# Automated Land Cover Classification from Satellite Images Using Deep Transformer Networks

Ketan Kanjiya[1], Piyush Sonani[2], Upendrasinh Zala[3]
[1]*Chief Research Officer, Kshatra infotech Pvt Ltd, Ahmedabad, Gujarat, India*
[2]*Chief Technology Officer, Kshatra infotech Pvt Ltd, Ahmedabad, Gujarat, India*
[3]*Chief Executive Officer, Kshatra infotech Pvt Ltd, Ahmedabad, Gujarat, India*

***Abstract*—Accurate land cover classification from satellite imagery is essential for environmental monitoring, urban planning, and sustainable resource management. Traditional machine learning and convolutional neural network approaches have achieved considerable success in remote sensing tasks; however, they often struggle to capture long-range spatial dependencies present in high-resolution satellite images. This study proposes a transformer based deep learning framework for automated land cover classification using semantic segmentation. The proposed approach utilizes the Mask2Former architecture with a Swin-Large backbone to perform pixel-level classification of multiple land cover categories. A patch-based training strategy is adopted to efficiently process large satellite images, while data augmentation techniques are applied to improve model generalization. To generate full resolution predictions, a sliding-window inference strategy is employed. The model is evaluated using standard semantic segmentation metrics, including Pixel Accuracy, Intersection over Union, Precision, Recall, and Dice coefficient. Experimental results show that the proposed framework achieves a mean Intersection over Union (mIoU) of 0.5701 and a pixel accuracy of 0.8716, indicating reliable segmentation performance across multiple land cover categories. These findings highlight the effectiveness of transformer-based architectures for large scale geospatial analysis and automated land cover mapping from satellite imagery.**

***Index Terms*—Deep Learning, Geospatial Analysis, Land Cover Classification, Remote Sensing, Satellite Imagery, Semantic Segmentation**

## I. INTRODUCTION

Satellite imagery has become an essential resource for monitoring and understanding the Earth's surface. Accurate identification of land cover types from satellite images plays a critical role in various applications such as environmental monitoring, urban planning, agricultural management, and sustainable development. Land cover classification involves categorizing different regions of an image into predefined classes such as urban areas, agricultural land, forests, water bodies, and barren land. With the increasing availability of high-resolution satellite imagery, automated and efficient methods for land cover classification have become increasingly important.

Traditional approaches to land cover classification relied heavily on manual interpretation or classical machine learning techniques using handcrafted features. These methods often struggle with large-scale datasets and complex spatial patterns present in satellite imagery. Moreover, variations in illumination, seasonal changes, and heterogeneous landscapes make accurate classification a challenging task. As a result, there has been a growing interest in deep learning-based techniques that can automatically learn hierarchical feature representations from large volumes of image data.

In recent years, convolutional neural networks (CNNs) have significantly improved the performance of image classification and segmentation tasks. Architectures such as U-Net, DeepLab, and SegNet have been widely used for semantic segmentation in remote sensing applications. However, CNN based models often have limitations in capturing long-range dependencies and global contextual information, which are important for accurately segmenting large and complex satellite scenes.

Transformer based models have recently emerged as a powerful alternative for computer vision tasks. By leveraging self-attention mechanisms, transformer architectures can effectively model global relationships within an image. One such model, Mask2Former, has demonstrated strong performance in various segmentation tasks, including semantic, instance, and panoptic segmentation. Mask2Former combines the strengths of transformer-based architectures with mask-based prediction mechanisms, allowing it to produce more accurate and flexible segmentation outputs.

In this study, we investigate the effectiveness of Mask2Former for land cover classification using high-resolution satellite imagery. The experiments are conducted on the DeepGlobe Land Cover Classification dataset, which contains satellite images annotated with multiple land cover categories including urban land, agriculture land, rangeland, forest land, water, barren land, and unknown regions. The proposed approach utilizes a transformer-based backbone along with advanced data preprocessing and augmentation techniques to improve segmentation performance.

The model is trained to perform multi-class semantic segmentation of satellite images, enabling pixel-level classification of different land cover types. Performance is evaluated using commonly used segmentation metrics such as mean Intersection over Union, precision, recall, Dice coefficient, and pixel accuracy. The results demonstrate the effectiveness of transformer-based architectures for accurate land cover classification in satellite imagery and highlight their potential for large-scale geospatial analysis.

