundaries accurately.

DeepLabV3+ achieves the highest pixel accuracy, which is particularly beneficial for large-area segmentation tasks where overall classification consistency is critical. Its use of atrous convolution and multi-scale context aggregation contributes to its strength in handling spatially diverse features.

Mask R-CNN provides instance-level segmentation, which is valuable for distinguishing between multiple occurrences of the same object class, such as separate buildings or vehicles. However, it shows a slightly lower IoU due to challenges in dealing with complex and noisy background textures commonly found in natural landscapes. This indicates a trade-off between instance-level precision and overall semantic coherence.

Swin Transformer achieves the best overall performance across metrics, benefiting from its hierarchical vision transformer design and self-attention mechanisms that effectively model long-range spatial dependencies. This makes it especially powerful for capturing subtle patterns and context in high-resolution satellite images. However, this superior accuracy comes at a higher computational cost, which may limit its practical deployment in resource-constrained environments, such as real-time onboard satellite processing or edge devices.

Despite promising results, several challenges remain:

- Misclassification in boundary regions (small objects such as narrow rivers are sometimes misidentified as roads);
- Variability in lighting and atmospheric conditions (shadows and haze in satellite images introduce noise);
- Data scarcity for specific regions (the model generalizes well for well-represented landscapes but struggles with less common environments).

5. Discussions

The experimental results demonstrate the effectiveness of deep learning models for satellite image segmentation, revealing notable variations in performance across different architectures. High Intersection over Union (IoU) and Dice coefficient scores confirm that the models can accurately differentiate between various land cover types, such as rivers, forests, and buildings. Among the evaluated models, the Swin Transformer consistently outperformed traditional CNN-based architectures, benefiting from self-attention mechanisms that effectively capture complex spatial relationships in satellite imagery. U-Net, despite its relatively simple design, delivered competitive results and remains a practical choice for large-scale segmentation tasks due to its computational efficiency and ease of training. DeepLabV3+ excelled in capturing fine details, which is especially advantageous for segmenting narrow rivers and small structures. In contrast, Mask R-CNN proved useful for instance segmentation but encountered difficulties with semantic segmentation of natural landscapes, primarily due to the complexity and variability of background textures.

Several key observations emerged from the analysis. Boundary regions between different land types presented consistent challenges, often resulting in misclassifications at the edges. Data imbalance also impacted model performance, as areas with fewer training examples — such as sparsely represented forest zones, tended to be segmented less accurately. Moreover, model

generalization was found to depend heavily on dataset diversity; models trained on geographically limited data often struggled to accurately segment landscapes from unfamiliar regions. These findings highlight both the strengths and current limitations of AI-based segmentation methods when applied to real-world satellite imagery.

Traditional satellite image segmentation methods, such as thresholding, edge detection, and classical machine learning techniques (e.g., Random Forests, SVM), have been widely used in remote sensing applications. However, these methods often struggle with complex, high-resolution images due to their limited ability to capture hierarchical spatial relationships. They typically rely on handcrafted features and shallow representations, which makes them less effective in handling variations in texture, lighting, and object scale. As a result, their performance tends to degrade in heterogeneous landscapes or when applied to large and diverse satellite datasets.

The list of advantages and disadvantages of models is shown in Table 4.

Table 4
Advantages and disadvantages of models

Method	Advantages	Disadvantages
Thresholding & Edge Detection	Simple, fast, interpretable	Sensitive to lighting conditions and noise
Random Forests & SVM	Effective for small datasets, interpretable	Requires handcrafted features, limited scalability
CNN-based Models (U-Net, DeepLabV3+)	High accuracy, learns spatial hierarchies	Computationally expensive
Transformer-based Models (Swin Transformer)	Captures long-range dependencies, state-of-the-art performance	Requires large datasets and computational power

The results show that deep learning methods significantly outperform classical approaches in terms of segmentation accuracy and robustness. Transformer-based architectures, in particular, demonstrate superior capability in handling complex satellite imagery, suggesting a shift towards these models in remote sensing applications.

