



OPEN Effects of transfer learning on seedling detection from drone imagery: a layer freezing study with faster R-CNN

Hyungsik Jeong¹, Joungwon You¹, Dayoung Kim^{1,4}, Peter Sang-Hoon Lee²✉ & Hee Han³✉

Effective decision-making in forest management relies on accurate and timely data, yet conventional manpower-based methods often fail to provide sufficient coverage. This study develops a cost-effective, deep learning-based monitoring technique for assessing replanted areas following timber harvesting, utilizing a small, field-verified dataset. Orthophotos of a two-year-old *Larix kaempferi* plantation were acquired using drones, and seedling locations were identified through field surveys. The Faster R-CNN model with a ResNet-50 backbone was trained under different layer freezing settings and with or without pretrained weights. To improve experimental accuracy, data augmentation, *k*-fold cross-validation, and hyperparameter random search were applied. Freezing some of the initial layers while fine-tuning the remaining layers with pretrained weights resulted in the best performance, achieving an F1 score of 0.87-surpassing the 0.84 score obtained without pretrained weights. These findings demonstrate that transfer learning and layer-freezing strategies effectively enhance seedling detection performance, even with limited field validation data. The proposed method offers a cost-effective approach to forest management data analysis by reducing reliance on traditional survey methods while simultaneously improving data collection efficiency and utilization.

Keywords Transfer learning, Limited dataset, Faster R-CNN model, Layer freezing, Reforestation monitoring, Seedling detection

Reforestation following timber harvesting is essential for restoring forest ecosystems and ensuring sustainable resource management for future generations¹. However, newly planted seedlings in reforested areas face multiple environmental challenges that threaten their survival. To promote seedling establishment, continuous monitoring and timely management interventions are required². The early stages of seedling growth are particularly critical, as effective management during this period can significantly influence the success of forest regeneration³. The implementation of appropriate measures—such as promoting natural regeneration and controlling invasive species—necessitates reliable monitoring systems to track seedling survival and growth.

Traditionally, forest data collection has relied on labor-intensive field surveys. However, the complex topography and dense vegetation of forested landscapes make large-scale, high-frequency data acquisition both costly and time-consuming. This challenge is especially pronounced in reforested areas, where frequent monitoring is needed to assess seedling establishment and survival. This reliance on labor-intensive approaches not only limits the scalability of monitoring efforts but also raises concerns about data consistency and accuracy⁴. Furthermore, the relatively low productivity of Korea's forestry sector⁵, combined with the high costs of traditional data collection methods, imposes significant economic constraints on forest management.

Recent advancements in information technology (IT) and artificial intelligence (AI) have introduced remote sensing and deep learning-based methods as powerful tools for forest surveying. These techniques offer cost-effective solutions that address many of the limitations of traditional field-based methods, particularly in challenging terrain. However, ensuring data reliability remains a significant challenge, as deep learning models often lack interpretability and require extensive field validation. The so-called 'black box' nature of these models

¹Department of Agriculture, Forestry and Bioresources, Seoul National University, Seoul 08826, Republic of Korea.

²Department of Urban & Regional Development, Hanyang University, Seoul 04763, Republic of Korea. ³Department of Agriculture, Forestry and Bioresources, Seoul National University and Research Institute of Agriculture and Life Science, Seoul National University, Seoul 08826, Republic of Korea. ⁴Forest Management Research Division, National Institute of Forest Science, Seoul 02455, Republic of Korea. ✉email: peter337@hanyang.ac.kr; hee.han@snu.ac.kr

makes it difficult to assess the basis of their predictions⁶. Therefore, an integrated approach that combines the efficiency of AI-driven methods with the accuracy of traditional field surveys is essential for improving forest monitoring.

One key area where these technologies can make a significant impact is in reforestation monitoring, particularly in tracking seedling establishment and survival. Effective post-planting assessment is essential for evaluating the success of reforestation efforts and informing timely management interventions. Traditional monitoring methods, which rely on manual field surveys, are often labor-intensive and impractical for large-scale assessments. Furthermore, the vast spatial extent and heterogeneity of forested landscapes present persistent challenges for collecting high-quality training data for AI models in forestry.

Transfer learning has emerged as a promising solution to this issue, enabling models to leverage knowledge from large-scale datasets and adapt it to new, data-scarce environments through fine-tuning⁷. Compared to training a model from scratch, transfer learning enhances performance while requiring significantly less labeled data, making it a viable approach for improving seedling detection in forestry applications.

UAV-based remote sensing has demonstrated potential as an alternative to conventional methods for seedling detection and post-harvest reforestation site monitoring. Early studies utilizing UAV remote sensing data have substantiated this potential. For instance, Feduck et al. reported a 75.8% detection rate using consumer-grade RGB cameras with simple spectral decision rules, achieving high correlation ($r=0.98$) with plot-level field surveys, primarily validating performance under leaf-off conditions⁸. Buters et al. implemented individual and cluster-level detection and tracking through automatic image classification and multi-sensor fusion (RGB/multispectral/point cloud)^{9,10}. However, these approaches exhibited substantial dependencies on rules, timing, and sensors, constraining inter-site generalization and operational scalability, while demonstrating limited sophistication and extensibility compared to learning-based end-to-end object detection methods.

Advances in artificial intelligence technologies, including machine learning and deep learning, have further enhanced UAV-based seedling detection. Through proven machine learning techniques such as Random Forests, Guo et al. demonstrated high success rates in seedling recognition and localization within plantations, albeit requiring extensive feature extraction for training¹¹. Subsequently, the development of CNNs, which exhibit particular strength in image processing, reduced this dependency on manual feature extraction. CNN-based end-to-end learning has improved applicability across diverse reforestation environments by learning features directly from data. Building upon CNN advancements, various methods for rapid, large-scale application of CNNs to RGB orthoimages have been proposed; however, these require substantial labeled datasets for supervised model training^{12–14}. Furthermore, attempts have been made to enhance CNN model accuracy by incorporating multispectral data beyond high-resolution UAV imagery alone^{15,16}. As CNN object detection models evolved, efforts emerged to apply advanced segmentation-based techniques such as Mask R-CNN^{17–19}. While these endeavors improved accuracy, data requirements remained substantial, with many studies leveraging multispectral data in addition to requiring pixel-level annotations for supervised learning. To leverage deep learning benefits, attempts were made to augment data through techniques such as data augmentation²⁰ and StyleGAN²¹. Additionally, methods utilizing UAV-acquired point cloud data have been proposed, though these similarly require extensive computational resources for processing voluminous point cloud datasets^{22,23}. Despite achieving improved precision through novel technological applications, these attempts have not adequately addressed practical constraints of data availability and computational resources inherent in field conditions.

In this context, transfer learning has also been applied in forestry-related UAV monitoring. Osco et al. leveraged ImageNet pre-trained weights for CNN initialization in citrus tree counting from multispectral UAV imagery, achieving improved convergence and performance despite domain differences¹⁶. Similarly, Pearse et al. fine-tuned a Faster R-CNN with a VGG16 backbone pre-trained on PASCAL VOC for seedling detection, which substantially reduced data requirements and improved accuracy under limited training conditions¹³. These cases highlight the practical value of transfer learning in mitigating data scarcity and enhancing model performance in forestry applications.

While transfer learning has been partially adopted as a pragmatic alternative to overcome these data and resource limitations, fundamental methodological considerations regarding the optimal ways to apply transfer learning remain relatively unexplored. Particularly, research examining the impact of systematic fine-tuning strategies—specifically, determining which layers of pre-trained models should be frozen versus retrained (fine-tuned)—on seedling detection performance remains largely underexplored. This suggests an insufficient understanding of methods to maximize model training efficiency and final performance beyond mere technological application in field environments. Therefore, this study aims to address this gap by experimentally investigating the effects of various layer-freezing strategies on deep learning model performance in UAV-based seedling detection and proposing optimal methodological approaches.

Building on these advancements, this study evaluates the effectiveness of transfer learning for seedling detection using UAV-derived imagery and a small, field-verified dataset. A Faster R-CNN model with a ResNet-50 backbone is employed under different layer-freezing configurations to identify the optimal approach for enhancing detection accuracy. This study aims to provide a key methodological contribution by experimentally identifying which specific freezing strategies can enhance model performance, going beyond the mere application of transfer learning. The findings provide insights into the practical application of deep learning in forest management and contribute to the development of cost-effective AI-driven monitoring techniques for reforestation.