## II. LITERATURE REVIEW

Land cover monitoring is a dynamic activity essential for understanding the interaction between human activities and environmental changes. As industrial and technological development accelerates environmental transformations worldwide, frequent land cover monitoring has become crucial for urban planning, biodiversity conservation, and sustainable resource management. Remote sensing imagery provides an effective data source for such studies by enabling the classification of different land cover types such as forests, urban regions, and agricultural areas [1].

The evolution of land cover classification methodologies reflects significant advancements in remote sensing technologies, computational algorithms, and standardization frameworks [2]. Historically, land cover classification relied on visual interpretation and manual digitization of aerial photographs, methods that were labor-intensive, subjective, and prone to human error [3-4]. To improve these processes, the field transitioned toward machine learning algorithms, such as support vector machines, random forests, K-nearest neighbor, and decision trees [5-7]. Among these, support vector machines and random forests became particularly popular due to their ability to process high-dimensional datasets and model complex nonlinear relationships [8-9]. However, traditional machine learning methods often require extensive manual feature engineering and are limited by their capacity to handle the massive volumes of data provided by modern satellite sensors [10].

The emergence of deep learning has revolutionized remote sensing by offering superior performance through automatic feature extraction [11]. Convolutional Neural Networks have become the leading paradigm in this domain, demonstrating remarkable success in image recognition and semantic segmentation tasks like building extraction and land cover change detection [12]. Architectures such as U-Net, DeepLabV3+, and LinkNet have set new benchmarks for high-resolution land cover mapping [13]. Despite their success, CNNs face inherent challenges, including a heavy reliance on local convolutional kernels, which can limit their ability to model long-range spatial dependencies [14]. This limitation often leads to a "semantic gap" in complex scenes where global context is necessary to distinguish between spectrally similar classes, such as different types of vegetation or fragmented urban interfaces [15].

To address these shortcomings, Deep Transformer Networks have recently emerged as a powerful alternative [16]. Originally designed for natural language processing, Vision Transformers utilize self-attention mechanisms to capture global correlations and semantic relationships between distant image grids [17]. Research has shown that Transformers can extract complex correlation

information between labels more effectively than traditional RNNs or CNNs [16]. For instance, "Cropformer" was proposed to accomplish both global and local feature extraction for multi-scenario crop classification [14]. Similarly, the "Extended Vision Transformer" has been developed as a multimodal deep learning framework to enhance classification precision [18]. Recent studies also explore fused architectures that combine the inductive biases of CNNs with the global modeling capabilities of Transformers [17]

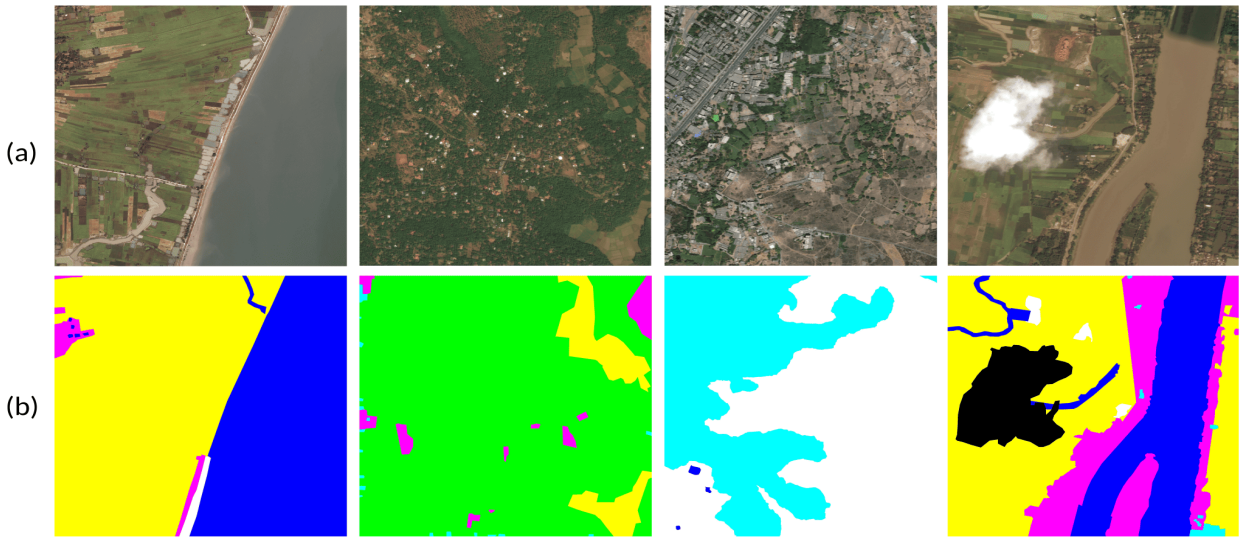



Figure 1: Sample images from the dataset: (a) satellite images and (b) corresponding ground truth segmentation masks representing the seven land cover classes.

