

Super-Resolution of Sentinel-2 Imagery for Accurate Forest Segmentation in Deforestation Detection

Khagan Ahad

Received August 27, 2025

Accepted December 4, 2025

Electronic access January 15, 2026

Accurate forest mapping is a critical remote sensing application that aims to delineate forest cover precisely, but is often limited by the native spatial resolution of widely available satellite imagery (e.g., Sentinel-2's 10m resolution). While deep learning segmentation models are commonly applied directly to such moderate-resolution data, few systematically investigate Super-Resolution (SR) pre-processing to enhance the input. In this study, we integrate an SR step to enhance Sentinel-2 imagery from 10m to 2.5m effective resolution before forest segmentation. We evaluate paired (Sentinel-2/SPOT) and unpaired (BSRGAN-based degradation simulation) data strategies for training the SWINIR SR model, finding that the unpaired approach yields significantly better quantitative results (PSNR 23.46 vs 19.34) and visual quality by avoiding artefacts. Subsequently, a U-Net segmentation model with an EfficientNet encoder is trained on the resulting 2.5m effective resolution imagery, using labels meticulously derived from this enhanced data and verified against high-resolution references. We conduct experiments using data covering diverse geographical regions within Azerbaijan. Experimental results on a hold-out validation set demonstrate that our segmentation approach achieves high performance with 95.8% accuracy and a 0.91 F1-score for the forest class. This indicates that leveraging SR pre-processing significantly improves the potential for detailed and accurate forest mapping from Sentinel-2 data, thereby enhancing boundary delineation and the detection of small forest features that are often missed at native resolution.

Keywords: Deep Learning, Machine Learning, Neural Networks, Satellite Images, Super-Resolution, Transformers

Introduction

The Copernicus Sentinel-2 mission provides a valuable resource for comprehensive global forest monitoring. It offers high spatial resolution multispectral imagery with a commendable revisit frequency. Sentinel-2 provides a rare combination of global reach, multispectral depth, and five-day revisit frequency. Despite advantages of Sentinel-2, its native 10 m spatial resolution imposes a critical bottleneck when forest inventories must resolve detail at the sub-hectare scale. Features smaller than a 10×10 m ground cell (e.g. logging skid trails, narrow riparian buffers, or small canopy gaps) collapse into “mixed pixels” whose spectral signatures blend forest, soil, water, and built surfaces. Along ecotonal boundaries, this spectral amalgamation erodes class separability, forcing segmentation networks to predict unnaturally smooth edges and masking the subtle spectral shifts that mark early-stage degradation. Small-area disturbances, selective-logging corridors, and micro-fragments can therefore evade detection or be misclassified as homogeneous forest.

To overcome these constraints, we embed a super-resolution (SR) transformation in the pre-processing pipeline, enlarging Sentinel-2 inputs from 10 m to an effective 2.5 m grid ($4\times$ enhancement). The SR network¹ learns cross-scale pri-

ors that reconstruct plausible high-frequency structure latent in the source imagery, producing outputs that are spectrally consistent with native high-resolution sensors while preserving radiometric integrity.

This resolution uplift serves two mission-critical objectives. First, boundary fidelity: disaggregating one mixed edge pixel into sixteen finer cells supplies the downstream U-Net² with spatially purer inputs, enabling the network to learn (and subsequently predict) sharper, topologically faithful forest/non-forest transitions. Second, micro-feature visibility: clear-cuts, skidder tracks, narrow shelterbelts, and isolated fragments now span several pixels, elevating their signal-to-noise ratio and improving the likelihood of correct classification. While super-resolution preprocessing shows promise for enhancing forest monitoring capabilities, the field lacks systematic evaluation of different SR approaches and their actual impact on downstream segmentation performance.

Critical gaps still remain despite growing interest in SR for remote sensing. First, transformer-based architectures like SwinIR¹, which excel at capturing long-range dependencies have received limited attention for operational forest monitoring, with GANs³ and CNNs dominating the literature. Second, paired versus unpaired SR training strategies remain poorly quantified for segmentation tasks. Third, few studies

rigorously compare SR-enhanced models against well-tuned native-resolution baselines to isolate SR's true contribution.

This study addresses these gaps by enhancing Sentinel-2 imagery from 10m to 2.5m resolution using transformer-based SR before forest segmentation in Azerbaijan's diverse landscapes. We make four principal contributions: (i) the first comprehensive evaluation of SwinIR for Sentinel-2 forest mapping, benchmarked against ESRGAN⁴ and HAT⁵; (ii) direct comparison of paired (Sentinel-2/SPOT) versus unpaired (BSRGAN-simulated) training strategies, demonstrating that the unpaired approach significantly outperforms paired training (PSNR (Peak Signal-to-Noise Ratio) 23.46 vs 19.34) by avoiding artifacts and geographic overfitting; (iii) rigorous controlled comparison of our SR-enhanced U-Net against an identical baseline trained on native 10m imagery, both evaluated on the same test set; and (iv) demonstration across Azerbaijan's underrepresented ecosystems, extending SR-based forest monitoring beyond well-studied tropical regions. To rigorously validate these contributions and quantify the actual benefits of super-resolution preprocessing, we designed a controlled experimental framework.