Methodology

Study site

The study site is located in a reforested area within Hongcheon-gun, Gangwon Province, Republic of Korea (37.7990°N, 127.9385°E–37.8002°N, 127.9425°E) (Fig. 1). This region has a temperate climate, with an average

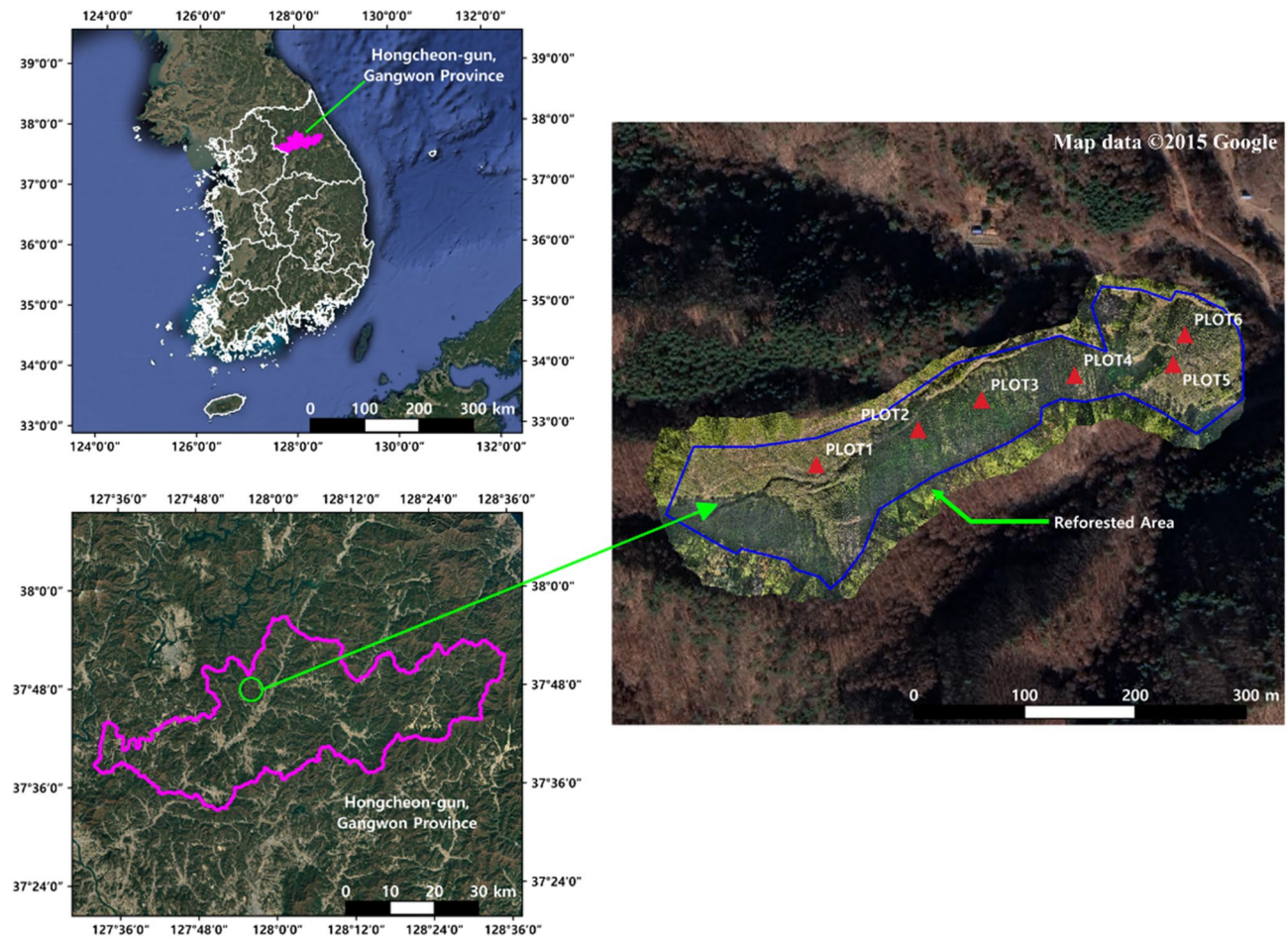


Fig. 1. Location of the study area in Hongcheon-gun, Gangwon Province, Republic of Korea. The figure was created by the authors in QGIS (version 3.40; <https://qgis.org/>) by overlaying a UAV orthomosaic generated from drone imagery using DJI Terra (version 5.0.2; <https://enterprise.dji.com/kr/dji-terra>) onto Google Earth satellite imagery.

annual precipitation of approximately 1338 mm and an average annual temperature of 10.8 °C. The total area of the site spans 4.7 hectares and is situated on a north-facing slope at an elevation of 260–415 m, with an average gradient of 29.8°. Following clearcutting in the spring of 2023, two-year-old *Larix kaempferi* seedlings were planted at 1.8 m intervals, with a planting density of 3,000 trees per hectare. The seedlings currently range in height from 0.3 to 1.5 m, with an average height of 1.0 m. To support early-stage growth, vegetation clearing was conducted during the summers of 2023 and 2024.

Field survey

Field surveys were conducted in November 2024. Six randomly selected plots were established within the reforested area (Fig. 1). Although each plot was initially designed as a circular area with an 8 m radius (Fig. 2b), additional data collection along the plot boundaries yielded irregularly shaped polygons. The plots contained between 65 and 97 seedlings, with a total of 444 seedlings surveyed across all plots.

During the deep learning model training phase, Plots 1–5 were used for training and validation, while Plot 6 was designated as the test set. Plot 6 was specifically chosen for testing due to its northeast-facing orientation, which differed from the other plots, allowing for an assessment of the model's generalization capability. Additionally, its relatively high seedling count (97) contributed to improved statistical stability, reducing variability in performance metrics. A summary of the six plots is provided in Table 1.

To ensure accurate alignment between field survey data and UAV-derived orthophotos, each plot center was marked with red forestry tape, serving as a ground control point (GCP) (Fig. 2a). Within each plot, the relative spatial coordinates of all seedlings were recorded (Fig. 2b). Horizontal distances were measured using a Haglöf Vertex rangefinder, while azimuth measurements were obtained with a Suunto compass, enabling precise localization of seedlings relative to the plot center.

UAV data collection

Following the field survey, UAV imagery was acquired using a DJI Mavic 3 Multispectral drone. The drone operated in autonomous mode along a predefined flight path (Fig. 3) at an altitude of 30 m above ground level,

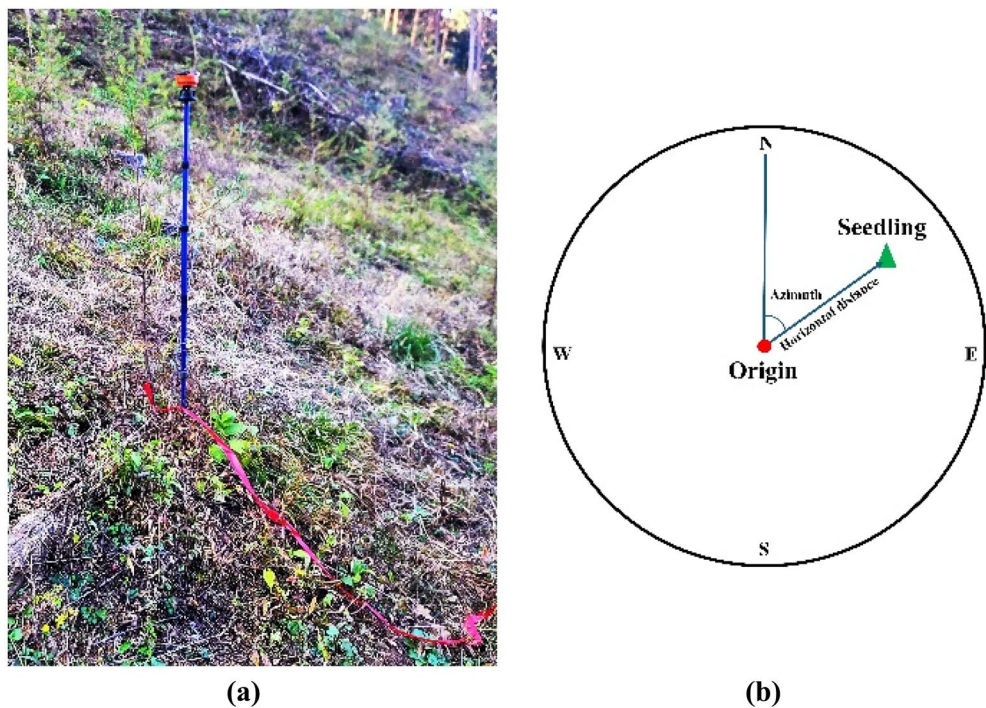


Fig. 2. Field Survey Methods: **(a)** GCP Installation; **(b)** Schematic of Seedling Location Measurement at the Plot Center.

