Research Article

pISSN 1225-6161 • eISSN 2287-9307
Korean Journal of Remote Sensing 2025, 41(6), 1011-1020
https://doi.org/10.7780/kjrs.2025.41.6.7

Korean Journal of Remote Sensing
KJRS

Application of Fine-Tuned Segment Anything Model with Google Maps to Forest Road Extraction from CAS500-1 Satellite Images: A Case Study

Hayoung Lee[1], Kwangseob Kim[2], Kiwon Lee[3]*

[1]Master Student, Department of Convergence Security, Hansung University, Seoul, Republic of Korea
[2]Assistant Professor, Department of Computer Software, Kyungmin University, Uijeongbu, Republic of Korea
[3]Professor, Department of Convergence Security, Hansung University, Seoul, Republic of Korea

Abstract: This study investigates the feasibility of a fine-tuned Segment Anything Model (SAM) for forest road segmentation, utilizing high-resolution imagery from the Compact Advanced Satellite 500-1 (CAS500-1) as a case study. Due to the lack of labeled or domain-specific datasets for forest roads, a custom training dataset was constructed from Google Maps (GM) imagery, enabling the fine-tuning of SAM with a Vision Transformer-Huge (ViT-H) backbone. The results were evaluated against manually labeled ground truth, OpenStreetMap (OSM), and the Forest Big Data Exchange Platform (FBDEP) in Korea. Evaluation metrics included mean Intersection over Union (mIoU) and area-weighted mIoU (AWmIoU). The fine-tuned model achieved notable improvements compared to the baseline SAM (mIoU: 0.6190 to 0.7786; AWmIoU: 0.6970 to 0.8369). These findings confirm the value of domain-specific fine-tuning and demonstrate the potential of SAM for extracting forest roads.

Keywords: CAS500-1, Fine-tuning, Forest road, Google maps, SAM

Received: November 10, 2025
Revised: November 24, 2025
Accepted: November 26, 2025
Published: December 31, 2025

Corresponding author:
Kiwon Lee
E-mail: kilee@hansung.ac.kr

1. Introduction

Roads can be classified by their shape, material, structure, and intended use. Each type presents distinct challenges for automated detection. Although recent advances in deep learning have made significant progress in road extraction (Liu et al., 2024a), several challenges remain. These include occlusions from trees, shadows, and buildings; decreased accuracy in complex backgrounds or under adverse weather conditions; the high cost and time required to generate large-scale labeled datasets; and difficulty generalizing across roads of different widths and surfaces. These challenges are especially pronounced for forest roads, which are typically narrow, unpaved, and often obscured by dense vegetation.

This study addresses these challenges by evaluating the applicability of Meta's Segment Anything Model (SAM), fine-tuned with domain-specific data, to improve segmentation accuracy for forest road extraction. Forest roads play a critical role in management and protection by enabling logging, transportation, and wildfire prevention. These roads include national forest roads managed by the central government, public forest roads maintained by local governments, and private roads constructed by landowners. Depending on their purpose, forest roads may be designed as arterial, operational, or fire-prevention roads, each contributing to the resilience of the forest ecosystem.

These roads can be further categorized based on their specific purpose. Arterial forest roads connect major roads or other forest roads and serve as the main transportation routes within forest areas. Operational forest roads branch off from arterial roads to enable specific forest operations in targeted areas. Establishing

fire-prevention forest roads enables rapid response to forest fires and other forest-related disasters, enhancing overall forest safety and resilience.

Çalışkan and Sevim (2022) applied deep learning models to automatically extract forest roads from orthophotos with a spatial resolution of 10 cm or 5 cm. Kelesakis et al. (2024) noted that road condition classification accuracy may be compromised due to limitations in forest/rural training datasets and vegetation cover, extracting forest/rural road networks using high-resolution Worldview-3 satellite imagery. The dataset's extensive nature enables the model to perform well across various segmentation tasks. SAM is a significant advancement in general-purpose segmentation tools because it can segment virtually any object in an image with minimal user intervention (Kirillov et al., 2023). Osco et al. (2023) examined SAM methodologies for remote sensing applications and noted that fine-tuning is critical. Sultan et al. (2024) demonstrated that adapting to specific data distributions and learning the unique features of a target domain through fine-tuning improves accuracy. Lee et al. (2024) applied the SAM method to surface water extraction using Compact Advanced Satellite 500-1 (CAS500-1) Images.