The automatic classification of land cover using satellite imagery has numerous real-world applications across various domains. In environmental monitoring [33], AI-based segmentation enables the detection of changes in river paths due to climate change or deforestation, allowing researchers to track the degradation of natural landscapes over time. It also facilitates the assessment of flood-prone areas, contributing to disaster prevention strategies. Similarly, the ability to analyze forest cover loss and land degradation helps environmental organizations and policymakers take appropriate conservation measures.

Urban planning and infrastructure development [34] also greatly benefit from automated segmentation methods. By analyzing satellite images, city planners can monitor urban expansion, identify informal settlements, and evaluate changes in land use. This data is essential for designing sustainable cities and ensuring efficient infrastructure growth. Automated segmentation allows authorities to track the development of new buildings and road networks, supporting informed decision-making in large-scale construction projects.

Despite the advancements in AI-based satellite image segmentation, several challenges remain that hinder widespread adoption and practical implementation. One of the primary issues [35] is the generalization of models across different geographic regions. Satellite images vary significantly based on atmospheric conditions, vegetation types, and urban structures, making it difficult for a model trained on one dataset to perform well in other locations. This limitation necessitates domain adaptation techniques or the collection of diverse training data to improve robustness [36-37]. Another significant challenge [38] is the issue of class imbalance and rare object detection. In many satellite datasets, certain land cover types, such as rivers or buildings, are underrepresented compared to dominant classes like forests or open land. This imbalance leads to biased model predictions, where rare classes are often misclassified or ignored. Addressing this problem requires specialized techniques such as data augmentation, focal loss, and synthetic data generation to ensure balanced learning [39].

6. Conclusions

This study assessed the application of artificial intelligence techniques for object segmentation in satellite imagery, with a specific focus on the automatic classification of land cover types such as rivers, forests, and buildings. A comprehensive comparison was conducted between traditional image processing methods and modern deep learning architectures, including convolutional neural networks (U-Net, DeepLabV3+, Mask R-CNN) and transformer-based models (Swin Transformer).

Experimental results demonstrated that deep learning methods significantly outperform traditional approaches in terms of segmentation accuracy, boundary delineation, and generalization across diverse landscapes. Among the tested models, the Swin Transformer achieved the highest accuracy metrics, while U-Net remained a computationally efficient and competitive baseline. However, the performance gains of advanced models come with higher computational costs and increased demand for annotated data.

Despite outcomes, the study identified key limitations in current AI-based segmentation approaches. These include: reduced model performance in regions with limited training representation, difficulty in accurately classifying boundary zones and rare object classes, and challenges in generalizing to unseen geographic areas. The research also highlighted the importance of selecting appropriate models based on deployment scenarios — particularly when balancing performance with computational efficiency.

In conclusion, the findings underscore the practical potential of deep learning in satellite image segmentation and emphasize the necessity of addressing current challenges to facilitate broader adoption in environmental monitoring, urban development, and disaster response scenarios.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. Sharma, S. R. Chopra, S. G. Sapate, K. Arora, M. K. I. Rahmani, S. Jha, S. Ahmad, M. E. Ahmed, H. A. M. Abdeljaber, J. Nazeer, Artificial intelligence techniques for landslides prediction using satellite imagery, IEEE Access (2024) 1. doi:10.1109/access.2024.3446037.

[2] C. Marrocco, A. Bria, F. Tortorella, S. Parrilli, L. Cicala, M. Focareta, G. Meoli, M. Molinara, Illegal microdumps detection in multi-mission satellite images with deep neural network and transfer learning approach, IEEE Access (2024) 1. doi:10.1109/access.2024.3409393.

[3] L. Xu, Y. Liu, S. Shi, H. Zhang, D. Wang, Land-Cover classification with high-resolution remote sensing images using interactive segmentation, IEEE Access (2022) 1. doi:10.1109/access.2022.3205327.

[4] S. Shakya, S. Kumar, M. Goswami, Deep learning algorithm for satellite imaging based cyclone detection, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 827–839. doi:10.1109/jstars.2020.2970253.

[5] Z. Zhan, X. Zhang, Y. Liu, X. Sun, C. Pang, C. Zhao, Vegetation land use/land cover extraction from high-resolution satellite images based on adaptive context inference, IEEE Access 8 (2020) 21036–21051. doi:10.1109/access.2020.2969812.