PLOT	Number of seedlings	Coordinate		Aspect	Classification
		N	E		
Plot 1	72	37.7991	127.9386	Southeast	Training/Validation set
Plot 2	74	37.7994	127.9397	Northwest	Training/Validation set
Plot 3	69	37.7996	127.9403	Northwest	Training/Validation set
Plot 4	65	37.7998	127.9413	Northwest	Training/Validation set
Plot 5	67	37.7999	127.9423	Northwest&Southeast	Training/Validation set
Plot 6	97	37.8001	127.9424	Northeast	Test set
Total	444				

Table 1. Characteristics of the sample plots.

maintaining a consistent height through its automatic altitude control system. At this altitude, the ground sampling distance (GSD) was 0.81 cm per pixel. To ensure comprehensive coverage, the image overlap was set to 70% forward and 80% lateral, and the flight speed was maintained at 5 m/s. To mitigate the risk of data loss and ensure the availability of backup data, repeated image acquisitions were conducted over the same area under identical flight conditions. The collected images were subsequently processed to generate an orthomosaic. The key flight parameters for image acquisition are summarized in Table 2.

To integrate the field survey results with the UAV imagery, a multi-step georeferencing procedure was applied. First, GCPs marked with red forestry tape during the field survey were identified in the orthomosaic to precisely locate the plots. The field survey data were then converted into point data and aligned with the orthomosaic. Only seedlings verified during the field survey were extracted and labeled for supervised learning. Corresponding segments of the orthomosaic surrounding these survey plots were then used to generate annotated training data (Fig. 4).

For data processing, QGIS (open-source GIS software) was used to align the point data with the orthomosaic, DJI Terra was utilized for orthomosaic generation, and GIMP and Label Studio (both open-source) were employed for image preprocessing and labeling. The overall research workflow is presented in Fig. 5.

Experimental setup

Deep learning architecture

This study employed the Faster R-CNN framework, a two-stage object detection model proposed by Ren et al.²⁴, for seedling detection. Faster R-CNN consists of a Region Proposal Network (RPN), which efficiently identifies potential regions of interest (RoIs), followed by a classification stage that refines predictions. Compared to earlier

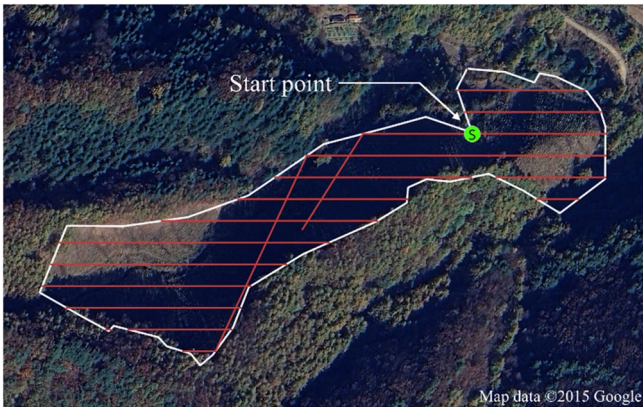


Fig. 3. Predefined autonomous flight path ensuring systematic coverage of the 4.7 ha study site. The arrow indicates the mission starting point. The figure is a schematic illustration of the flight path and was manually drawn by the authors in QGIS (version 3.40; <https://qgis.org/>) using Google Earth satellite imagery as the basemap.

Flight parameters	
Distance	4,838 m
Flight time	17 m 8 s
Flight Altitude	30 m
Flight Speed	5 m/s
Number of images	774
Area	4.7 ha
GSD	0.81 cm/pixel

Table 2. Key flight and data acquisition parameters for the UAV survey.

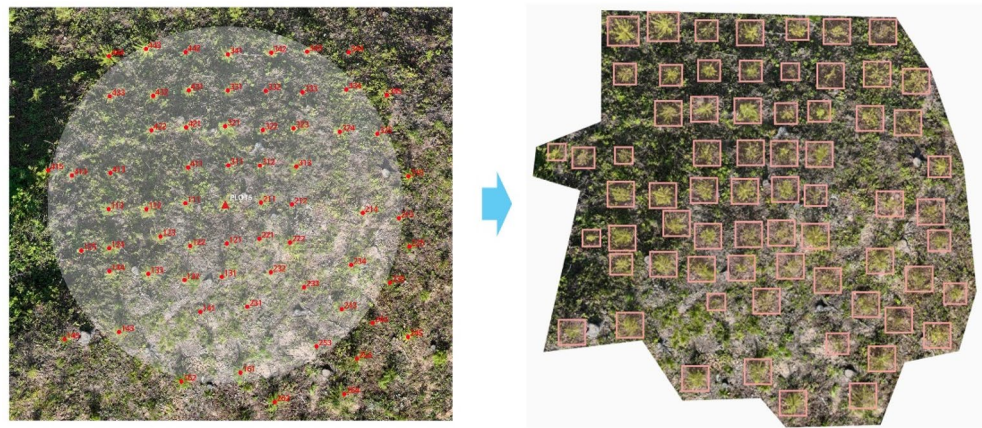


Fig. 4. Overlay of field survey data and UAV orthomosaic images.

models such as R-CNN and Fast R-CNN, this architecture improves both speed and accuracy. While more recent variants, such as Mask R-CNN and Cascade R-CNN, offer enhanced performance in specific applications, they introduce additional computational complexity and make it more challenging to isolate the impact of backbone network modifications^{25,25}.

While numerous object detection models have recently been introduced, including one-stage detectors like the YOLO series and Transformer-based DETR, Faster R-CNN was selected for three main reasons. First, its two-stage detection framework has been widely reported to perform well in small-object detection tasks, which matches the requirements of seedling detection from UAV imagery²⁷. Second, in data-constrained environments such as ours, Faster R-CNN has shown relatively stable convergence and robustness compared with more recent but data-hungry models. Finally, the central objective of this study is to evaluate transfer learning strategies

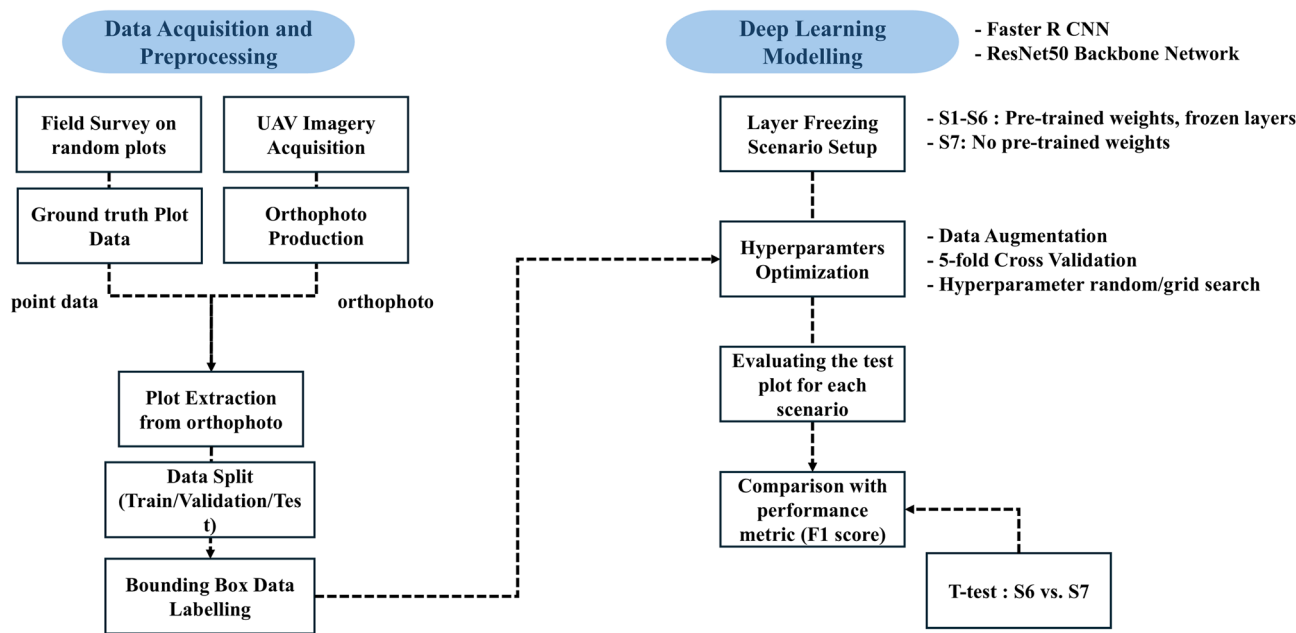


Fig. 5. Research workflow diagram.

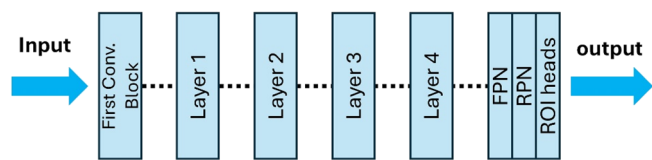


Fig. 6. Schematic architecture of the faster R-CNN with a ResNet-50 backbone.