To evaluate the impact of the super-resolution preprocessing step, a direct comparison will be performed against a baseline model trained under identical conditions but using the original 10m resolution Sentinel-2 imagery. Specifically, a second U-Net model, mirroring the same EfficientNet encoder architecture, training parameters, and data splits, will be trained directly on the 10m input data paired with appropriately downsampled ground truth labels. Both the super-resolution-trained model and the baseline 10 m-trained model will then be evaluated on the same hold-out test set. For pixel-wise metric calculation against the 2.5m resolution ground truth, the output predictions from the baseline 10m model will be upsampled to 2.5m using nearest-neighbor interpolation. Performance will be compared using key metrics, including overall accuracy and mean Intersection over Union (mIoU). Furthermore, qualitative visual comparisons will be conducted, focusing specifically on the models' abilities to delineate complex forest edges, capture small forest patches or clearings, and handle fragmented landscapes, thereby directly assessing the practical benefits conferred by the super-resolution enhancement.

In essence, super-resolution functions as a task-oriented data-enrichment layer that offsets Sentinel-2's spatial deficit. By injecting sub-pixel detail and sharpening boundary cues before learning commences, the pipeline targets three downstream gains: (i) higher-precision forest-area estimates, (ii) more reliable quantification of fragmentation metrics, and (iii) heightened sensitivity to incipient deforestation and canopy-thinning events (Figure 1).

Related Work

Super Resolution Techniques

The aim of Super-resolution techniques is to create a high-resolution (HR) image from low-resolution (LR) images. Better spatial resolution results in better distinction of the boundaries of deforestation patches, better ability to distinguish between different types of forest disturbances, and a reduction in the number of false positives and false negatives. However, these benefits must be weighed against computational trade-offs. Super-resolution processing adds computational overhead, both in terms of processing time and resource requirements. Some SR techniques might introduce artefacts or slightly alter the spectral characteristics of the imagery⁶. In order to determine if the improvement in deforestation detection accuracy can justify the additional complexity and potential effects on spectral fidelity, a careful evaluation is required. This is especially true for applications that depend on accurate spectral information for in-depth analysis.

Various super-resolution methods have emerged to enhance Sentinel-2 imagery from 20-60m to 10m or better resolution, ranging from channel attention mechanisms to generative adversarial networks and transfer learning approaches. These methods can be broadly categorized based on their architectural paradigms and optimization objectives. The trade-offs between computational efficiency and image quality remain evident. This can be seen from comparison of GANs and CNN-based methods where GANs achieving superior perceptual realism (LPIPS=0.28) while CNN-based methods maintain better radiometric accuracy for quantitative forest analysis.⁷⁻¹⁰

SR methods based on GANs perform well at generating realistic high-resolution images. This is especially important for tasks that require attention to small details. Although they are good at creating HR images, they might sometimes create artefacts or not perfectly preserve the spectral fidelity of the original data⁶. Within the GAN family of methods, SRGAN (Super-Resolution Generative Adversarial Network) was a pioneering work in applying GANs for photo-realistic single image super-resolution, particularly for large upscaling factors¹¹. A key contribution of SRGAN was the introduction of a perceptual loss function, which combines an adversarial loss (based on the discriminator's output) with a content loss calculated on high-level feature maps of a pre-trained VGG network. This perceptual loss encourages the generator to produce super-resolved images that are not only pixel-wise similar to the ground truth but also perceptually realistic, with finer texture details. ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) was subsequently developed to address some of the artefacts produced by SRGAN and to generate visually even better high-resolution images¹².

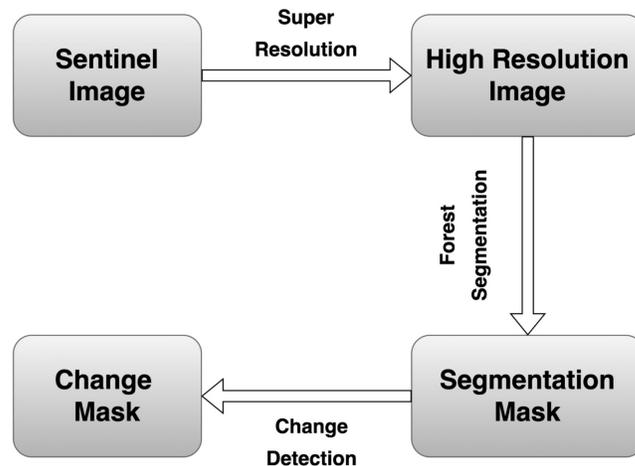


Fig. 1 General pipeline for detecting change mask using images with Super Resolution

Recognizing the unique requirements of remote sensing applications, RS-ESRGAN is an addition of ESRGAN, created specifically for remote sensing imagery. Improvements include removing the upsampling layers in the generator which helps facilitate training with co-registered WorldView and Sentinel-2¹² image pairs. This allows the model to learn the mapping between lower-resolution Sentinel-2 and higher-resolution WorldView imagery. These changes allow RS-ESRGAN to have better performance on Sentinel-2 imagery that can be observed both visually and with quantitative metrics, and it preserves spectral information⁶.