This study is the first to apply SAM to forest road extraction and further analyzes fine-tuning results. Forest roads differ from general roads in several ways that challenge the original SAM model. Their boundaries are often ambiguous due to surrounding vegetation, occlusion by tree canopies, narrow and winding shapes, seasonal changes, and low contrast with the background. These conditions reduce SAM's ability to generate accurate masks because it was primarily trained on large-scale, object-centric datasets where boundaries are more precise and object shapes are less irregular. As a result, when SAM is applied directly to forest-road imagery, it can be expected to produce fragmented or over-segmented masks.

To address these limitations, fine-tuning is necessary. This study focuses on adapting SAM's mask decoder to better recognize the unique visual patterns of forest roads, which are underrepresented in the pretraining data. By fine-tuning only the decoder, we preserve the strong generalization ability of the pretrained image encoder while enabling the model to learn domain-specific cues, such as subtle linear textures, partially occluded road segments, and low-contrast edges.

Regarding applied satellite imagery, CAS500-1 is an Earth observation satellite that was launched on March 22, 2021. Its mission is to monitor land, study the environment, and observe natural disasters. CAS500-1 currently operates in a sun-synchronous orbit approximately 500 km above Earth. The satellite has a panchromatic resolution of 0.5 m, a multispectral resolution of 2 m, and a swath width of approximately 13 km. According to the 2024 Regulations concerning the Construction and Maintenance of Forest Roads (https://www.law.go.kr/LSW/admRulLsInfoP.do?admRulSeq=2100000218498), the effective width of forest roads for wildfire prevention must be at least 3 m. In areas where vehicles must pass, the width may be increased to 5 m. However, roads with curved alignments require a minimum width of 8 m. Additionally, shoulders and roadside ditches can be constructed with widths ranging from 50 cm to 1 m. Due to their narrow width, forest roads cannot be accurately detected using medium- or low-resolution images. High-resolution satellite images with a resolution better than 5 m are required.

There are many publicly available datasets for training and validating machine learning and deep learning schemes. Despite the abundance of public datasets, few are dedicated to training or validating the extraction of object features from forest roads using satellite imagery. Therefore, this study used forest roads extracted from Google Maps (GM) in the East Asian region, which has similar environmental conditions, as the training and validation data. In 2019, the Forest Big Data Exchange Platform (FBDEP) developed a digital map of national forest roads. This data was revised and updated in 2021 before being made publicly available. However, since the map was constructed in a polyline format, it does not represent the condition or width of the forest roads (https://www.bigdata-forest.kr/product/FRT002601). This study is an experimental investigation to verify whether SAM can be fine-tuned to extract forest roads using a small number of annotated images derived from GM, given the lack of publicly available training datasets for this class of object.

2. Materials and Methods

2.1. Study Area

To apply an artificial intelligence (AI)-based scheme, the ideal location for extracting forest roads should be an area with multiple forest roads, allowing for the use of multiple sources to establish ground truth data. Incorporating roads of varying widths within the area is also recommended to facilitate evaluation of the model under more challenging conditions. A complex background environment is also necessary to assess the model's performance in realistic situations.

Mount Chiak is located on the border between Wonju City and Hoengseong County in Gangwon Special Self-Governing

https://doi.org/10.7780/kjrs.2025.41.6.7

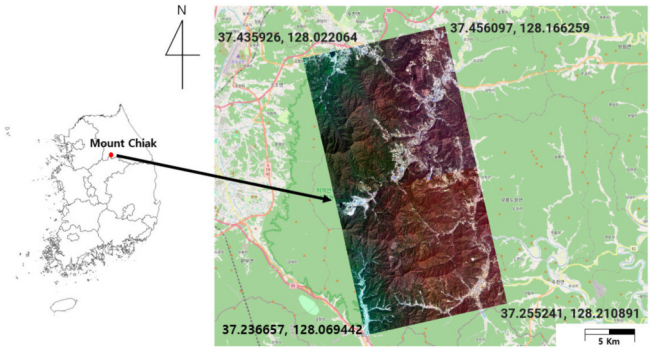

Fig. 1. Study area and applied satellite images of CAS500-1 with 2 m resolution.

Province, South Korea (Fig. 1). It is located at the southern end of the Charyeong Range, which branches southwest from the main Taebaek Mountains. The target region for forest road extraction, as shown in the satellite image, spans approximately 289.6 km².

2.2. Applied Methods

2.2.1. SAM and Workflow

SAM is a pre-trained vision model designed for general-purpose segmentation of a wide range of objects. One of its core components is the Vision Transformer (ViT). The architecture of SAM comprises three main components: an image encoder, a prompt encoder, and a mask decoder, as shown in Eq. (1). The image encoder extracts meaningful features from the input image, while the prompt encoder processes various user inputs to guide the segmentation process.

$$M = MaskDecoder(ImageEncoder(\mathrm{I}), PromptEncoder(\mathrm{P})) \quad (1)$$

where M is the predicted segmented mask, and I is the input image. P means user prompt. ImageEncoder(I) and PromptEncoder(P) are image embedding and prompt embedding, respectively. Based on these inputs, the mask decoder generates accurate masks that clearly separate the target object from the background.