Scenario	Pretrained weights	Layer-freezing scheme (●:Frozen, ○:Trained)						
		First Conv Block	Layer1	Layer 2	Layer3	Layer4	FPN ROI heads	RPN
S1	Yes	●	●	●	●	●	○	○
S2	Yes	●	●	●	●	○	○	○
S3	Yes	●	●	●	○	○	○	○
S4	Yes	●	●	○	○	○	○	○
S5	Yes	●	○	○	○	○	○	○
S6	Yes	○	○	○	○	○	○	○
S7	No	○	○	○	○	○	○	○

Table 3. Scenario Summary of Layer-Freezing Strategies.

through selective layer freezing. Therefore, Faster R-CNN—one of the most widely used and benchmarked models—was adopted, providing a stable and widely recognized foundation for methodological evaluation.

For the backbone network, ResNet-50—a deep convolutional neural network introduced by He et al.²⁸—was utilized to leverage the benefits of residual learning, which mitigates the vanishing gradient problem in deep networks and enables stable, high-performance training (Fig. 6). Additionally, pretrained weights from the COCO (Common Objects in Context) dataset were incorporated to enhance feature extraction. The COCO dataset comprises approximately 120,000 images annotated across 80 object categories and is widely used for image classification and object detection tasks²⁹. The use of pretrained weights was expected to improve model generalization, particularly given the limited size of the field-verified dataset used in this study.

Layer-freezing scenarios

To evaluate the impact of different layer-freezing strategies on seedling detection performance, multiple scenarios were designed by varying the number of trainable layers in the ResNet-50 backbone and assessing the effect of using pretrained weights from the COCO dataset (Table 3).

In Scenario S1, all backbone layers were frozen, and only the final detection modules—including the region of interest (ROI) heads, region proposal network (RPN), and feature pyramid network (FPN)—were trained. In Scenarios S2–S5, progressively more layers of the backbone were unfrozen to allow deeper feature learning. Finally, in Scenarios S6 and S7, the entire network was made trainable. The only difference between S6 and S7 was whether pretrained COCO weights were used, enabling a direct comparison between training from scratch and fine-tuning with transfer learning.

The primary objective of this study is to systematically investigate the impact of various layer-freezing strategies on model performance within a transfer learning framework for seedling detection. To effectively isolate the variable of interest—the fine-tuning strategy—we deliberately adopted a single, well-established backbone architecture, ResNet-50, as a controlled baseline.

Structural optimization of deep learning backbones is certainly an important research direction that could lead to performance improvements. However, the purpose of this study was not to develop a new architecture, but to systematically evaluate layer-freezing strategies within a controlled transfer learning framework. For this reason, the standard ResNet-50 was used as a fixed baseline to isolate the effects of freezing strategies without confounding factors from structural modifications.

Training and evaluation process

The model was fine-tuned for each scenario using the field-collected dataset. UAV orthomosaics were prepared for six plots, with each plot represented by a cropped image containing 65–97 seedlings annotated with bounding boxes. The dataset was split into four training images, one validation image, and one test image. Due to the dataset's limited size, fivefold cross-validation was used to assess generalization performance.

To improve training efficiency while minimizing overfitting, data augmentation techniques were applied to the training set. These included horizontal flipping, brightness/contrast adjustments ($\pm 20\%$), and 90° rotations³⁰. This augmentation process doubled the training dataset from four to eight images. The validation and test sets remained unchanged to preserve real-world conditions for evaluation.

Considering computational constraints, the optimizer, learning rate scheduler, and batch size were fixed during hyperparameter tuning. The learning rate, weight decay, and number of training epochs were optimized through a two-stage process:

- (1) Grid search: Learning rates from $1e-2$ to $1e-5$ and weight decay values from $1e-2$ to $1e-6$ were systematically tested in 25 combinations, each trained for 10 epochs. The top three configurations were selected based on validation-set mean Average Precision (mAP).
- (2) Random search: The three best-performing configurations were refined by exploring $\pm 20\%$ variations in learning rate and weight decay, generating 30 additional parameter combinations. Each was trained for 20 epochs, and the combination yielding the highest mAP was selected as the final configuration.

Once optimal hyperparameters were identified, the models were trained for up to 50 epochs, repeated five times for stability. Early stopping was not applied to allow full convergence. The best-performing epoch for each fold (based on mAP) was recorded, and the median across folds was used to determine the final epoch count.

For final model training, the training and validation sets were combined, and models were trained using the optimized hyperparameters. For each scenario, we repeated the entire training procedure 30 times, producing 30 independently trained models; each model was evaluated once on the same fixed test set. Mean and standard deviation were computed for each performance metric across the 30 runs.

All deep learning experiments were conducted using the PyTorch framework in Google Colab Pro environment. To ensure reproducibility, random seeds were set to 42 across all scenarios, and the computational environment was reinitialized between runs to prevent variable carryover.

Accuracy assessment

Model performance was evaluated using Recall (R), Precision (P), and F1 Score (F):

$$R = TP / (TP + FN) . \quad (1)$$

$$P = TP / (TP + FP) . \quad (2)$$

$$F = 2 \times R \times P / (R + P) . \quad (3)$$

where TP (True Positives) represents correctly detected seedlings, FN (False Negatives) denotes seedlings present in the ground truth but not detected, and FP (False Positives) refers to seedlings incorrectly identified by the model. A detection was classified as correct (TP) only if all of the following conditions were met:

- (1) The Intersection over Union (IoU) between the predicted and ground-truth bounding boxes was ≥ 0.4 .
- (2) The predicted bounding box center was within 10% of the bounding box size from the ground-truth center.
- (3) The aspect ratio of the predicted bounding box fell within $\pm 20\%$ of the ground-truth bounding box's aspect ratio.

Results

Hyperparameter optimization

Table 4 presents the optimized hyperparameter settings for each scenario, including the selected learning rate, weight decay, number of epochs, and mean mAP recorded on the validation set. The final learning rates ranged from $8.2e-3$ to $1.17e-2$, while weight decay values varied between $1e-6$ and $1.1506e-2$. The optimal number of

Scenario	Pretrained weights	Layer-freezing scheme (●:Frozen, ○:Trained)						Learning Rate	Weight decay	Average Epoch of Median Best mAP	Average Map of Median Best Epochs
		First Conv Block	Layer1	Layer 2	Layer3	Layer4	FPN ROI heads RPN				
S1	Yes	●	●	●	●	●	○	0.008643	0.000008	20	0.4777
S2	Yes	●	●	●	●	○	○	0.011719	0.000085	19	0.4815
S3	Yes	●	●	●	○	○	○	0.008203	0.011506	26	0.4857
S4	Yes	●	●	○	○	○	○	0.008643	0.000001	29	0.5066
S5	Yes	●	○	○	○	○	○	0.009036	0.001174	28	0.5031
S6	Yes	○	○	○	○	○	○	0.008643	0.000827	23	0.5023
S7	No	○	○	○	○	○	○	0.010094	0.000943	27	0.4031

Table 4. Result of Hyperparameter Optimization.

Scenario	Mean ± Standard Deviation					
	TP	FP	FN	Precision	Recall	F1 Score
S1	75.3 ± 4.5	4.5 ± 2.2	21.7 ± 4.5	0.9444 ± 0.0258	0.7759 ± 0.0459	0.8510 ± 0.0288
S2	72.0 ± 8.1	3.6 ± 3.1	25.0 ± 8.1	0.9547 ± 0.0372	0.7419 ± 0.0832	0.8317 ± 0.0539
S3	68.6 ± 3.3	0.9 ± 1.0	28.4 ± 3.3	0.9871 ± 0.0146	0.7076 ± 0.0344	0.8238 ± 0.0251
S4	76.2 ± 2.7	1.1 ± 0.9	20.8 ± 2.7	0.9864 ± 0.0117	0.7852 ± 0.0280	0.8741 ± 0.0173
S5	75.6 ± 3.2	1.3 ± 1.3	21.4 ± 3.2	0.9832 ± 0.0169	0.7797 ± 0.0334	0.8694 ± 0.0232
S6	76.1 ± 3.4	2.0 ± 1.6	20.9 ± 3.4	0.9750 ± 0.0190	0.7842 ± 0.0347	0.8687 ± 0.0200
S7	73.0 ± 4.8	2.4 ± 1.6	24.0 ± 4.8	0.9680 ± 0.0211	0.7529 ± 0.0496	0.8461 ± 0.0348

Table 5. Test set performance results by scenario.

epochs across scenarios fell within the range of 19 to 29, and the mean validation set mAP ranged from 0.4031 to 0.5066.