For example, researchers have proposed network-level fusion architectures that integrate inverted bottleneck residuals with self-attention layers to improve local information extraction [17]. Such hybrid models aim to leverage the spatial detail captured by convolutional kernels alongside the global context provided by attention mechanisms [16].

Despite these technical advances, label scarcity remains a significant barrier, as high-quality manual annotation is time consuming and costly [19]. To mitigate this, techniques such as transfer learning, data augmentation, and self-supervised learning have been widely adopted [20]. Self-supervised strategies like "Bootstrap Your Own Latent" allow encoders to learn meaningful representations from large amounts of unlabeled imagery, significantly reducing the labeled data required for high-resolution mapping [21]. Furthermore, the integration of Explainable AI, such as Grad-CAM and SHAP, has become crucial for making these "black-box" models more transparent and trustworthy for critical decision making in urban planning and environmental monitoring [22]. Collectively, these developments emphasize the growing importance of Transformer based methodologies in enhancing the precision, scalability, and interpretability of automated land cover classification [23].

## III. DATASET

This study utilizes the publicly available DeepGlobe Land Cover Classification Dataset, which is widely used for research on land cover segmentation from satellite imagery. The dataset contains satellite images collected by DigitalGlobe with a spatial resolution of 50 cm per pixel, enabling detailed observation of various land cover types. Each image has a resolution of 2448 × 2448 pixels and is provided in RGB format.

The dataset consists of 803 training images, 171 validation images, and 172 test images. For the training set, each satellite image is paired with a corresponding ground truth mask that provides pixel-level annotations of land cover classes. The masks

are represented as RGB images where each color corresponds to a specific land cover category.

The dataset contains seven land cover classes with distinct colors in the segmentation masks: urban land represented by cyan, agricultural land by yellow, rangeland by magenta, forest land by green, water bodies by blue, barren land by white, and unknown regions by black. This annotation scheme allows the task to be formulated as a multi-class semantic segmentation problem, where each pixel in the satellite image is assigned to one of the predefined land cover categories. Figure 1 illustrates sample satellite images and their corresponding segmentation masks from the dataset.

## IV. METHODOLOGY

This study proposes a deep learning-based framework for automatic land cover classification from high-resolution satellite imagery. The methodology consists of four main stages: data preprocessing, model architecture, training strategy, and evaluation. The overall workflow involves converting RGB mask annotations into semantic labels, training a transformer-based segmentation model, and evaluating the predicted land cover maps using standard segmentation metrics.

### A. Data Preprocessing

The satellite images in the dataset have a resolution of 2448 × 2448 pixels, which is computationally expensive for direct training. Therefore, a patch-based training strategy was adopted. Each image and its corresponding mask were randomly cropped into smaller patches of 512 × 512 pixels during training. For validation, center cropping was applied to ensure consistent evaluation.

The ground truth masks in the dataset are provided as RGB images with color coded classes. These masks were converted into integer label maps, where each pixel corresponds to a class index ranging from 0 to 6. Pixels that do not match any predefined class color are assigned an ignore label (255), which is excluded during loss computation.

To improve the robustness and generalization ability of the model, several data augmentation techniques were applied during training. These include horizontal flipping, vertical flipping, random brightness and contrast adjustment, and hue-saturation variations. Augmentations were implemented using the Albumentations library and applied only to the training samples.

### B. Model Architecture

The proposed approach utilizes the Mask2Former architecture, a transformer based universal segmentation model designed for semantic, instance, and panoptic segmentation tasks. Specifically, the pretrained Mask2Former with Swin-Large backbone was employed as the base model.