Unlike traditional convolutional neural network (CNN)-based encoders, ViT divides an image into fixed-size patches and processes them as a sequence. This allows ViT to capture global contextual information across the entire image effectively. In SAM, ViT serves as the image encoder, converting high-resolution input images into rich visual feature representations. Through this process, ViT learns the relationships between image patches, enabling it to accurately recognize the boundaries and shapes of objects, even when they are spatially distant.

SAM employs different ViT backbones: ViT-B (Base), ViT-L (Large), and ViT-H (Huge). These backbones primarily differ in model size and computational capacity. Osco et al. (2023) reported that the ViT-H model achieved higher accuracy than other models for object extraction tasks, such as roads and buildings, based on SAM. Similarly, empirical studies by Baker and Handmann (2023) and Liu et al. (2024b) demonstrated that ViT-H produced superior results. Therefore, this study adopted the ViT-H backbone for the proposed approach.

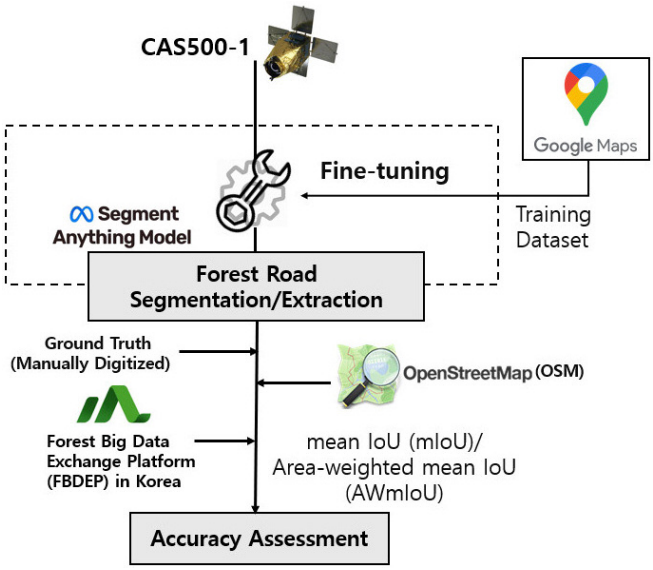

Fig. 2. Workflow and research scheme.

Fig. 2 shows the overall workflow for forest road segmentation and accuracy assessment using high-resolution satellite imagery. CAS500-1 imagery was used as the primary data source for extracting forest roads because it offers a high spatial resolution suitable for capturing narrow and occluded features. It is known that SAM's performance can be significantly improved by updating specific parts of the pre-trained model with a small amount of labeled data from the target domain. Fine-tuning is especially helpful in fields that require high precision or involve objects that differ significantly from those in the original training dataset.

2.2.2. Fine-Tuning Scheme

The architecture shown in Fig. 3 comprises three major components in SAM: image encoding, interactive segmentation, and decoder fine-tuning. In the first stage, the input remote sensing imagery is divided into a set of non-overlapping patches, which are subsequently passed through the image encoder of the SAM architecture to generate high-dimensional image features. These encoded features are stored as feature files and serve as reusable inputs for both inference and training.

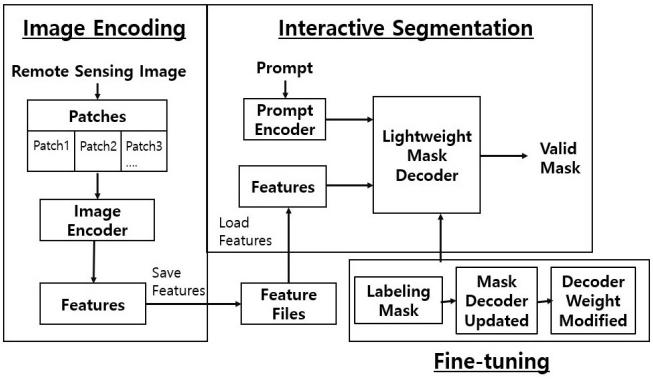

Fig. 3. SAM architecture of image encoding and interactive segmentation with fine-tuning.