In contrast to purely GAN-based approaches, hybrid networks that combine different types of neural network architectures have also demonstrated potential in remote sensing image super-resolution. These methods aim to leverage complementary strengths of different architectures. EHNNet (Efficient Super-Resolution Hybrid Network) uniquely combines a CNN-based encoder with a Swin Transformer-based decoder in a UNet-like structure for efficient remote sensing image super-resolution¹³. This hybrid method intends to leverage the inductive biases of convolutional layers for local feature extraction and the long-range modelling capabilities of transformer networks.

Similarly, SARNet⁸ (Spectral Attention Residual Network) is another example of a hybrid network designed specifically for the super-resolution of Sentinel-2 satellite images⁸. SARNet incorporates Residual Channel Attention Blocks (RCABs) within a residual learning framework, enabling the network to focus on the most informative spectral channels by exploiting interdependencies among them. This spectral attention mechanism is particularly advantageous for multispectral imagery such as Sentinel-2. SARNet is trained and evalu-

ated using real-world pairs of PlanetScope (high-resolution) and Sentinel-2 (low-resolution) images, addressing the challenge of super-resolving real-world satellite imagery where ideal ground truth is often unavailable¹⁴.

Transformer-based models can be used for both deforestation detection and image super-resolution. A prime example for that is SwinIR¹, which utilises Swin Transformer architecture for image restoration tasks. SwinIR employs Residual Swin Transformer Blocks (RSTBs) for deep feature extraction. Each RSTB contains multiple Swin Transformer Layers (STLs). The STLs are based on a shifted window partitioning scheme, allowing for the capture of both local and long-range dependencies in the image. This ability is important for effective super-resolution.

Deforestation Detection techniques

Convolutional Neural Networks (CNNs) are a popular choice for distinguishing deforested areas from intact forests¹⁵ since they are very effective in learning hierarchical features from image data. Several studies have reported that high accuracy rates for deforestation detection tasks were achieved using CNN-based approaches¹⁵.

While CNNs provide strong baseline performance, the UNet architecture, with its distinctive encoder-decoder structure and skip connections, has gained immense popularity for image segmentation tasks, including the precise delineation of deforested areas in satellite imagery¹⁵. The encoder path progressively downsamples the input image, learning a compressed representation of its features, while the decoder path upsamples these features to produce a segmentation map at the original resolution.

Skipping connections between corresponding layers in the encoder and decoder helps to preserve fine-grained spatial information, which is important for accurate boundary delineation in deforestation mapping. Various modifications and improvements to the basic U-Net architecture have been proposed to further boost its efficiency in deforestation mapping, such as the integration of attention mechanisms or the use of different backbone networks¹⁶. Studies utilising U-Net with Sentinel-2 data have shown its value in mapping forest cover, detecting forest cover changes over time, and identifying specific instances of deforestation^{16–18}.

Moving beyond convolutional architectures, in addition to CNNs and U-Net variants, transformer-based networks have recently been explored for deforestation detection, offering a different methodology for examining satellite imagery. Transformer networks, originally developed for natural language processing, excel at capturing long-range dependencies in sequential data. When adapted for image analysis, such as in the ChangeFormer architecture, they can effectively analyse spatial and temporal patterns in bi-temporal satellite imagery through the use of attention mechanisms¹⁹.

These mechanisms allow the model to weigh the importance of different parts of the image when making a prediction, potentially improving its ability to detect subtle changes associated with deforestation. A study applying ChangeFormer to deforestation detection in the Brazilian Amazon using Sentinel-2 imagery achieved an overall accuracy of 93%, with a corresponding F1 score of 90% and an Intersection over Union (IoU) score of 82%, demonstrating the potential of transformer-based networks for this task¹⁹.

The impact of spatial resolution on detection accuracy has been a recurring theme across these architectural approaches. While there are many studies on deforestation mapping with native resolution images^{14,20–23}, several studies have investigated the direct impact of applying super-resolution techniques to Sentinel-2 imagery on the accuracy of deep learning models for deforestation detection.

Research findings generally suggest that improving the spatial resolution of Sentinel-2 data can result in higher detection accuracy²⁴. One study concluded that the higher resolution of Sentinel-2 images, even without further super-resolution, improves the segmentation of deforestation polygons both quantitatively (in terms of F1-score) and qualitatively compared to lower-resolution imagery like Landsat-8²⁵. This benefit of Sentinel-2's 10-meter resolution over coarser data shows the potential for further gains through super-resolution.

Further evidence for the importance of spatial resolution comes from comparative studies: comparison between deforestation detection accuracy using native resolution Sentinel-2 imagery and higher-resolution imagery like PlanetScope (with a resolution of 3-4 meters) also shows the advantages of increased spatial detail²⁶. One experiment showed that Plan-

etScope imagery provided a higher quality of segmentation for deforestation detection compared to Sentinel-2 data, suggesting that finer spatial resolution allows for better discrimination of deforested areas²⁶. This observation gives a credible reason for using super-resolution techniques to bridge the resolution gap between Sentinel-2 and higher-resolution sources.