Mask2Former integrates a transformer decoder with masked attention mechanisms, allowing the model to predict segmentation masks through a set of learned queries. Unlike traditional convolutional

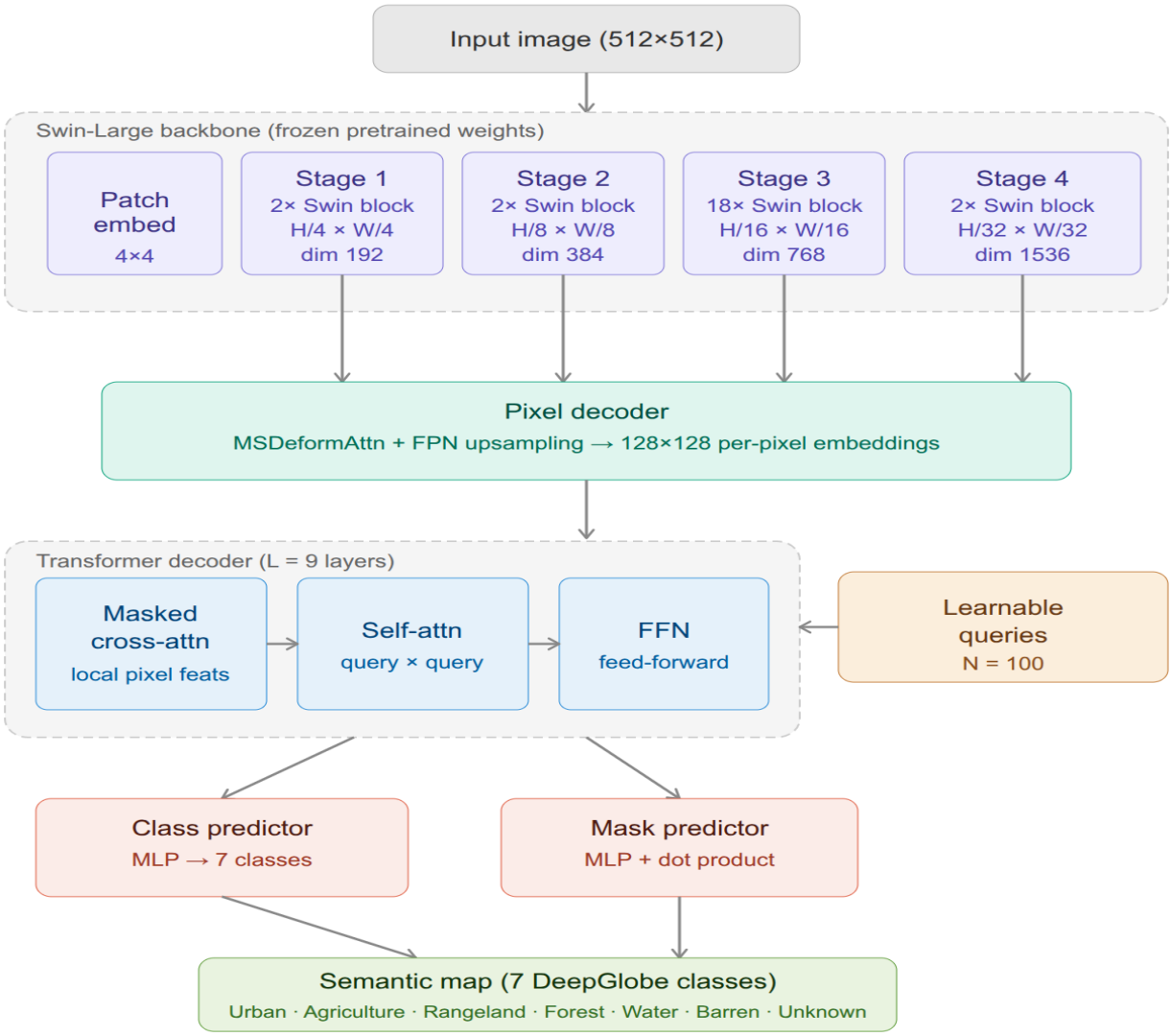



Figure 2: Architecture of the Mask2Former model with a Swin-Large backbone used for land cover semantic segmentation.

segmentation networks that rely on dense pixel-wise classification, Mask2Former generates a set of mask predictions and corresponding class labels, which are combined to produce the final semantic segmentation map.

In this study, the pretrained model originally trained on the ADE20K dataset was adapted to the land cover classification task by replacing the final classification head with a 7-class prediction layer corresponding to the land cover categories in the dataset. The Hugging Face Transformers implementation of Mask2Former was used to facilitate model initialization and training.

C. Training Strategy

The model was trained using the AdamW optimizer with a learning rate of $6 \times 10^{-5}$ and a weight decay of 0.01. A Cosine Annealing Warm Restarts learning rate scheduler was employed to improve convergence and stabilize training.

To reduce GPU memory consumption and accelerate computation, mixed precision training was implemented using automatic mixed precision (AMP). Gradient clipping was also applied to prevent gradient explosion and stabilize training.