In the interactive segmentation stage, user prompts, such as point- or box-based cues, are processed by the prompt encoder to produce prompt embeddings. Together with the pre-computed image features, these embeddings are forwarded to the lightweight mask decoder to generate the final segmentation mask. This modular structure allows the system to load pre-extracted features efficiently without re-encoding the image at every interaction, thereby reducing computational cost during large-scale remote sensing applications.

The fine-tuning stage focuses on adapting the mask decoder to domain-specific characteristics of forest roads. During training, the stored image features are paired with manually prepared labeling masks to update the decoder weights. Only the decoder parameters are optimized, while the image encoder and prompt encoder remain frozen. This lightweight fine-tuning strategy reduces the overall training burden and enables effective domain adaptation from heterogeneous imagery, including GM, to target sensor data of CAS500-1. The updated decoder subsequently improves segmentation accuracy by learning sensor-specific spatial patterns and road geometries present in the annotated training samples.

Table 1 lists the hyperparameters for the fine-tuning process. In this study, we employed the SAM model with a ViT-H backbone and used the pre-trained checkpoint "sam_vit_h_4b8939.pth" for fine-tuning. The model's image encoder was frozen and excluded from training, while only the prompt encoder and mask decoder were set to trainable to optimize the model for the forest road dataset.

The input data consisted of forest road images with Common Objects in Context (COCO) format annotations, with polygon information converted into binary masks for use as ground-truth labels during training. All images were resized to 1,024 × 1,024 pixels and normalized using the mean and standard deviation. The training prompts included both points and bounding boxes, and the mask decoder was configured to produce a single mask. The low-resolution masks produced by the decoder were upsampled to 1,024 × 1,024 pixels using bilinear interpolation. For optimization, the AdamW optimizer was adopted. The learning rate was adjusted using a LambdaLR scheduler, which linearly increased during the initial 250 steps (warm-up phase) and subsequently decayed by a factor of 0.1 at 66.7% and 86.7% of the total training steps.

Table 1. Hyperparameter list for the fine-tuning process

Entries	Detailed setting
Checkpoint	SAM ViT-H (sam_vit_h_4b8939.pth)
Batch size	2
Normalization	Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]
Optimizer	AdamW
Learning rate	1e-4
Weight decay	1e-4
Learning rate scheduler	LambdaLR
Loss function	20× Focal + Dice + IoU (MSE)
Frozen layers	Image encoder

The loss function was designed to improve training stability on imbalanced datasets, enhance boundary precision, and improve Intersection over Union (IoU) prediction quality. It comprised a weighted sum of the sigmoid focal loss (with a weighting factor of 20), the Dice loss, and the mean squared error (MSE) between the predicted IoU and the ground truth IoU. Model performance was evaluated using the per-mask IoU metric (micro-imagewise, threshold=0.5) provided by the segmentation_models_pytorch library. Training was conducted in FP32 precision with a batch size of 2 for 10 epochs.

2.2.3. Weighted Mean IoU

The mean IoU (mIoU) method is among the most widely used approaches for evaluating object detection accuracy. The mIoU is a metric used in object detection and segmentation to measure performance across classes. The primary function of this algorithm is to calculate the average IoU of all objects detected within the target area. A notable feature of this calculation is its equitable treatment of all objects, irrespective of their size.

Tetteh et al. (2021) introduced an overall segmentation quality metric for delineating agricultural parcels from satellite imagery, which is equivalent to the area-weighted mIoU (AWmIoU) used

https://doi.org/10.7780/kjrs.2025.41.6.7

in this study. Lee et al. (2024) applied the SAM method to extract water bodies and reported that the AWmIoU technique yields more practically meaningful results than the conventional mIoU. As for AWmIoU, applying the area of each detected object as a weight constitutes a more rational approach in Eq. (2), as it reflects the influence of larger and more significant objects on the overall average accuracy.

$$\textit{Area-weighted mean IoU} - \frac{\sum_{i=1}^{n} \omega_i \cdot IoU_i}{\sum_{i=1}^{n} \omega_i} \quad (2)$$

where N means the total number of classes, IoU_i and ω_i are IoU for class i and weight for class i, typically the area of overlap between the ground truth and the detected region for class i, respectively.

2.3. Data Sources

2.3.1. Applied Data

We applied the fine-tuned SAM to the CAS500-1 imagery to automatically segment and extract forest roads. The extracted results were evaluated against multiple reference datasets to assess accuracy. Several reference datasets were used to evaluate the accuracy of the results, including OpenStreetMap (OSM) and FBDEP. This integrated approach validates segmentation results against expert annotations and publicly available geospatial data. This framework presents a scalable, adaptable methodology for applying vision foundation models to remote sensing tasks, particularly for infrastructure mapping in forested areas. The ground truth was manually digitized by an expert interpreter based on the CAS500-1 image.