[6] C.-J. Zhang, J.-X. Guo, L.-M. Ma, X.-Q. Lu, W.-C. Liu, TCCL-DenseFuse: infrared and water vapor satellite image fusion model using deep learning, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2023) 1–25. doi:10.1109/jstars.2023.3277842.

[7] H. Yi, X. Chen, D. Wang, S. Du, B. Xu, F. Zhao, An epipolar resampling method for multi-view high resolution satellite images based on block, IEEE Access 9 (2021) 162884–162892. doi:10.1109/access.2021.3133664.

[8] K. K. Jena, S. K. Bhoi, S. R. Nayak, R. Panigrahi, A. K. Bhoi, Deep convolutional network based machine intelligence model for satellite cloud image classification, Big Data Min. Anal. 6.1 (2023) 1–12. doi:10.26599/bdma.2021.9020017.

[9] Z. Xu, Y. Jiang, J. Wang, Y. Wang, A dual branch multi-scale stereo matching network for high-resolution satellite remote sensing images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2024) 1–17. doi:10.1109/jstars.2024.3502842.

[10] H. Mansourifar, A. Moskowitz, B. Klingensmith, D. Mintas, S. J. Simske, GAN-based satellite imaging: A survey on techniques and applications, IEEE Access (2022) 1. doi:10.1109/access.2022.3221123.
[11] E. Cho, E. Kim, Y. Choi, Cloud cover prediction model using multi-channel geostationary satellite images, IEEE Trans. Geosci. Remote Sens. (2024) 1. doi:10.1109/tgrs.2024.3473992.
[12] M. F. Humayun, F. A. Nasir, F. A. Bhatti, M. Tahir, K. Khurshid, YOLO-OSD: optimized ship detection and localization in multi-resolution SAR satellite images using a hybrid data-model centric approach, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2024) 1–20. doi:10.1109/jstars.2024.3365807.
[13] H. Yi, X. Chen, D. Wang, S. Du, N. Guo, Methods for the epipolarity analysis of pushbroom satellite images based on the rational function model, IEEE Access 8 (2020) 103973–103983. doi:10.1109/access.2020.2999393.
[14] S. Shende, CNN based missing object detection, Int. J. Res. Appl. Sci. Eng. Technol. 11.4 (2023) 956–959. doi:10.22214/ijraset.2023.50138.
[15] K. Karwowska, D. Wierzbicki, Using super-resolution algorithms for small satellite imagery: A systematic review, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2022) 1. doi:10.1109/jstars.2022.3167646.
[16] Z. Hu, K. Zhang, Y. Liu, Edge constrained DSM refinement based on shading from high resolution multi-view satellite images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2025) 1–12. doi:10.1109/jstars.2025.3526817.
[17] IEE Satellite Systems & Applications Professional Network, Personal broadband satellite: seminar, tuesday, january 2002, IEE, savoy place, WC2R 0BL, UK, IEE Professional Networks, London, 2002.
[18] H. Ouchra, A. Belangour, A. Erraissi, Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study, IEEE Access (2023) 1. doi:10.1109/access.2023.3293828.
[19] Z. Chen, W. Li, Z. Cui, Y. Zhang, Surface depth estimation from multi-view stereo satellite images with distribution contrast network, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2024) 1–10. doi:10.1109/jstars.2024.3457616.
[20] K. Sasaki, T. Sekine, W. Emery, Enhancing the detection of coastal marine debris in very high resolution satellite imagery via unsupervised domain adaptation, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2024) 1–16. doi:10.1109/jstars.2024.3364165.
[21] X. Zuo, J. Teng, F. Su, Z. Duan, K. Yu, Multi-model combination bathymetry inversion approach based on geomorphic segmentation in coral reef habitats using icesat-2 and multispectral satellite images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2024) 1–13. doi:10.1109/jstars.2024.3523296.
[22] T. D. Nguyen, A. Shinya, T. Harada, R. Thawonmas, Segmentation mask refinement using image transformations, IEEE Access 5 (2017) 26409–26418. doi:10.1109/access.2017.2772269.
[23] S. Yoneda, G. Irie, M. Nishiyama, Canonical plane segmentation without annotating pixel-level object regions for image registration, IEEE Access (2024) 1. doi:10.1109/access.2024.3373463.
[24] A. Naseer, N. A. Mudawi, M. Abdelhaq, M. Alonazi, A. Alazaib, A. Algarni, A. Jalal, CNN-based object detection via segmentation capabilities in outdoor natural scenes, IEEE Access (2024) 1. doi:10.1109/access.2024.3413848.
[25] H. Li, G.-L. Yuan, C. Xu, Siamese contour segmentation network for multi-state object tracking, SSRN Electron. J. (2022). doi:10.2139/ssrn.4303230.
[26] Y. Liang, Y. Zhang, Y. Wu, S. Tu, C. Liu, Robust video object segmentation via propagating seams and matching superpixels, IEEE Access 8 (2020) 53766–53776. doi:10.1109/access.2020.2981140.
[27] Y. Niu, C. Su, W. Guo, Salient object segmentation based on superpixel and background connectivity prior, IEEE Access 6 (2018) 56170–56183. doi:10.1109/access.2018.2873022.
[28] T.-W. Yu, M. A. Sarwar, Y.-A. Daraghmi, S.-H. Cheng, T.-U. Ik, Y.-L. Li, Spatiotemporal activity semantics understanding based on foreground object segmentation: icounter scenario, IEEE Access (2022) 1. doi:10.1109/access.2022.3178609.
[29] Real-time object segmentation based on convolutional neural network with saliency optimization for picking, J. Syst. Eng. Electron. 29.6 (2018) 1300. doi:10.21629/jsee.2018.06.17.
[30] B. Ray, A simple guide to semantic segmentation, 2019. URL: https://medium.com/beyondminds/a-simple-guide-to-semantic-segmentation-effcf83e7e54.