Among the scenarios, S4 achieved the highest mean mAP of 0.5066 at epoch 29, with a learning rate of $8.643\text{e-}3$ and a weight decay of $1\text{e-}6$. Although S4 converged more slowly compared to other scenarios that used pretrained weights, its small weight decay facilitated gradual and precise parameter updates, effectively leveraging pretrained knowledge for seedling detection. In contrast, S1 and S2 exhibited faster convergence but lower overall performance, likely due to excessive layer freezing, which restricted the model's ability to learn seedling-specific features.

Scenario S7, which was trained without pretrained weights, had the lowest mean mAP (0.4031). Since S7 started with random initial weights, it required a higher learning rate and more training epochs. Despite these adjustments, training from scratch consistently underperformed, highlighting the superiority of transfer learning for seedling detection.

The validation set results in this section serve as an intermediate assessment for hyperparameter tuning. While these findings helped identify promising parameter configurations, final performance evaluation was conducted on the test set, which remained unused during training to ensure unbiased assessment.

Final performance by scenario

Table 5 summarizes the seedling detection performance for each layer-freezing scenario, evaluated on the test set. Results are reported as Mean ± Standard Deviation. A graphical comparison is provided in Fig. 7.

Across all scenarios, precision remained consistently high (>0.94), indicating that most identified seedlings were correct detections, with a low false-positive rate. However, recall ranged from 0.70 to 0.78 and exhibited greater variability, suggesting that some actual seedlings were missed. This indicates that while the model is highly precise in classifying detected seedlings, its ability to detect all existing seedlings remains somewhat limited. As a result, the F1 score, which balances precision and recall, ranged from 0.82 to 0.87.

Among the scenarios, S4 achieved the highest F1 score (0.8741 ± 0.0173), aligning with its superior performance observed during hyperparameter selection. S5 (0.8694 ± 0.0232) and S6 (0.8687 ± 0.0200) followed closely. While S3 attained the highest precision, it also recorded the lowest recall, leading to the lowest F1 score (0.8238 ± 0.0251). Scenario S7 (0.8461 ± 0.0348), which was trained without pretrained weights, exhibited intermediate performance, comparable to S1 (0.8510 ± 0.0288) and S2 (0.8317 ± 0.0539).

Scenarios that incorporated pretrained weights and allowed broader trainable layers (S4, S5, S6) demonstrated both higher average F1 scores and lower variability, highlighting their superior performance and greater stability compared to S7. Notably, while S6 and S7 had identical layer-freezing configurations, S6 outperformed S7 across all metrics, benefiting from pretrained weights, as indicated by its higher means and smaller standard deviations.

Statistical significance of performance differences across scenarios

To systematically validate the findings, the analysis was structured in two stages.

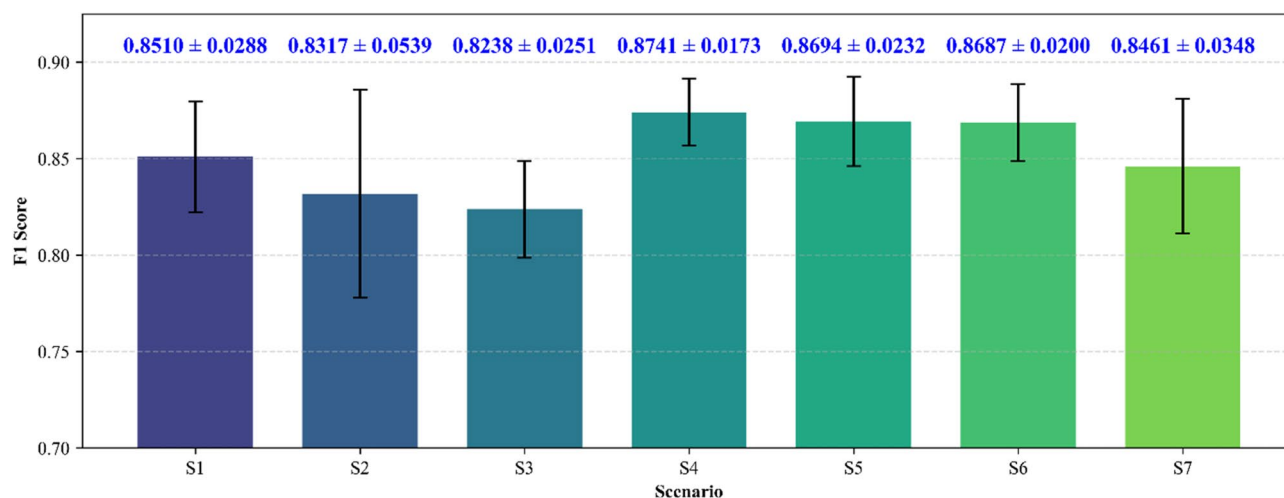


Fig. 7. F1 Score comparison by scenario with standard deviation.

Classification	Levene's test		S6(a)		S7(b)		Mean difference	t	p
	F	P	Mean	S.D	Mean	S.D	(a)-(b)		
F1 score	2.149	0.148	0.8687	0.02	0.8461	0.0348	0.0221	3.076	0.003

Table 6. Levene's test and independent-samples t-test results for F1 score comparison between S6 (pretrained) and s7 (trained from scratch).

Test	Statistic	p-value	Interpretation	Post hoc (Dunn test)
One-way ANOVA	F = 14.57	p < .001	Significant group differences	S4 > S1, S2, S3 S5, S6 > S2, S3 S1 > S3
Welch ANOVA	F = 20.11	p < .001		
Kruskal–Wallis test	$\chi^2 = 60.49$	p < .001		

Table 7. ANOVA Results for F1 score comparisons across fine-tuning strategies (S1–S6).

First, to establish the fundamental benefit of transfer learning, the fully fine-tuned model (S6) was compared against its counterpart trained from scratch (S7). These scenarios shared an identical configuration of trainable layers, differing only in weight initialization. An independent samples t-test was conducted on the F1 scores to verify the significance of the performance gap. Levene's test confirmed the equality of variances ($p = 0.148$), and the subsequent t-test revealed that the mean F1 score of S6 was significantly higher than that of S7 ($p < 0.01$). This result provides clear statistical evidence for the superiority of using pre-trained weights for this task (Table 6).

Second, having confirmed the overall value of pre-training, the study sought to identify the optimal fine-tuning strategy among the various approaches (S1–S6). A comprehensive series of hypothesis tests was conducted on this specific set of scenarios. A one-way ANOVA indicated highly significant differences among the groups ($F = 14.57$, $p < 0.001$). However, as the assumptions of ANOVA were not met in some groups, additional tests were performed. Both Welch's ANOVA ($F = 20.11$, $p < 0.001$) and the non-parametric Kruskal–Wallis test ($\chi^2 = 60.49$, $p < 0.001$) yielded consistent results, confirming the robustness of these findings.

A post-hoc Dunn's test was then performed to pinpoint the specific sources of these differences. The results showed that S4 significantly outperformed S1, S2, and S3 (all $p < 0.01$), while S5 and S6 also performed better than S2 and S3 ($p < 0.01$), and S1 exceeded S3 ($p < 0.05$). These results provide strong statistical backing for the main conclusion that a balanced fine-tuning strategy, as represented by S4, is the most effective approach. The detailed post-hoc comparisons are summarized in Table 7.

Failure analysis and error characteristics

To address the precision–recall gap, we conducted a failure analysis focusing on false negatives (FNs). We found that missed seedlings were strongly associated with smaller bounding-box sizes. Using a representative detection output (Fig. 8) as the basis for quantifying box-scale differences, we compared the ground truth bounding-box areas of true positives (TPs) and FNs across the test runs. The results show that the mean bounding-box area of TPs was approximately 1.9 times larger than that of FNs. This indicates that the model's sensitivity decreases as the seedling scale diminishes, leading to higher omission rates among smaller individuals. This distributional

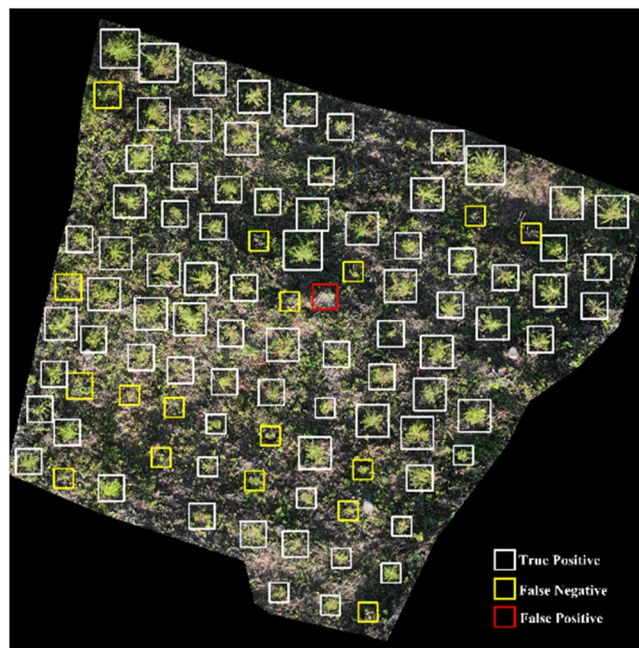


Fig. 8. Representative test-set detection outputs used for failure analysis.