The CAS500-1 image used in this study is a level 2G (geometrically corrected) product. It was captured on April 20, 2022, with a ground sampling distance (GSD) of 2 m. The OSM data used for validation is a polyline shapefile obtained with the QuickOSM plugin (https://docs.3liz.org/QuickOSM/) using the "track" tag, dated August 23, 2023. The FBDEP data is a shapefile dated November 17, 2021. These references provide forest road data in polyline format, while SAM outputs data in polygon format. All validation and extracted data were converted into shapefiles and processed using QGIS.

2.3.2. Fine-Tuning Data

To fine-tune the training process, satellite images were automatically collected using the Google Static Maps application programming interface (API) in conjunction with the "requests" library. Latitude and longitude coordinates were passed to the API, and images were saved based on the specified center points. To avoid overfitting, the training data were extracted from areas outside the study region. Since the ViT models used in SAM are designed to accept inputs at a default resolution of 1,024 × 1,024, all fine-tuning data was resized to this resolution. The collected images were loaded in RGB format using OpenCV and the Python Imaging Library (PIL), then manually labeled. Based on these labels, a forest road dataset was created in the COCO format using JavaScript Object Notation (JSON), a widely used data format in computer vision applications.

The GM training samples for forest roads were collected from selected regions in South Korea, central and northern Japan, and the northeastern and northern areas of China. These regions were chosen because they share environmental characteristics similar to those of the study area, including temperate mixed forests of broadleaf and conifer species, mountainous and hilly terrain, comparable slope conditions, and potentially similar road-network patterns. The spatial sampling strategy first identified candidate areas centered on forested zones. To avoid geographic bias toward specific locations, a regular-interval random-stratified sampling approach was used. This was followed by a visual inspection process to determine the presence of roads, and all road masks were manually annotated to ensure label accuracy. Because forest roads are often connected to general roads and their boundaries are not clearly distinguishable, creating high-quality training data was not straightforward. As a result, the number of available samples was limited.

Osco et al. (2023) found that a fine-tuned SAM model for remote sensing imagery can improve performance, even with a small dataset of a few hundred annotated masks. However, there are no existing training datasets explicitly designed for forest roads. Xie et al. (2024) reported experimental results in the medical imaging domain demonstrating that fine-tuning with fewer than one hundred mask annotations can still yield significant improvements in segmentation accuracy for extracting specific anatomical structures. Instead, we manually generated masks from GM. We created a total of 604 masks and split them into training, validation, and test sets at 8:1:1 for the experiments. Of the 604 masks, 404 were simple and 200 were complex. Simple objects refer to masks with straightforward shapes, while complex objects refer to masks with irregular boundaries or fragmented segments.

Fig. 4 shows subsets of forest roads extracted from GM imagery along with the corresponding binary masks used to train and validate the model. The criteria for classifying the dataset as simple