Underpinning all these methodological advances is the critical role of data and evaluation. There are a lot of datasets that are used for training and evaluating these methods. These datasets vary in terms of geographic location, the type of deforestation being studied, and the availability of high-quality ground truth data. A model's capability to generalise strongly depends on the quality of the datasets. Chosen evaluation metrics (e.g., overall accuracy, F1-score, IoU, PSNR, SSIM (Structural Similarity Index Measure)) are equally important to dataset choice in order to correctly assess the performance of both super-resolution and deforestation detection models.

Methods

Data Acquisition and Preparation

The study area within Azerbaijan covers approximately ~31,000 km². We specifically selected diverse geographical and ecological regions across the country to ensure the model's generalisability. These regions encompass a variety of forest types, including temperate broadleaf and mixed forests in the Greater Caucasus mountains, subtropical lowland forests, and riparian forests. This diversity in topography, forest density, and species composition provides a robust testbed that represents a challenging range of conditions relevant for forest monitoring in many temperate and subtropical regions.

Our data acquisition protocol involved a pre-selection step where we only downloaded Sentinel-2 L2A scenes (level 2A - bottom of atmosphere) with less than 5% cloud cover. During the manual labeling phase, any remaining small clouds or topographic shadows were manually masked out and excluded from both the training and evaluation datasets. This ensures the model was trained on clean, high-quality data.

To prevent any data leakage and ensure a truly independent test set, a geographical split was implemented. The Sentinel-2 scenes were divided into training, validation, and testing groups. All tiles from a given scene belong exclusively to one set. This ensures that the model is evaluated on entirely unseen geographical areas, providing a more robust estimate of its generalization performance.

Low-resolution (LR) input data for the SR task consisted of multispectral images acquired from the Sentinel-2 satellite constellation (10m native resolution bands). High-resolution (HR) reference data were sourced from the SPOT satellite constellation (native 1.5m GSD). Since the task is 4x super-

resolution to generate the 2.5m effective resolution imagery required for detailed segmentation, two super-resolution training strategies were investigated: paired and unpaired. The paired dataset comprised 15000 Sentinel-2/SPOT tiled pairs covering approximately 25000 km². These pairs were sourced from different eco-regions (e.g., montane forests, riparian corridors, agricultural-forest mosaics) within Azerbaijan. While this provided some landscape variety, the limited number and geographic scope constrained the representativeness of the paired model compared to the unpaired dataset, which benefitted from synthetic diversity.

Paired Data Approach: This method aimed to create direct LR-HR pairs. SPOT 1.5m images were downsampled to 2.5m using bilinear resampling to serve as the HR target. Near-temporal Sentinel-2 images were selected as potential LR counterparts. Pairs underwent visual inspection for landscape similarity, followed by co-registration for spatial alignment and histogram matching for radiometric normalisation. This process yielded a dataset of approximately 100 GB. Obtaining sufficient quantities of perfectly aligned, contemporaneous, and radiometrically consistent pairs across diverse landscapes proved difficult.

Unpaired Data Approach: This method circumvented the need for direct pairing. SPOT imagery, resampled to 2.5m, constituted the HR domain. To create a realistic LR domain mimicking Sentinel-2, the 2.5m SPOT images were artificially degraded using a process inspired by BSRGAN²⁷, involving simulated noise, blurring, and downsampling to 10m resolution.

Segmentation Dataset Preparation

The 2.5m effective resolution images generated by the selected unpaired SWINIR model served as the input for segmentation.

Labelling: Ground truth labels designating ‘forest’ and ‘non-forest’ pixels were created using QGIS software based on visual interpretation of the super-resolved 2.5m imagery. This allowed for more precise boundary delineation compared to using native 10m data. The generated labels were rigorously verified by cross-referencing them against independent high-resolution (HR) reference imagery, primarily the original high-resolution SPOT data, ensuring label accuracy, especially for fine details enhanced by SR. To evaluate the consistency of the manually created forest/non-forest masks, a subset of image tiles was independently labeled by three annotators. Across the sample, agreement was generally high, with only minor differences observed at forest boundary regions where canopy edges are less distinct. In these cases, the final label was resolved through consensus discussion. This step provided confidence that the labeling process was reliable overall, while also highlighting the areas where natural

ambiguity in land-cover transitions can occur.

The large super-resolved images and corresponding label masks were divided into 50,000 non-overlapping tiles of 256×256 pixels using a custom script. Each tile represents a ground area of 640m × 640m. The dataset was divided into training, validation, and testing sets (e.g., following a 70%/15%/15% ratio) for model development, hyperparameter tuning, and final unbiased evaluation. The pixel values within each image tile were standardised (normalised) by subtracting the mean and dividing by the standard deviation, calculated across the training dataset, to achieve zero mean and unit variance.

Super-Resolution Model and Training

Three leading SR backbones ESRGAN⁴, HAT⁵ and SwinIR¹ were benchmarked on a pilot subset. We selected SwinIR-L after a preliminary benchmark against ESRGAN and HAT. Expert visual review revealed that SwinIR-L consistently produced fewer artefacts and more realistic high-frequency details. This was especially clear along complex forest boundaries and narrow canopy gaps features that are critical for reliable segmentation. While ESRGAN often yielded perceptually sharp results, it was more prone to hallucinated textures. On the other hand, HAT required larger training datasets to reach its potential and tended to generate oversharpened outputs with speckled or serrated artefacts, which reduced their reliability in ecological mapping tasks.