The model was trained for 20 epochs using a batch size of 4. During training, the loss provided by the Mask2Former framework was used, which combines classification and mask prediction losses. The best

model was selected based on the highest validation mean Intersection over Union (mIoU) score.

D. Inference Strategy

Since the original satellite images are large, a sliding-window inference approach was used during prediction. Each full-resolution image was divided into overlapping patches of 512 × 512 pixels with a stride of 384 pixels. Predictions from overlapping regions were aggregated and averaged to generate the final segmentation map for the entire image.

This strategy ensures that high-resolution spatial information is preserved while maintaining feasible memory usage during inference.

E. Evaluation Metrics

The performance of the proposed model is evaluated using several standard semantic segmentation metrics, including Pixel Accuracy (PA), Precision, Recall, Intersection over Union (IoU), and Dice Coefficient. These metrics measure the similarity between the predicted segmentation maps and the ground truth annotations.

Pixel Accuracy measures the ratio of correctly classified pixels to the total number of pixels in the image and is defined as

$$\mathrm{PA} = \frac{\text{Number of correctly classified pixels}}{\text{Total number of pixels}} \qquad (1)$$

Precision and Recall evaluate the correctness and completeness of predicted pixels, respectively:

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \qquad (2)$$

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad (3)$$

Intersection over Union (IoU) measures the overlap between predicted and ground truth regions:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \qquad (4)$$

The mean IoU (mIoU) is computed by averaging IoU across all classes. Additionally, the Dice coefficient measures segmentation similarity:

$$\mathrm{Dice} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \qquad (5)$$

Here, TP, FP, FN, and TN denote true positive, false positive, false negative, and true negative pixels, respectively.

## V. RESULTS AND DISCUSSION

This section presents the experimental results obtained using the proposed transformer-based land cover classification framework. The Mask2Former model with a Swin-Large backbone was trained and evaluated on the DeepGlobe Land Cover Classification dataset to assess its effectiveness in performing multi-class semantic segmentation of satellite imagery.

A. Quantitative Evaluation

The performance of the proposed model was evaluated using several widely used semantic segmentation metrics, including Pixel Accuracy (PA), mean Intersection over Union (mIoU), Precision, Recall, and Dice coefficient. These metrics provide a comprehensive assessment of segmentation performance by measuring both classification correctness and spatial overlap between predicted and ground truth regions. Table 1 summarizes the overall performance metrics obtained on the validation dataset. The results indicate that the transformer based Mask2Former architecture is capable of effectively learning complex spatial relationships in satellite imagery.

Table 1. Overall segmentation performance

| Metric | Value |
|---|---|
| Pixel Accuracy | 0.8716 |
| Mean IoU | 0.5701 |
| Mean Precision | 0.6541 |
| Mean Recall | 0.7067 |
| Mean Dice Coefficient | 0.6698 |

In particular, the achieved mean IoU demonstrates strong segmentation performance across multiple land cover classes. High pixel accuracy further confirms that a large proportion of pixels in the validation images were correctly classified.

Compared to traditional convolutional architectures, the self-attention mechanism in transformer networks enables the model to capture long range contextual dependencies. This ability is particularly beneficial for satellite image analysis where spatial relationships between distant regions can provide important contextual information for accurate classification.

B. Class-wise Performance Analysis

To further analyze model behavior, segmentation performance was evaluated separately for each land cover class. Table 2 presents the class-wise Intersection over Union, Precision, Recall, and Dice scores.

Table 2. Class-wise segmentation performance

| Class | IoU | Precision | Recall | Dice |
|---|---|---|---|---|
| Urban Land | 0.8081 | 0.913 | 0.8755 | 0.8939 |
| Agriculture | 0.8765 | 0.972 | 0.8993 | 0.9342 |
| Rangeland | 0.357 | 0.4115 | 0.7293 | 0.5261 |
| Forest | 0.5654 | 0.7747 | 0.6766 | 0.7223 |
| Water | 0.8582 | 0.9257 | 0.9217 | 0.9237 |
| Barren | 0.5251 | 0.5815 | 0.8442 | 0.6886 |
| Unknown | 0 | 0 | 0 | 0 |