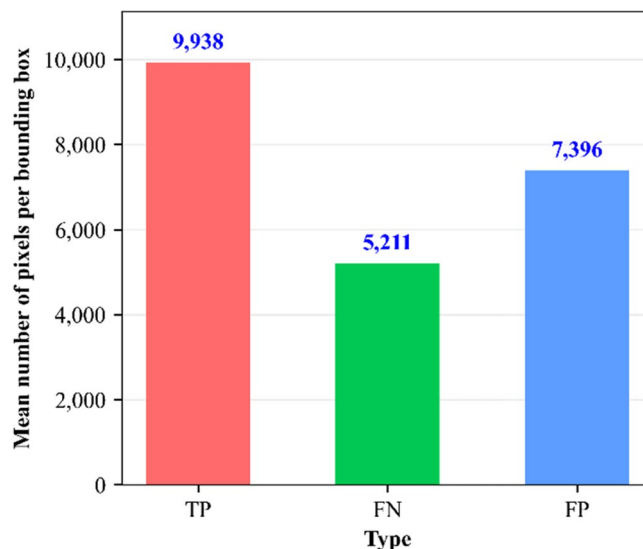


Fig. 9. Distributions of ground-truth bounding-box areas for TPs and FNs.

gap is visualized in Fig. 9, confirming that limited detectability of small-scale seedlings in complex backgrounds is the primary driver of the lower recall.

Visual inspection further supports this interpretation (Fig. 10.). In TP examples, the seedling's characteristic radial stem structure is clearly visible, providing strong and consistent cues for detection. In contrast, FN examples tend to be smaller and visually ambiguous: the radial structure is less distinct, and the seedlings are more easily confused with surrounding herbaceous vegetation.

Summary of results

Scenario S4 emerged as the optimal configuration, achieving the highest mean F1 score (0.8741 ± 0.0173). Statistical analyses confirmed the significant superiority of S4 over S1–S3 ($p < 0.01$) and the effectiveness of transfer learning via S6 ($p < 0.01$). These findings identify S4 as a reliable and automated configuration for seedling detection, effectively addressing the performance-stability trade-off in forestry data environments. Furthermore, our failure analysis indicates that the lower recall is primarily driven by the limited detectability of small-scale seedlings. Rather than a simple increase in dataset size, focusing on the acquisition and learning

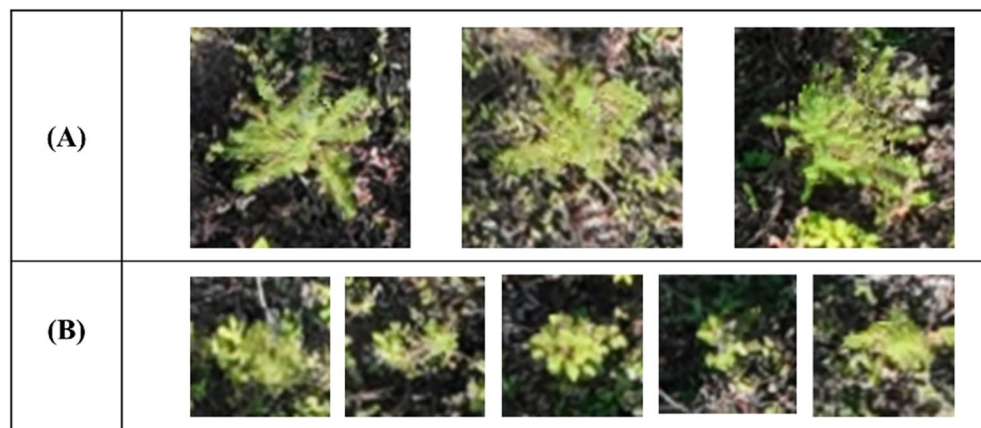


Fig. 10. Visual comparison of representative true-positive (A) and false-negative (B) Examples from the test set.

of small-scale seedling data represents a more efficient approach to improving detection robustness in complex forest environments.

Discussion

The findings of this study demonstrate that pretrained weights significantly enhance seedling detection performance, reinforcing the effectiveness of transfer learning in forestry applications. Despite being trained on the COCO dataset, which primarily consists of everyday objects and scenes, the pretrained model successfully transferred knowledge to the specialized domain of seedling detection²⁸. These results show that even partially dissimilar datasets can provide useful feature representations for forestry. Transfer learning can therefore bridge general-purpose and domain-specific tasks. Especially in data-limited forestry environments where acquiring large-scale annotated datasets remains challenging and resource-intensive, transfer learning constitutes an effective alternative.

Beyond the impact of pretrained weights, the range of trainable layers also played a crucial role in detection performance. While increasing the number of trainable layers generally led to better results, the relationship was not strictly linear. Scenario S4, which fine-tuned intermediate and later layers, yielded superior results, indicating that mid-level representations capture essential seedling features. This aligns with the common understanding that CNNs extract low-level general features (e.g., edges and textures) in early layers, while mid-to-late layers capture more complex, domain-specific attributes³¹. This suggests that leveraging pretrained knowledge for low-level general features while focusing computational resources on capturing seedling-specific characteristics leads to more efficient resource allocation.

More specifically, our results suggest that the second residual block group of ResNet-50 (Layer2) plays a particularly important role in seedling detection. Scenarios S4–S7, which fine-tuned layers beyond Layer2, consistently outperformed Scenarios S1–S3, in which Layer2 remained frozen. This indicates that the mid-level features encoded in Layer2 capture critical textural and structural cues that help distinguish seedlings from complex natural backgrounds.

By contrast, making all layers trainable in Scenarios S6 and S7 did not yield superior performance compared with S4, suggesting that for small objects such as seedlings, very high-level semantic abstractions may provide limited additional benefit. This suggests that indiscriminately unfreezing more layers does not always yield better results and that certain pretrained layers are already well-suited for seedling detection. Simply allowing more layers to be trained without a structured strategy can lead to suboptimal feature adaptation, underscoring the need for a more nuanced approach to layer freezing.

Interestingly, S1, which had extensive layer freezing, outperformed S2 and S3, which adopted wider unfreezing (fine-tuning) ranges. This suggests that seedlings may have visual characteristics that are largely captured by the general representations learned in a pretrained backbone. At the same time, in data-limited settings, expanding the trainable range without enough data to guide stable learning may increase sensitivity to incidental variations in the training set or site-specific background patterns, thereby increasing the risk of overfitting rather than delivering consistent performance gains. Therefore, when labeled data are scarce and the optimal unfreezing range is uncertain, a conservative fine-tuning strategy that largely preserves pretrained knowledge may serve as a more reliable and stable baseline.

While this study utilized ResNet-50 to isolate and verify the effectiveness of fine-tuning strategies, the proposed approach is expected to be applicable to more modern backbones, such as ResNet-101 or Vision Transformer (ViT) architectures. The core principle—preserving early-layer generic features while adapting deeper representations to task-specific requirements—remains a robust strategy for seedling detection. However, the optimal 'cut-point' for layer freezing may shift depending on the specific block composition and depth of the backbone.

From a resource utilization perspective, fine-tuning should prioritize layers that effectively capture seedling characteristics while minimizing unnecessary updates to layers already optimized through pretraining. This

Item	Manual survey	Drone-based survey
Survey area	4.7 ha	
Required personnel	2 persons (1 team)	1 person
Basis of calculation	25 ha/day	10 min (Drone setup and takeoff preparation)
		35 min (Flight & image capture)
		10 min (Equipment wrap-up)
Total person-hours	3.01	0.92
Labor reduction	Baseline	69.5% reduction
Reproducibility	Low (Operator-dependent)	High (Consistent data)
Data utility	Limited (Numerical records, low repeatability)	High (Imagery, time-series analysis)
Risk exposure	High (Weather, terrain)	Low (Remote operation)

Table 8. Comparative analysis of manual and drone-based survey methods. The reported time savings refer specifically to the operational phase (on-site survey and data processing). The initial time investment for data annotation and model training, which are non-recurring development tasks, is not included in this comparison.

targeted use of computational resources makes drone-based remote sensing more viable for forest monitoring in data- and resource-constrained settings. Therefore, beyond validating algorithmic improvements, it is essential to evaluate the practical implications that this approach offers to forestry management practices.