From an operational perspective, SwinIR-L proved significantly more practical: it trained stably, maintained predictable memory usage, and supported tile-wise inference on a single prosumer GPU, enabling large-area processing without reliance on specialized hardware. Overall, SwinIR-L provided the most effective balance of benchmark accuracy, spectral and textural fidelity, and computational efficiency, making it the most suitable choice for routine large-scale mapping.

A lightweight U-Net discriminator enforced texture realism in the adversarial phase. Training progressed in three stages. (i) Pixel pre-training: the generator was optimised with mean-squared-error loss to initialise content fidelity. (ii) Adversarial refinement: GAN training introduced an adversarial loss, an L1 content loss, and a perceptual loss computed on VGG features. (iii) Metric-oriented fine-tuning: a final pass with pure L2 loss minimised residual artefacts and boosted PSNR/SSIM. Both paired and unpaired datasets followed the same curriculum, differing only in their LR–HR sampling.

The generator was configured as follows:

Training the adopted SwinIR-L model on an NVIDIA RTX A6000 (48 GB VRAM) required approximately 120 GPU hours, with peak memory usage around 34 GB. Inference was considerably lighter, averaging 0.6 seconds per 256×256 tile (~640×640 m) with a memory footprint of ~8 GB, corre-

Table 1 Parameters for Super Resolution training

Parameter	Value
Image size	64
Window size	8
Embedding dimension	240
Depths	[6, 6, 6, 6, 6, 6, 6, 6]
Number of heads	[8, 8, 8, 8, 8, 8, 8, 8]
MLP ratio	2
Upsampler	nearest + conv
Residual connection	3 conv

sponding to ~ 1.4 megapixels per second throughput. A full Sentinel-2 scene of $\sim 100 \text{ km}^2$ can therefore be processed in roughly 3–4 GPU hours on a single A6000. These values indicate that, although training is resource-intensive, inference is tractable for operational deforestation monitoring, particularly when parallelised on multi-GPU or cloud infrastructures.

Models from the paired and unpaired pipelines were compared on the held-out test tiles using standard quality indices (Table 1) and blind visual ranking by three remote-sensing experts. The unpaired SwinIR-L checkpoint delivered the best trade-off between edge sharpness, spectral faithfulness, and quantitative metrics, and was consequently selected for downstream forest-segmentation experiments.

Table 2 Super-Resolution Model Performance Metrics (on held-out test data)

Training Approach	PSNR	SSIM
Unpaired (BSRGAN-simulated LR)	23.46	0.67
Paired (SPOT-Sentinel)	19.34	0.40

As indicated by the quantitative metrics (Table 2) and visual inspection, the SWINIR model trained using the unpaired approach (leveraging BSRGAN-simulated degradation) achieved significantly better performance. Its paired-data counterpart, in contrast, exhibits edge ringing, checkerboard noise, and conspicuous “hallucinations” over spectrally complex classes—most often urban fabric and open water (Figure 2). We attribute this deficit to (i) the scarcity of perfectly coregistered Sentinel-2/SPOT pairs, which forces the network to overfit a geographically narrow domain, and (ii) class imbalance in the paired archive, whereby high-contrast built-up and hydrographic features are under-represented. By decoupling LR and HR domains, the unpaired route circumvents these biases, regularises the generator via diverse synthetic degradations, and ultimately yields a model better generalised to real-world Sentinel-2 inputs.

Consequently, the SWINIR model trained via the unpaired



Fig. 2 Super Resolution results for both Paired and Unpaired data. Hallucinations caused by the paired model can be observed in second image

method was selected to generate the 2.5m super-resolved imagery used for the subsequent forest segmentation task (Figure 3).



Fig. 3 Difference between original Unpaired Low Resolution images and images with Super Resolution applied

Forest segmentation and change detection

We chose the U-Net² architecture for our semantic segmentation task due to its effective encoder-decoder structure and skip connections, which help in capturing detailed spatial information. EfficientNet²⁸ was selected as the encoder backbone, as it provides a great balance between computational efficiency and accuracy. To speed up training and enhance the model’s feature representation capabilities, we initialised the encoder with pre-trained weights from ImageNet. The Adam optimizer²⁹ was used with a constant learning rate of 5×10^{-4} , as it adapts well during the learning process. To distinguish between forest and non-forest areas at the pixel level, we applied Binary Cross-Entropy (BCE) loss. The model was trained for 50 epochs, with each epoch covering a complete pass through the training data.

Application to Forest Change Detection

Beyond generating static forest cover maps, the developed high-resolution segmentation model provides a powerful foundation for monitoring forest dynamics over time, specifically for change detection tasks such as identifying deforestation or afforestation. A post-classification comparison approach can be implemented using the trained segmentation model:

Temporal Image Selection. Two Sentinel-2 Level-2A scenes (T_1 and T_2) are selected such that (i) both fall within the same seasonal phenophase to reduce phenological noise, (ii) cumulative cloud/shadow coverage is $< 5\%$ over the AOI.

Pre-processing and Super-Resolution. Both the T_1 and T_2 Sentinel-2 images undergo the same pre-processing steps, the application of the trained SWINIR super-resolution model to enhance their effective spatial resolution to 2.5m.