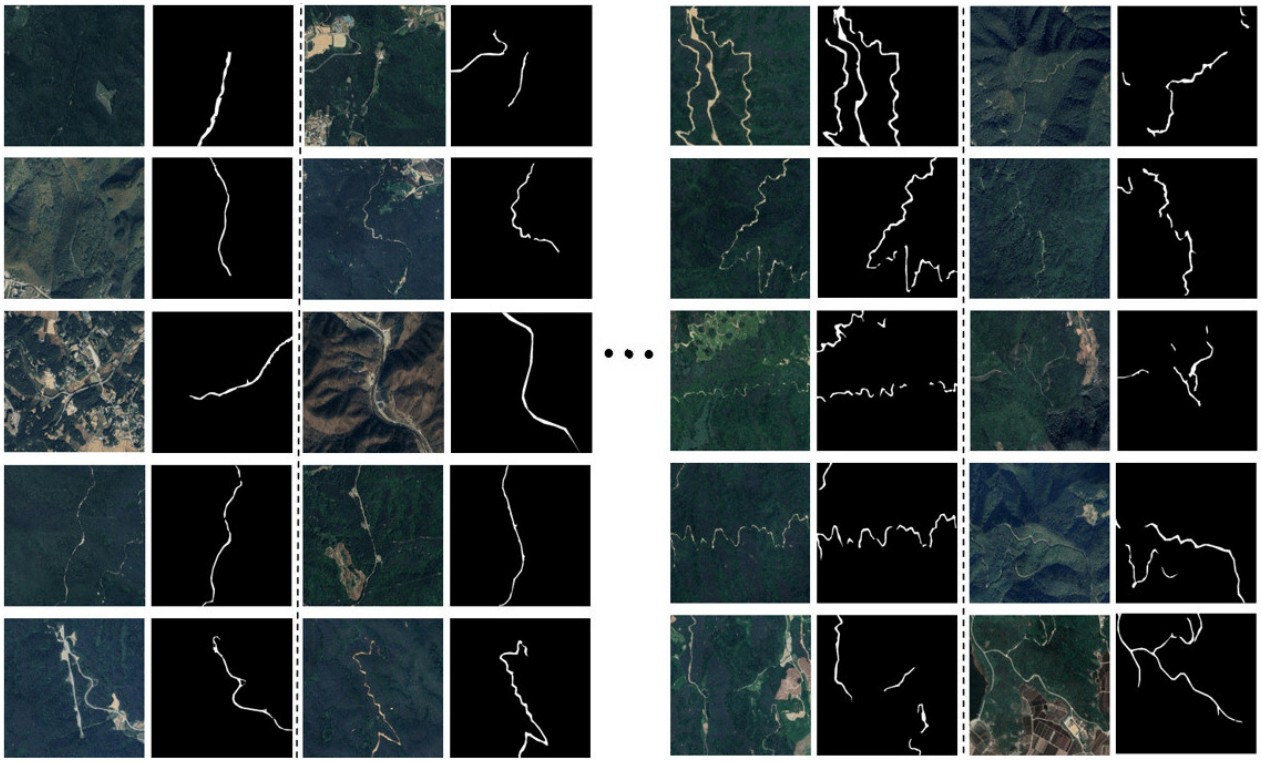

Fig. 4. Examples of forest roads for fine-tuning: visible forest road segments captured in a high-resolution satellite image, along with corresponding manually annotated binary masks.

or complex types were established based on visual interpretation. First, data exhibiting a continuous shape with minimal fragmentation were categorized as the simple type, while data containing several fragmented elements within the feature were classified as the complicated type. Second, training samples consisting of straight forms or objects with low curvature were classified as the simple type, while those exhibiting complex geometries with varying curvatures were classified as the complicated type. Although the training data were categorized as simple or complicated, all data were used together during fine-tuning in this study. Therefore, the research results reflect the combined contribution of both types of data.

Fig. 5 presents the training dynamics of the proposed model, showing loss and accuracy over 10 epochs. Fig. 5(a) shows the training and validation loss curves. The training loss initially declines precipitously and then gradually converges, indicating rapid parameter optimization. It is noteworthy that the validation loss closely follows the training loss throughout the learning process, with minimal divergence. This suggests that the model generalizes well to unseen data and does not overfit. Fig. 5(b) shows the training and validation accuracy curves. Both metrics demonstrate a consistent, gradual increase over time, reaching a

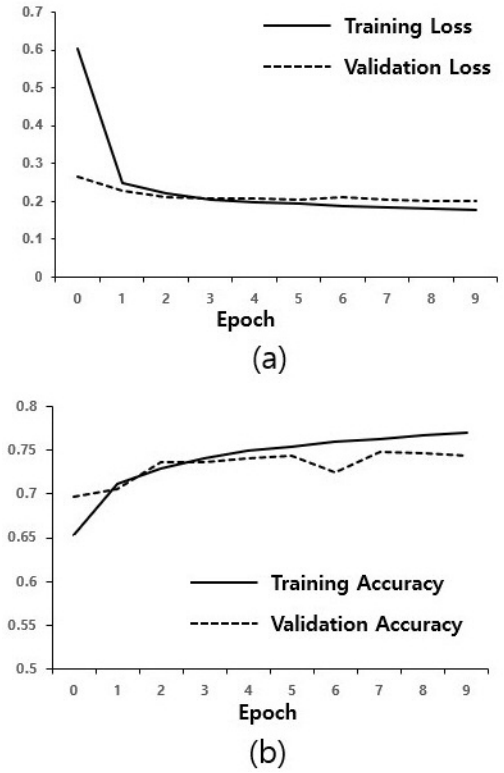

Fig. 5. Fine-tuning processes: (a) training and validation loss and (b) training and validation accuracy.

https://doi.org/10.7780/kjrs.2025.41.6.7

plateau after approximately 5 epochs. The accuracy values for the training and validation sets remain closely aligned, providing further evidence of the model's stability and capacity for generalization.

Collectively, the findings demonstrate that the model converges stably with a relatively small number of training epochs. The similarity between the model's performance during training and validation suggests that the architecture and hyperparameters are appropriate. These results suggest that the model has the potential to be applied to real-world tasks, such as satellite image segmentation, where accuracy and computational efficiency are critical.