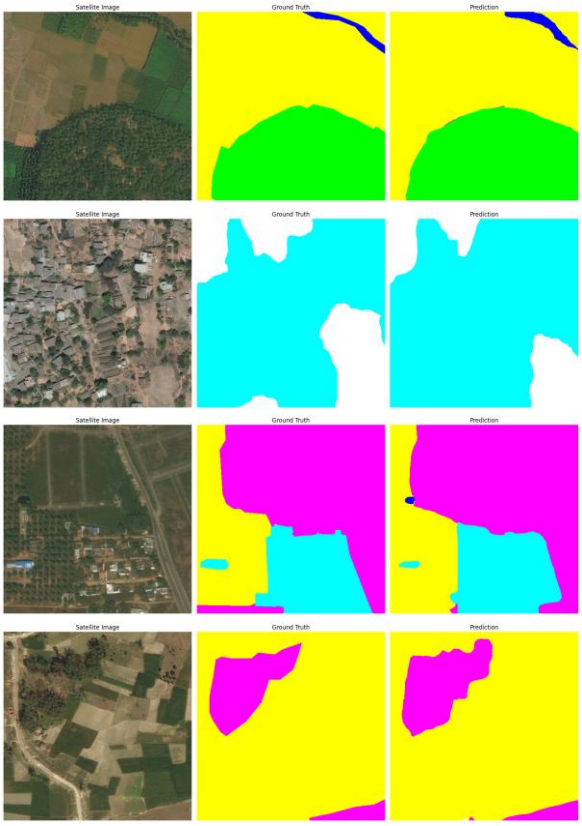
The results show that classes with distinctive spectral characteristics, such as water bodies and urban land, tend to achieve higher segmentation accuracy. These categories have relatively clear boundaries and unique visual patterns, which allows the model to distinguish them more effectively.

In contrast, classes such as rangeland and barren land sometimes exhibit lower IoU values due to their similar spectral appearance and complex spatial patterns. Such ambiguity is common in remote sensing datasets where vegetation types may have overlapping characteristics.

The unknown class obtained zero scores across all metrics. This is primarily due to the limited presence or absence of this class in the validation samples used during evaluation. As a result, the model was not able to learn sufficient representations for this category, leading to no predicted pixels belonging to this class during inference.

Overall, the transformer-based architecture demonstrates strong performance across most classes, highlighting its ability to capture both local spatial features and global contextual relationships.

C. Qualitative Results




(a) (b) (c)

Figure 3: Qualitative segmentation results showing (a) satellite image, (b) ground truth mask, and (c) predicted segmentation map.

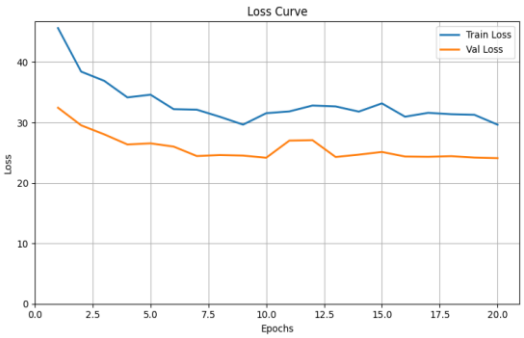
In addition to quantitative evaluation, qualitative analysis was performed by visualizing predicted segmentation maps alongside the corresponding ground truth masks. Figure 3 presents several examples of satellite images, ground truth annotations, and predicted segmentation outputs generated by the proposed model.

The visual results demonstrate that the model successfully identifies major land cover categories such as urban areas, forests, water bodies, and barren land. The predicted segmentation maps closely match the ground truth annotations, indicating that the model effectively captures spatial patterns present in high-resolution satellite imagery.

The sliding-window inference strategy also enables accurate prediction on large images by aggregating overlapping patches. This approach preserves high-resolution spatial details while maintaining feasible memory usage during inference.

D. Training Behavior

(a)




(b)

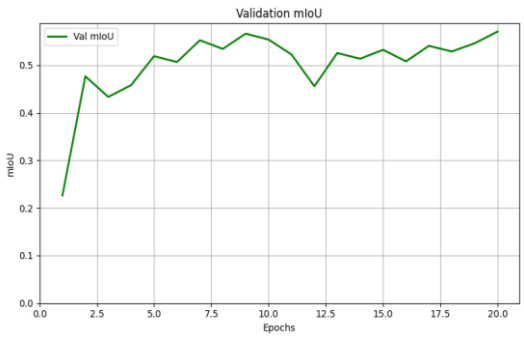



Figure 4: Training curves of the model showing (a) training and validation loss and (b) validation mIoU across training epochs.

The training process was monitored using loss curves and validation mIoU values across epochs. Figure 4 shows the training and validation loss curves along with the validation mIoU progression.