Sub-pixel co-registration. Precise spatial alignment between the super-resolved T_1 and T_2 images is essential for accurate change detection. Semi-automated image co-registration techniques are applied to ensure that corresponding pixels in both images represent the same ground location.

Independent segmentation and uncertainty masking. The EfficientNet-U-Net predicts binary masks M_1 and M_2 . Softmax probabilities are retained; a pixel is committed to ‘forest’ only if $p \geq 0.55$. To examine robustness to threshold selection, we computed pixel-wise precision, recall, and F1-score across a range of probability thresholds ($p = 0.40\text{--}0.70$) on the same validation set. As expected, recall declined while precision increased with higher thresholds. Importantly, the F1-score remained stable ($\approx 0.90\text{--}0.91$) for thresholds between 0.50 and 0.60 (Figure 4), demonstrating that the model’s performance is not overly sensitive to the exact cutoff point. This stability supports the use of $p = 0.55$ as a balanced operating point, while also providing assurance that small variations in the decision rule would not affect the change detection outcomes considerably.

Post-classification comparison & transition logic A four-state transition matrix is computed: stable forest, stable non-forest, deforestation ($M_1 = 1 \rightarrow M_2 = 0$), and afforestation/regrowth ($M_1 = 0 \rightarrow M_2 = 1$). To suppress salt-and-pepper noise, connected-component filtering eliminates patches smaller than 3 contiguous pixels (≈ 0.005 ha).

Results

The performance of the trained forest segmentation model was evaluated on the hold-out validation dataset, comprising super-resolved 2.5m image tiles and ground truth labels unseen during training. After 50 epochs of training, the U-Net model with the EfficientNet encoder achieved the following key performance metrics on the test set (Table 3):

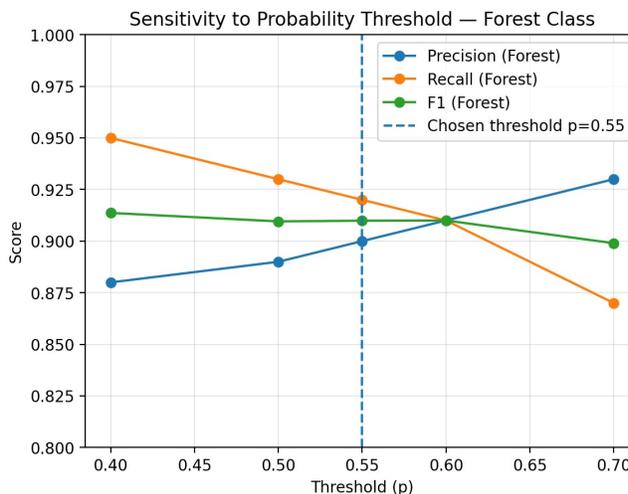


Fig. 4 Sensitivity to Probability threshold for forest class

Table 3 Segmentation metrics

Metric	Value
Test Accuracy	95.8%
Test F1-Score (Forest Class)	91%

These results indicate a high overall pixel classification accuracy and, more importantly, a strong performance in accurately identifying and delineating forest areas, as reflected by the high F1-score. This suggests that the combined approach of using super-resolved imagery and the chosen deep learning architecture is effective for detailed forest mapping in this context (Table 4).

Table 4 Per class Segmentation metrics

	Forest (positive class)	Non-Forest
Precision	0.900	0.976
Recall	0.920	0.969
F1	0.910	0.973

The resulting change map highlights areas of forest transition. The use of 2.5m super-resolved imagery in this process allows for the detection of finer-scale changes compared to using native 10m resolution data, potentially revealing smaller clearing events or initial stages of regeneration that might otherwise be missed (Table 5). This methodology provides a valuable tool for quantifying forest cover changes across Azerbaijan, supporting efforts in monitoring deforestation rates, assessing reforestation projects, and understanding land cover dynamics.

Table 5 Change detection metric comparison

Image	Accuracy	IOU	F1
10m resolution images	0.932	0.795	0.843
Images using SR as a preprocessing step	0.958	0.856	0.896

Images in Figures 5, 6 and 7 demonstrate the performance difference between Sentinel and high resolution; the first one is the result of the original image, and the second one is processed using a super-resolution algorithm.

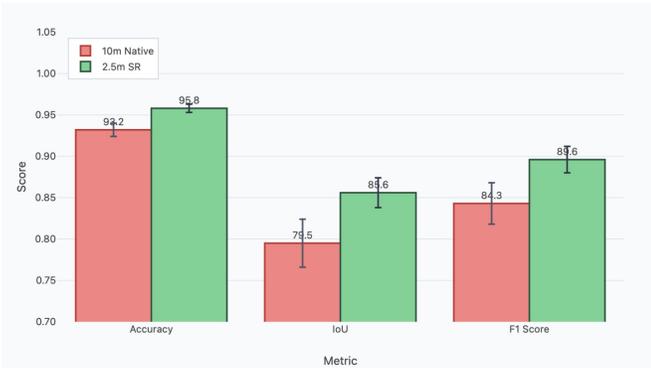


Fig. 5 Comparative performance of change-detection metrics (10m vs 2.5m SR) with 95% CI

The integration of the super-resolution step, enhancing Sentinel-2 imagery to an effective 2.5m resolution, significantly impacts the accuracy and detail of the forest change detection process compared to performing the analysis directly on native 10m resolution data.