3. Results

Fig. 6 presents a visual comparison of forest road vector layers from various sources and extraction methods within the same geographic region. The boxed area in Fig. 6 corresponds to the study region shown in the satellite image in Fig. 1. Figs. 6(a–c) show the overlay of forest road data from (a) manually digitized ground truth, (b) FBDEP, and (c) OSM, respectively. This demonstrates the integration of diverse data sources. These reference layers serve as baseline vectors for validating object extraction performance.

Figs. 6(d) and (e) show the segmentation results obtained by applying the SAM without (d) and with (e) fine-tuning. The extracted polygons were overlaid on the same base map to facilitate direct visual comparison. As previously mentioned, the fine-tuned SAM delineates forest road features more completely and continuously than the base SAM, particularly in the central and southern regions of the study area. This suggests that refining the SAM with domain-specific training data substantially enhances its ability to identify narrow and occluded road segments in complex forest environments.

Furthermore, the enhanced geometric alignment and reduced fragmentation of road features in the refined output substantiate

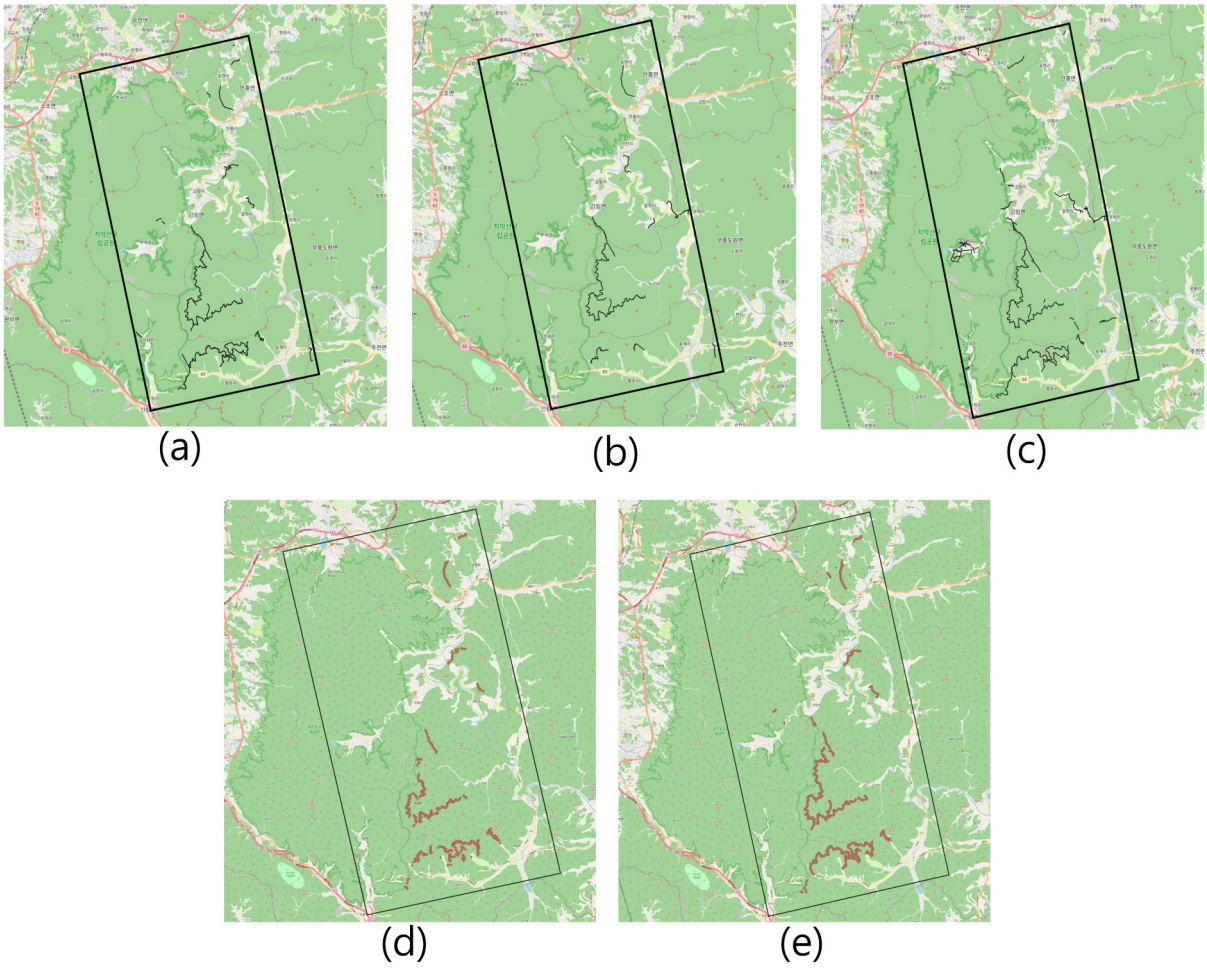

Fig. 6. Forest roads for validation and extracted results in the study area on OSM: (a) forest roads as ground truth, (b) forest roads of FBDEP, (c) forest roads extracted from OSM, (d) results using SAM, and (e) results using fine-tuned SAM.