The curves indicate stable convergence during training. The training loss gradually decreases across epochs, while validation loss follows a similar trend, suggesting that the model effectively learns meaningful representations. Additionally, the validation mIoU improves steadily during the training process, confirming that the model generalizes well to unseen validation data.

The use of mixed precision training and the AdamW optimizer with cosine annealing warm restarts contributed to efficient optimization and improved training stability.

E. Discussion

The experimental results demonstrate that transformer-based architectures such as Mask2Former can achieve high performance for land cover classification tasks using satellite imagery. The integration of self-attention mechanisms allows the model to capture global context and long-range dependencies, which are often difficult to model using traditional convolutional neural networks.

Furthermore, the patch-based training strategy combined with sliding-window inference enables efficient processing of high-resolution satellite images without exceeding GPU memory limitations. Data augmentation techniques also improved the robustness of the model by exposing it to diverse variations in illumination and spatial orientation.

Overall, the proposed approach provides an effective framework for automated land cover classification and highlights the potential of transformer-based models for large-scale geospatial analysis.

## VI. CONCLUSION

This study presented a transformer-based framework for automated land cover classification from high-resolution satellite imagery. A semantic segmentation model based on Mask2Former with a Swin-Large backbone was employed to perform pixel-level classification of multiple land cover categories. The proposed approach incorporated patch-based training, data augmentation, and a sliding-window inference strategy to efficiently process large satellite images while preserving important spatial details.

Experimental results demonstrated that the model achieved a Pixel Accuracy of 0.8716 and a mean Intersection over Union of 0.5701, indicating reliable segmentation performance across several land cover classes. Strong performance was observed for classes such as agriculture, water, and urban areas, while visually similar categories such as rangeland and barren land remained more challenging.

Overall, the findings highlight the potential of transformer-based architectures for accurate and scalable land cover classification. Future work may explore larger datasets, improved class balancing strategies, and hybrid CNN-Transformer

architectures to further enhance segmentation performance.

## REFERENCES


[1] Y. Liu, K. Yang, Z. Peng, T. Zou, D. Su, R. Sun, and J. Ma, "Land use and land cover classification and terrestrial ecosystem carbon storage changes in Vietnam based on Sentinel images," Scientific Reports, vol. 15, no. 1, p. 22114, Jul. 2025. doi: 10.1038/s41598-025-04765-z.

[2] J. C. Campos, A. V. Liz, L. Patkó, A. Abdulkarem, L. Van Essen, M. El-Bana, A. Al-Ansari, O. Al-Attas, and J. C. Brito, "An optimised land-use land-cover classification approach for general application in deserts and arid regions," Science of Remote Sensing, vol. 12, p. 100334, 2025, doi: 10.1016/j.srs.2025.100334.

[3] D. Phiri and J. Morgenroth, "Developments in Landsat land cover classification methods: A review," Remote Sensing, vol. 9, no. 9, p. 967, 2017, doi: 10.3390/rs9090967.

[4] M. C. Hansen, R. S. DeFries, J. R. G. Townshend, and R. Sohlberg, "Global land cover classification at 1 km spatial resolution using a classification tree approach," International Journal of Remote Sensing, vol. 21, no. 6–7, pp. 1331–1364, 2000, doi: 10.1080/014311600210209.

[5] G. Amin, I. Imtiaz, E. Haroon, N. U. Saqib, M. I. Shahzad, and M. Nazeer, "Assessment of machine learning algorithms for land cover classification in a complex mountainous landscape," Journal of Geovisualization and Spatial Analysis, vol. 8, no. 2, p. 34, 2024, doi: 10.1007/s41651-024-00195-z.

[6] S. Talukdar, P. Singha, S. Mahato, Shahfahad, S. Pal, Y.-A. Liou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—A review," Remote Sensing, vol. 12, no. 7, p. 1135, 2020, doi: 10.3390/rs12071135.

[7] A. Vitale and F. Lamonaca, "Enhancing GeoAI land cover classification via hyperparameter tuning and cross-validation: A case study in Ravenna, Italy," Measurement, vol. 257, p. 118662, 2026, doi: 10.1016/j.measurement.2025.118662.

[8] E. Nikolaou-Alavanou, G. P. Petropoulos, and K. Kalogeropoulos, "Combining ENMAP hyperspectral imaging and machine learning for land use/cover classification," Remote Sensing in Earth Systems Sciences, vol. 9, no. 1, p. 19, 2026, doi: 10.1007/s41976-026-00271-6.