In terms of field applicability, the methodology presented in this study has the potential to significantly improve the efficiency of seedling surveys in reforested areas compared to traditional manual methods. The efficiency gains from drone-based remote sensing have already been demonstrated in other ecological monitoring fields, such as wildlife population surveys^{32,33}. Furthermore, the application of transfer learning—particularly through strategic layer freezing in backbone networks—plays a critical role in enhancing the practicality of drone-based forest monitoring by reducing the time and computational resources required for constructing large-scale annotated datasets.

To benchmark the expected efficiency gains, we conducted a quantitative analysis using the standard unit labor costs presented by the Korea Forest Service, as specified in Chapter II, Article 5(2)(4) of the *Standards for Remuneration of Forestry-Technology Services*³⁴. These figures define the labor requirements for assessing seedling survival in reforestation sites via manual surveys and are routinely employed by the Korea Forest Service as a basis for estimating appropriate project costs in national forestry operations. Given that the majority of forestry operations in Korea are government-led, this standard serves as a meaningful benchmark for directly comparing the costs of alternative survey methods from a forest-management perspective.

This comparative analysis focused specifically on the time required for on-site fieldwork, which is the central component of a survey workflow that also includes survey preparation and post-survey data handling. According to the standard labor cost estimates, a two-person team conducting sample-based manual surveys—typically using circular plots that cover approximately 1% of the total area—can cover 25 hectares per day based on an 8-h workday. This corresponds to 0.64 person-hours per hectare. Applying this rate to a 4.7-hectare survey area in this study yields an estimated labor input of 3.01 person-hours for the manual method.

In contrast, the drone-based survey covered the same area with a single operator, including 10 min for preparation, 35 min for flight and image capture, and 10 min for wrap-up. This amounts to 0.92 person-hours. Overall, the drone-based approach resulted in a 69.5% reduction in on-site labor compared to the manual survey, indicating that the required survey objective can be performed with only 30% of the workforce required for conventional methods. Table 8 provides a summary of these findings.

Beyond the quantitative efficiencies, the proposed methodology fundamentally enhances the quality and reliability of the collected data. While manual surveys are susceptible to human-induced errors such as systematic error, random error, and blunders³⁵, the likelihood of such errors is significantly lower in drone-based surveys. This is highly advantageous for maintaining the consistency and reproducibility of survey results. Furthermore, the drone-based approach offers clear advantages for conducting repeated surveys to generate time-series data. It also enhances operational continuity and safety, particularly under extreme weather conditions—such as heatwaves—that are becoming more frequent and severe due to climate change.

Moreover, the computational resources required for post-processing drone data can be mitigated through the application of transfer learning optimized via a layer-freezing strategy. As the data processing algorithms and protocols become increasingly systematized for large-scale reforestation surveys, the initially high fixed costs per unit area are likely to decrease substantially. This suggests that the drone-based method will hold a comparative advantage in terms of overall efficiency. In summary, the drone-based approach represents a practical and effective alternative, capable of significantly enhancing the efficiency of seedling surveys in reforested areas.

To further refine this approach, adaptive learning rate strategies could further improve performance by assigning lower learning rates to layers that are well-suited for transfer learning and higher learning rates to layers requiring greater adaptation. Additionally, a more sophisticated fine-tuning approach could dynamically adjust hyperparameters such as epoch count, learning rate schedules, optimizer selection, and weight decay frequencies based on each layer’s contribution to detection performance.

Previous UAV-based forestry studies have already demonstrated the feasibility of deep learning for tree and seedling detection. For example, Osco et al.¹⁶ developed a CNN framework for citrus tree counting using

UAV multispectral imagery, achieving high accuracy through spectral combinations but without systematically examining transfer learning strategies. Similarly, Pearse et al.¹³ applied CNN-based object detection to map conifer seedlings from UAV RGB orthomosaics with impressive cross-site performance using large annotated datasets, yet their approach relied on a single fine-tuning scheme. In contrast, our study differs by explicitly investigating how selective layer-freezing strategies in ResNet-50 affect model performance under data-limited conditions, thereby addressing a previously underexplored methodological aspect in forestry applications.

The idea of avoiding unnecessary adjustments and focusing only on the essential components is not limited to forestry. In autonomous driving, where traffic sign recognition requires rapid detection in complex scenes, researchers have achieved real-time processing through lightweight models with limited fine-tuning³⁶. In medical imaging, to capture subtle changes (such as Alzheimer's indicators) in high-resolution MRI, approaches that adjust only upper layers or reduce feature dimensions based on pretrained representations are used to achieve stable performance even with small-scale data³⁷. Similar principles can be observed in non-visual domains: IoT intrusion detection and smart building energy prediction both emphasize extracting only essential features to simplify models and improve efficiency^{38,39}. These examples highlight a common principle: learn only what is necessary and maintain the rest. Ultimately, in situations with limited data and computation, strategies that focus on learning only the essential components rather than everything are more practical.

From this perspective, the transfer learning and selective layer freezing used in this study are well-suited for UAV-based seedling detection, where only small datasets are available, images are high-resolution, and practical field use is required. This approach is particularly effective when labeled data is scarce, computing resources such as memory and processing speed are constrained, and small adjustments to the upper layers are sufficient to capture task-specific patterns. Therefore, this study should be understood as a case study demonstrating the effectiveness of this approach under conditions of limited labeled data and constrained computational resources.

The dataset used in this study was collected under constrained environmental conditions within a specific region, potentially limiting the capture of adequate variability in topography, tree species, and seasonal dynamics. These limitations heighten the risk that the proposed seedling detection model may overfit to localized training data features, thereby reducing its generalization capacity across diverse forest environments. Therefore, to improve the generalizability of the proposed transfer learning strategy, additional validation using more heterogeneous datasets is essential. These datasets should cover natural-condition variability, including soil characteristics, vegetation density, lighting conditions, and seedling ages, as well as acquisition-condition variability, such as drone flight altitude (i.e., ground sampling distance).

More specifically, this validation should prioritize realistic field conditions that are likely to challenge S4. Among the most critical environmental factors are increased weed density around seedlings (i.e., stronger occlusion and background clutter) and a broader range of seedling ages and sizes, both of which are expected to intensify domain shift in appearance. In addition, acquisition settings should be tested at higher flight altitudes than the current 30 m to assess performance under reduced image quality (coarser GSD and lower spatial detail). Evaluating S4 under these stress-test conditions would help validate its robustness and generalizability.

Moreover, general-purpose datasets such as COCO have inherent limitations due to their lack of forestry-specific content, resulting in a domain gap that constrains the transferability of learned features to UAV-based seedling detection. To overcome this issue, future research should focus on constructing domain-specific visual datasets tailored to forestry and ecological applications, explicitly evaluating detection performance across multi-temporal UAV acquisitions, and applying self-supervised pre-training and advanced domain adaptation techniques to achieve representation learning better suited for seedling detection under diverse environmental factors.

Future research could benefit from Explainable Artificial Intelligence (XAI) tools to gain deeper insights into how each layer contributes to CNN-based seedling detection. Techniques such as Class Activation Mapping (CAM) and Gradient-weighted Class Activation Mapping (Grad-CAM) could visualize the specific features extracted at different layers, allowing researchers to systematically identify the most relevant layers for fine-tuning^{40,40}. A layer-wise optimization strategy informed by these insights could lead to more specialized, high-precision transfer learning models. This would not only enhance detection accuracy with greater computational efficiency but also support forest managers with reliable, data-driven tools—thereby transforming conventional monitoring practices into precision forestry.

Conclusion

This study evaluated the effectiveness of transfer learning, along with layer freezing strategies designed to maximize transfer learning performance, for seedling detection by integrating high-resolution drone imagery with field survey data. A Faster R-CNN model with a ResNet-50 backbone was trained under various layer-freezing scenarios, with and without pretrained weights, to assess the impact of transfer learning on model performance. Detection accuracy was measured using precision, recall, and F1 score to determine the optimal training strategy for forestry applications.

The results confirmed that scenarios utilizing pretrained weights consistently outperformed those trained from scratch, demonstrating the advantages of transfer learning in seedling detection. While models with a broader range of trainable layers generally performed better, the relationship was not strictly linear, suggesting that certain layers contribute more significantly to seedling detection than others. These findings highlight the importance of a structured fine-tuning approach, where selective layer freezing can optimize feature extraction rather than applying a uniform training strategy across all layers.

By demonstrating that deep learning models enhanced with transfer learning can effectively detect seedlings in reforested areas, this study presents a scalable, cost-efficient approach for forest monitoring. Conventional seedling surveys rely heavily on manual field assessments, which are labor-intensive and often impractical for

large-scale monitoring. The integration of UAV-based imagery and AI-driven detection models provides a viable alternative for improving the efficiency and accuracy of forest regeneration assessments.