We achieved a substantial increase of 6% IOU in change detection after applying the super-resolution step. By providing a finer-grained 2.5m representation, SR reduces pixel mixing, leading to sharper boundary delineation in the segmentation maps at both T_1 and T_2 . Consequently, the comparison between these maps is more likely to reflect true boundary shifts rather than noise or ambiguity inherent in the coarser resolution data, thereby increasing the geometric accuracy of detected changes (Figure 6).

While SR offers significant advantages, its use also introduces considerations: Any artefacts or inconsistencies introduced by the SR algorithm itself could potentially be misinterpreted as a change if they differ between the T_1 and T_2 processed images (Figure 7). Consistent application and validation of the SR model are necessary to minimise this risk. The heightened sensitivity might detect very minor or ephemeral changes (e.g., seasonal canopy variations misinterpreted as gain/loss if images are not carefully selected from similar phenological periods) that may need to be filtered out depending on the specific definition of significant change (e.g., by apply-

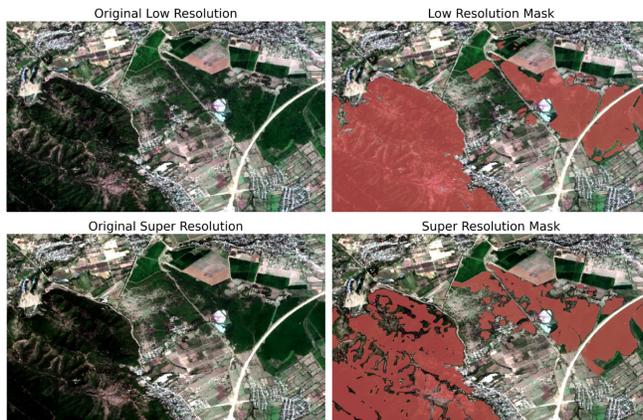


Fig. 6 Comparison between the mask quality of Super Resolution and Low Resolution satellite images

ing a minimum mapping unit threshold to the change results).



Fig. 7 Artefacts created during Super Resolution

In summary, leveraging super-resolved imagery within a post-classification change detection framework is expected to yield more accurate and detailed results, particularly in capturing fine-scale forest dynamics and improving the precision of boundary change localisation.



Fig. 8 Detection of non-forest area in Low Resolution and Super Resolution images

Discussion

This study demonstrates the efficacy of integrating super-resolution as a pre-processing technique to enhance the utility of Sentinel-2 imagery for detailed forest segmentation. By computationally increasing the effective spatial resolution to

2.5m using a SWINIR model trained on unpaired data, we provided the subsequent U-Net segmentation network with richer spatial information, particularly concerning forest boundaries and small-scale features. The resulting model achieved high accuracy (0.958) and a robust F1-score (0.91) for the forest class on the held-out test set, indicating a significant improvement in the ability to accurately map forest extent and structure. While the super-resolution process itself presents ongoing challenges and opportunities for further refinement, particularly in generating highly faithful high-frequency details without artefacts (Figure 9), our results indicate that even the current state-of-the-art SR methods can yield tangible benefits. This approach offers a promising pathway to leverage freely available Sentinel-2 data for more granular forest monitoring tasks previously hindered by native resolution constraints, although continued research into optimising the super-resolution component is warranted to unlock its full potential.

The practical implications of these findings extend to operational forest monitoring systems currently deployed at a global scale. While global monitoring systems such as Global Forest Watch rely on Sentinel-2's 10 m native resolution, our findings demonstrate that super-resolution preprocessing can bridge much of the gap to commercial high-resolution imagery while retaining the global accessibility of free Sentinel-2 data. This enables finer-scale detection of forest loss and boundary changes than current operational platforms, offering a practical pathway to enhance existing global monitoring frameworks with sub-hectare precision. Such improvements are particularly valuable for detecting small-scale deforestation events, illegal logging activities, and subtle forest degradation that may be missed at coarser resolutions.

An important consideration for the broader application of this methodology concerns its generalisability across diverse geographical contexts. While the model was trained exclusively on data from Azerbaijan, the diversity of the forest ecosystems included (from mountainous mixed forests to subtropical lowlands) provides a degree of robustness. We hypothesise that the model has learned fundamental features of forest structure (e.g., texture, edges, canopy patterns) that are transferable to other temperate and subtropical ecosystems. However, its performance in vastly different environments, such as dense tropical rainforests or sparse boreal forests, would likely require region-specific fine-tuning. We frame this as a key area for future investigation to assess the model's external validity on a global scale.

Beyond geographical generalisability, the primary focus of this research was to validate the effectiveness of SR preprocessing for the specific and critical task of delineating forest extent to monitor deforestation, which motivated our binary classification approach (forest vs. non-forest). However, we acknowledge that extending this framework to multi-

class segmentation scenarios, such as distinguishing between different forest types, shrubland, grassland, and agricultural land would demonstrate the broader applicability of the proposed methodology. Such extensions would not only test the robustness of the SR-enhanced segmentation pipeline across more complex classification schemes but would also provide valuable insights for diverse land cover mapping applications. This represents an important direction for future research, where the interplay between super-resolution fidelity and the discriminative capacity required for fine-grained land cover classes can be systematically explored.