[9] L. Ma, X. Li, and J. Hou, "An inclusive classification optimization model for land use and land cover classification," Scientific Reports, vol. 15, no. 1, p. 9847, 2025, doi: 10.1038/s41598-025-91260-0.

[10] S. Alikhanova, C. Tarantino, and J. W. Bull, "Improving land cover classification in drylands with MSAVI: Evidence from the South Aral Seabed," Journal of Arid Land, vol. 18, no. 2, pp. 185–201, 2026, doi: 10.1016/j.jaridl.2026.02.001.

[11] M. Beak, K. Ichii, Y. Yamamoto, R. Wang, B. Zhang, R. C. Sharma, and T. Hiyama, "Land cover classification for Siberia leveraging diverse global land cover datasets," Progress in Earth and Planetary Science, vol. 12, no. 1, p. 3, 2025, doi: 10.1186/s40645-024-00672-5.

[12] Z. Huang, J. Cheng, G. Wei, X. Hua, and Y. Wang, "An urban land cover classification method based on segments' multidimension feature fusion," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 5580–5593, 2024, doi: 10.1109/JSTARS.2024.3367626.

[13] T. F. Deressu, A. K. Bojer, T. G. Debelee, W. G. Negera, S. Nadarajah, and K. W. Gebissa, "Enhancing land use and land cover classification with deep learning-based satellite imagery segmentation," International Journal of Applied Earth Observation and Geoinformation, vol. 144, p. 104839, 2025, doi: 10.1016/j.jag.2025.104839.

[14] S. Zhao, K. Tu, S. Ye, H. Tang, Y. Hu, and C. Xie, "Land use and land cover classification meets deep learning: A review," Sensors, vol. 23, no. 21, p. 8966, 2023, doi: 10.3390/s23218966.

[15] A. Temenos, N. Temenos, M. Kaselimi, A. Doulamis, and N. Doulamis, "Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP,"

IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3251652.

[16] X. Huang, X. Liu, Y. Liu, Y. Dai, and Z. Li, "Multiscale and fine-grained feature mining model for land use and land cover classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 26561–26572, 2025, doi: 10.1109/JSTARS.2025.3620291.

[17] H. M. Albarakati, M. A. Khan, A. Hamza, F. Khan, N. Kraiem, L. Jamel, L. Almuqren, and R. Alroobaea, "A novel deep learning architecture for agriculture land cover and land use classification from remote sensing images based on network-level fusion of self-attention architecture," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 6338–6353, 2024, doi: 10.1109/JSTARS.2024.3369950.

[18] M. Aljebreen, H. A. Mengash, M. Alamgeer, S. S. Alotaibi, A. S. Salama, and M. A. Hamza, "Land use and land cover classification using river formation dynamics algorithm with deep learning on remote sensing images," IEEE Access, vol. 12, pp. 11147–11156, 2024, doi: 10.1109/ACCESS.2023.3349285.

[19] J. Yuan, L. Ru, S. Wang, and C. Wu, "WH-MAVS: A novel dataset and deep learning benchmark for multiple land use and land cover applications," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 1575–1590, 2022, doi: 10.1109/JSTARS.2022.3142898.

[20] S. Sierra, R. Ramo, M. Padilla, and A. Cobo, "Optimizing deep neural networks for high-resolution land cover classification through data augmentation," Environmental Monitoring and Assessment, vol. 197, no. 4, p. 423, 2025, doi: 10.1007/s10661-025-13870-5.

[21] D. Hester, V. S. Martins, L. B. Ferreira, and T. M. A. Lima, "Learning with less: Label-efficient land cover classification at very high spatial resolution using self-supervised deep learning," Science of Remote Sensing, vol. 13, p. 100397, 2026, doi: 10.1016/j.srs.2026.100397.

[22] R. A. Al-Falluji and M. Ali Albahar, "OEF-LULC: An optimized and explainable AI-based framework for land use land cover classification," IEEE Access, vol. 14, pp. 38745–38757, 2026, doi: 10.1109/ACCESS.2026.3672676.

[23] V. Pushpalatha, P. B. Mallikarjuna, H. N. Mahendra, S. Rama Subramoniam, and S. Mallikarjunaswamy, "Land use and land cover classification for change detection studies using convolutional neural network," Applied Computing and Geosciences, vol. 25, p. 100227, 2025, doi: 10.1016/j.acags.2025.100227.