Despite these promising results, further research is required to validate the model's generalizability across diverse environmental conditions, including different geographic regions, seasonal variations, and forest types. Additionally, future studies should explore whether the findings extend to newer detection models and backbone architectures beyond Faster R-CNN, such as transformer-based models and advanced vision architectures. Leveraging eXplainable AI (XAI) tools to analyze the contribution of individual layers could also refine transfer learning strategies, leading to more precise and computationally efficient deep learning applications in forestry.

This study provides a foundation for integrating deep learning and UAV-based monitoring into reforestation management, offering new possibilities for cost-effective, data-driven forest monitoring solutions. Advancing these techniques will be crucial for enhancing sustainable forest management and ecological restoration efforts worldwide.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 9 November 2025; Accepted: 7 April 2026



References

1. Paquette, A. & Messier, C. The role of plantations in managing the world's forests in the Anthropocene. *Front. Ecol. Environ.* **8**, 27–34 (2010).
2. Brancalion, P. H. S. & Chazdon, R. L. Beyond hectares: Four principles to guide reforestation in the context of tropical forest and landscape restoration. *Restor. Ecol.* **25**, 491–496 (2017).
3. Engel, V. L. & Parrotta, J. A. An evaluation of direct seeding for reforestation of degraded lands in central São Paulo state, Brazil. *For. Ecol. Manag.* **152**, 169–181 (2001).
4. White, J. C. et al. Remote sensing technologies for enhancing forest inventories: A review. *Can. J. Remote Sens.* **42**, 619–641 (2016).
5. Seok, H. S. & Ahn, S. J. Strategies for fostering forest industry clusters to revitalize regional forestry (translated title). *KREI Issue Report*, PRN071 (Korea Rural Economic Institute, 2013). Available at: <http://repository.krei.re.kr/handle/2018.oak/20487>. [in Korean]
6. Guidotti, R. et al. A survey of methods for explaining black box models. *ACM Comput. Surv.* **51** (05), Article 93 (2018).
7. Iman, M., Arabnia, H. R. & Rasheed, K. A review of deep transfer learning and recent advancements. *Technologies* **11**, 40 (2023).
8. Feduck, C., McDermid, G. J. & Castilla, G. Detection of coniferous seedlings in UAV imagery. *Forests* **9**, 432 (2018).
9. Buters, T., Belton, D. & Cross, A. Seed and seedling detection using unmanned aerial vehicles and automated image classification in the monitoring of ecological recovery. *Drones* **3**, 53 (2019).
10. Buters, T. M., Belton, D. & Cross, A. T. Multi-sensor UAV tracking of individual seedlings and seedling communities at millimetre accuracy. *Drones* **3**, 81 (2019).
11. Guo, X. et al. Tree recognition on the plantation using UAV images with ultrahigh spatial resolution in a complex environment. *Remote Sens.* **13**, 4122 (2021).
12. Fromm, M., Schubert, M., Castilla, G., Linke, J. & McDermid, G. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Remote Sens.* **11**, 2585 (2019).
13. Pearse, G. D., Tan, A. Y. S., Watt, M. S., Franz, M. O. & Dash, J. P. Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data. *ISPRS J. Photogramm. Remote Sens.* **168**, 156–169 (2020).
14. Lopatin, E. & Poikonen, P. Cost-effective aerial inventory of spruce seedlings using consumer drones and deep learning techniques with two-stage UAV flight patterns. *Forests* **14**, 973 (2023).
15. Imangholiloo, M. et al. A new approach for feeding multispectral imagery into convolutional neural networks: Improved classification of seedlings. *Remote Sens.* **15**, 5233 (2023).
16. Osco, L. P. et al. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **160**, 97–106 (2020).
17. Hao, Z. et al. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **178**, 112–123 (2021).
18. Yu, K. et al. Comparison of classical methods and Mask R-CNN for automatic tree detection and mapping using UAV imagery. *Remote Sens.* **14** (2), 295 (2022).
19. Fu, H. et al. Automatic detection tree crown and height using Mask R-CNN based on unmanned aerial vehicles images for biomass mapping. *For. Ecol. Manag.* **555**, 121712 (2024).
20. Moharram, D., Yuan, X. & Li, D. Tree seedlings detection and counting using a deep learning algorithm. *Appl. Sci.* **13**, 895 (2023).
21. Zhang, Y. et al. High-precision detection for sandalwood trees via improved YOLOv5s and StyleGAN. *Agriculture* **14**, 452 (2024).
22. Finn, A., Kumar, P., Peters, S. & O'Hehir, J. Unsupervised spectral-spatial processing of drone imagery for identification of pine seedlings. *ISPRS J. Photogramm. Remote Sens.* **183**, 363–388 (2022).
23. Hartley, R. J. L. et al. An assessment of high-density UAV point clouds for the measurement of young forestry trials. *Remote Sens.* **12**, 4039 (2020).
24. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Preprint at <https://arxiv.org/abs/1506.01497> (2015).
25. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. Preprint at <https://arxiv.org/abs/1703.06870> (2017).
26. Cai, Z. & Vasconcelos, N. Cascade R-CNN: High Quality Object Detection and Instance Segmentation. Preprint at <https://arxiv.org/abs/1906.09756> (2019).
27. Tseng, H.-H. et al. Rice seedling detection in UAV images using transfer learning and machine learning. *Remote Sens.* **14**(12), 2837 (2022).
28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual Learning for Image Recognition. Preprint at <https://arxiv.org/abs/1512.03385> (2015).
29. Lin, T.-Y. et al. Microsoft COCO: Common Objects in Context. Preprint at <https://arxiv.org/abs/1405.0312> (2014).
30. Ho, D., Liang, E., Chen, X., Stoica, I. & Abbeel, P. Population based augmentation: Efficient learning of augmentation policy schedules. In Proc. 36th Int. Conf. Mach. Learn. (ICML), PMLR 97, 2731–2741 (2019).
31. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **27**, 3320–3328 (2014).

32. Dundas, S. J., Vardanega, M., O'Brien, P. & McLeod, S. R. Quantifying waterfowl numbers: Comparison of drone and ground-based survey methods for surveying waterfowl on artificial waterbodies. *Drones* **5**, 5 (2021).
33. Murphy, N. K., Elmore, J. A., Boudreau, M. R., Dorr, B. S. & Rush, S. A. Monitoring active osprey nests with drones is more time efficient and less disturbing than conventional methods. *Wildl. Biol.* **2024**, e01341 (2024).
34. Korea Forest Service. Standards for remuneration of forestry-technology services. *Ministry of Forest Service Notice* No. 2024-116 (2024). Available at: <https://www.law.go.kr/LSW/admRull.sInfoP.do?admRulSeq=2100000250002>. [in Korean]
35. Ye, X., Xiao, X., Shi, J. & Ling, M. The new concepts of measurement error theory. *Measurement* **83**, 96–105 (2016).
36. Kandasamy, K., Natarajan, Y., Sri Preethaa, K. R. & Ali, A. A. Y. A robust TrafficSignNet algorithm for enhanced traffic sign recognition in autonomous vehicles under varying light conditions. *Neural Process. Lett.* **56**, 241 (2024).
37. Abd El-Latif, A. A., Chelloug, S. A., Alabdulhafith, M. & Hammad, M. Accurate detection of Alzheimer's disease using lightweight deep learning model on MRI data. *Diagnostics* **13**, 1216 (2023).
38. Sundaram, K., Subramanian, S., Natarajan, Y. & Thirumalaisamy, S. Improving performance of intrusion detection using ALO selected features and GRU network. *SN Comput. Sci.* **4**, 809 (2023).
39. Natarajan, Y. et al. Enhancing building energy efficiency with IoT-driven hybrid deep learning models for accurate energy consumption prediction. *Sustainability* **16**, 1925 (2024).
40. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning Deep Features for Discriminative Localization. *Preprint at https://arxiv.org/abs/1512.04150* (2015).
41. Selvaraju, R. R. et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **128**, 336–359 (2020).

Acknowledgements

This study was carried out with the support of 'R&D Program for Forest Science Technology (Project No. RS-2024-00402265)' provided by Korea Forest Service (Korea Forestry Promotion Institute).

Author contributions

Hyungsik Jeong designed the research, conducted the data analysis, and wrote the manuscript. Joungwon You contributed to the field survey and data acquisition. Dayoung Kim assisted in literature review and statistical validation. Hee Han and Peter Sang-Hoon Lee supervised the overall study, provided methodological guidance, critically revised the manuscript, and served as the corresponding author. All authors reviewed and approved the final version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.S.-H.L. or H.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026