Acknowledgements

I wish to express my sincere gratitude to Umit Tigrak, Hamid Askarov, and Asad Dashdamirli. Their insightful discussions, valuable feedback, and dedicated support were crucial throughout the various stages of this research, from conceptualisation to manuscript preparation. I deeply appreciate their contributions, which significantly enhanced the quality and rigour of this study.

References

- 1 J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool and R. Timofte, *Swinir: Image restoration using swin transformer*, 2021.
- 2 O. Ronneberger, P. Fischer and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, 2015.
- 3 I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, *Generative adversarial nets*, 2014.
- 4 X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao and C. C. Loy, *ESRGAN: Enhanced super-resolution generative adversarial networks*, 2018.
- 5 X. Chen, X. Wang, W. Zhang, X. Kong, Y. Qiao, J. Zhou and C. Dong, *HAT: Hybrid attention transformer for image restoration*, 2023.
- 6 Y. Li, Y. Wang, B. Li and S. Wu, *Super-resolution of remote sensing images for $\times 4$ resolution without reference images*, 2022.
- 7 X. Zhu, Y. Xu and Z. Wei, *Super-resolution of sentinel-2 images based on deep channel-attention residual network*, 2019.
- 8 M. Beaulieu, K. Foucher and C. Haberman, *Deep image-to-image transfer applied to resolution enhancement of sentinel-2 images*, 2018.
- 9 K. Zhang, G. Sumbul and B. Demir, *An approach to super-resolution of sentinel-2 images based on generative adversarial networks*, 2020.
- 10 C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias and K. Schindler, *Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network*, 2018.
- 11 C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang and W. Shi, *Photo-realistic single image super-resolution using a generative adversarial network*, 2017.
- 12 P. Kramer, A. Steinhardt and B. Pedretsch, *Enhancing sentinel-2 image resolution: Evaluating advanced techniques based on convolutional and generative neural networks*, 2024.
- 13 W. Zhang, Z. Tan, Q. Lv, J. Li, B. Zhu and Y. Liu, *An efficient hybrid CNN-transformer approach for remote sensing super-resolution*, 2024.
- 14 M. Zabalza and A. Bernardini, *Super-resolution of sentinel-2 images using a spectral attention mechanism*, 2022.
- 15 I. Md Jelas, M. A. Zulkifley, M. Abdullah and M. Spraggon, *Deforesta-*

-
- tion detection using deep learning-based semantic segmentation techniques: A systematic review, 2024.
- 16 D. Lee and Y. Choi, *A learning strategy for amazon deforestation estimations using multi-modal satellite imagery*, 2023.
 - 17 J. V. Solórzano, A. Mas, J. Gao and J. L. Gallardo-Cruz, *Land use land cover classification with u-net: Advantages of combining sentinel-1 and sentinel-2 imagery*, 2021.
 - 18 H. Zhang, J. Lin, P. Wang, H. Zhang and B. Ge, *Automated delineation of agricultural field boundaries from sentinel-2 images using recurrent residual u-net*, 2021.
 - 19 M. Alshehri, A. Ouadou and G. J. Scott, *Deep transformer-based network deforestation detection in the brazilian amazon using sentinel-2 imagery*, 2024.
 - 20 A. U. Waldeland, D. Trier and A. B. Salberg, *Forest mapping and monitoring in africa using sentinel-2 data and deep learning*, 2022.
 - 21 A. M. Pacheco-Pascagaza, Y. Gou, G. Louis, J. Roberts, P. Rodriguez-Gonzalez, F. Holguin, O. Gonzalez and G. Uribe, *Near real-time change detection system using sentinel-2 and machine learning: A test for mexican and colombian forests*, 2022.
 - 22 M. Szostak, P. Hawryło and D. Piela, *Using of sentinel-2 images for automation of the forest succession detection*, 2018.
 - 23 K. Isaienkov, M. Yushchuk, V. Khramtsov and O. Seliverstov, *Deep learning for regular change detection in ukrainian forest ecosystem with sentinel-2*, 2020.
 - 24 T. A. Lima, R. Beuchle, A. Langner, R. C. Grecchi, V. C. Griess and F. Achard, *Comparing sentinel-2 MSI and landsat 8 OLI imagery for monitoring selective logging in the brazilian amazon*, 2019.
 - 25 D. L. Torres, J. N. Turnes, P. J. Soto Vega, R. Q. Feitosa, D. E. Silva, J. Marcato Junior and C. Almeida, *Deforestation detection with fully convolutional networks in the amazon forest from landsat-8 and sentinel-2 images*, 2021.
 - 26 M. Gašparović, D. Medak, I. Pilaš, L. Jurjević and I. Balenović, *Fusion of sentinel-2 and planetscope imagery for vegetation detection and monitoring*, 2018.
 - 27 F. B. de Lima and E. G. da Silva, *Enhancing text recognition in OCR systems through image processing with BSRGAN*, 2025.
 - 28 B. Koonce, *Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization*, 2021.
 - 29 D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014.