[31] Kayumov O., Segmentation of forest fellings based on satellite imagery data using the maskformer model, 2023. URL: https://research-journal.org/archive/10-136-2023-october/10.23670/irj.2023.136.16.
[32] A. Vina, Using computer vision to analyze satellite imagery, 2024. URL: https://www.ultralytics.com/blog/using-computer-vision-to-analyse-satellite-imagery.
[33] X. Chen, W. Chen, L. Su, T. Li, Slender flexible object segmentation based on object correlation module and loss function optimization, IEEE Access (2023) 1. doi:10.1109/access.2023.3261543.
[34] X. Jiang, Y. Gao, Z. Fang, P. Wang, B. Huang, An end-to-end human segmentation by region proposed fully convolutional network, IEEE Access 7 (2019) 16395–16405. doi:10.1109/access.2019.2892973.
[35] K. Smelyakov, S. Smelyakov and A. Chupryna, "Advances in Spatio-Temporal Segmentation of Visual Data," in Adaptive Edge Detection Models and Algorithms. – Springer Nature Switzerland AG 2020, pp. 1–51. doi:10.1007/978-3-030-35480-0_1.
[36] S. Voloshyn, et al., "Big Data Analysis for Multispectral Images Recognition Based on Deep Learning," IEEE 16th International Conference on Computer Sciences and Information Technologies, vol. 1, pp. 160-170, 2021. doi: 10.1109/CSIT52700.2021.9648650.
[37] A. Sartiukova, et al., "The Multiclass Classification of Objects Based on Multispectral Images Recognition," IEEE 16th International Conference on Computer Sciences and Information Technologies, vol. 1, pp. 52-60, 2021. doi: 10.1109/CSIT52700.2021.9648719.
[38] K. Smelyakov, P. Dmitry, M. Vitalii and C. Anastasiya, "Investigation of network infrastructure control parameters for effective intellectual analysis," 2018 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, 2018, pp. 983-986, doi: 10.1109/TCSET.2018.8336359.
[39] S. Tchynetskyi, et al.,"A Neural Network Development for Multispectral Images Recognition," IEEE 16th International Conference on Computer Sciences and Information Technologies, vol. 2, pp. 278-284, 2021. doi: 10.1109/CSIT52700.2021.9648735.