Table 2. Confusion matrix for result validation

mIoU \ AWmIoU	SAM	Fine-tuned	Ground truth	FBDEP	OSM
SAM	1.0000	0.7844	0.6970	0.6563	0.6785
Fine-tuned	0.7342	1.0000	0.8369	0.7422	0.8201
Ground truth	0.6190	0.7786	1.0000	0.8384	0.9231
FBDEP	0.5707	0.6596	0.7214	1.0000	0.9005
OSM	0.5658	0.7177	0.8545	0.8419	1.0000

the model's adaptability to high-resolution satellite imagery when customized data are incorporated during training. The results of this study indicate that the road density in the evaluation area is relatively low. One objective of this research is to evaluate the detectability of sparsely distributed road objects. Low-density environments are an important testing ground for evaluating a model's ability to generalize to small, sparse objects. Additionally, the fine-tuned SAM model developed in this study primarily learns geometric and structural features, enabling it to perform well across road densities. Results from an evaluation area with sparsely distributed roads provide strong evidence of the model's applicability and generalizability.

Table 2 shows the mIoU and AWmIoU values for each pair of the five segmentation sources: the original SAM, the fine-tuned SAM, the manually labeled ground truth, the FBDEP, and the OSM. This matrix is used to quantitatively validate the spatial agreement between the different segmentation outputs. The results demonstrate that fine-tuning significantly enhances segmentation performance. Compared to the original SAM, the fine-tuned model achieved AWmIoU and mIoU values of 0.8369 and 0.7786, respectively, when validated against ground truth data, representing improvements of 25.78% and 20.06%. Similar improvements were observed when compared with OSM (AWmIoU: 0.8201) and FBDEP (AWmIoU: 0.7422). These results suggest the potential utility of the fine-tuned model in practical mapping and infrastructure monitoring applications.

Among the external vector datasets, OSM displays the highest consistency with the ground truth (AWmIoU: 0.9231), followed by FBDEP (AWmIoU: 0.8384). This suggests that both datasets can serve as reliable references for delineating forest roads. OSM and FBDEP also exhibit high mutual similarity (AWmIoU: 0.9005), highlighting the consistency of publicly available geographic datasets in forested environments. Furthermore, the AWmIoU agreement patterns are reflected in the mIoU matrix, confirming the robustness of the evaluation. Together, these findings support the fine-tuned SAM's ability to generalize to unseen data and approximate expert-labeled references. This reinforces its potential for semi-automated road extraction and validation using open-source geospatial data.

4. Discussion

The study area is outlined on a regional topographic base map, and a specific region of interest is selected for detailed evaluation, marked by a blue box (Fig. 7). Fig. 7(a) shows a digital topographic map with elevation contours, including existing road vectors in red. Fig. 7(b) shows a high-resolution optical satellite image with the exact road alignment superimposed, which enables visual verification of topographic alignment. Figs. 7(c–f) present zoomed-in views of key segments, comparing road extraction results from different segmentation methods or configurations. Blue and magenta lines indicate segmented road features from two distinct approaches, which are likely baseline and fine-tuned SAM models or alternative algorithms. Yellow bounding boxes highlight regions for qualitative comparison, emphasizing spatial consistency and boundary precision between the extracted road segments and the visible ground truth features in the imagery.

Although the fine-tuned SAM showed a clear performance gain on CAS500-1, the training–testing configuration inherently suffers from a domain shift, because GM is a composite product derived from multiple optical sensors, whereas CAS500-1 is a 2 m multispectral sensor. In this study, GM tiles were adopted as the primary training source to leverage readily available, visually consistent road patterns over large areas, thereby avoiding the considerable cost of manually annotating a large CAS500-1 training set. To reduce the domain gap, we restricted GM sampling to East Asian temperate forest regions whose land-cover and topographic characteristics are similar to those of the Korean CAS500-1 scenes. Nevertheless, differences in radiometric properties, contrast, and shadow patterns between GM and CAS500-1 may still lead to local segmentation errors, particularly for narrow or heavily occluded forest roads.

https://doi.org/10.7780/kjrs.2025.41.6.7

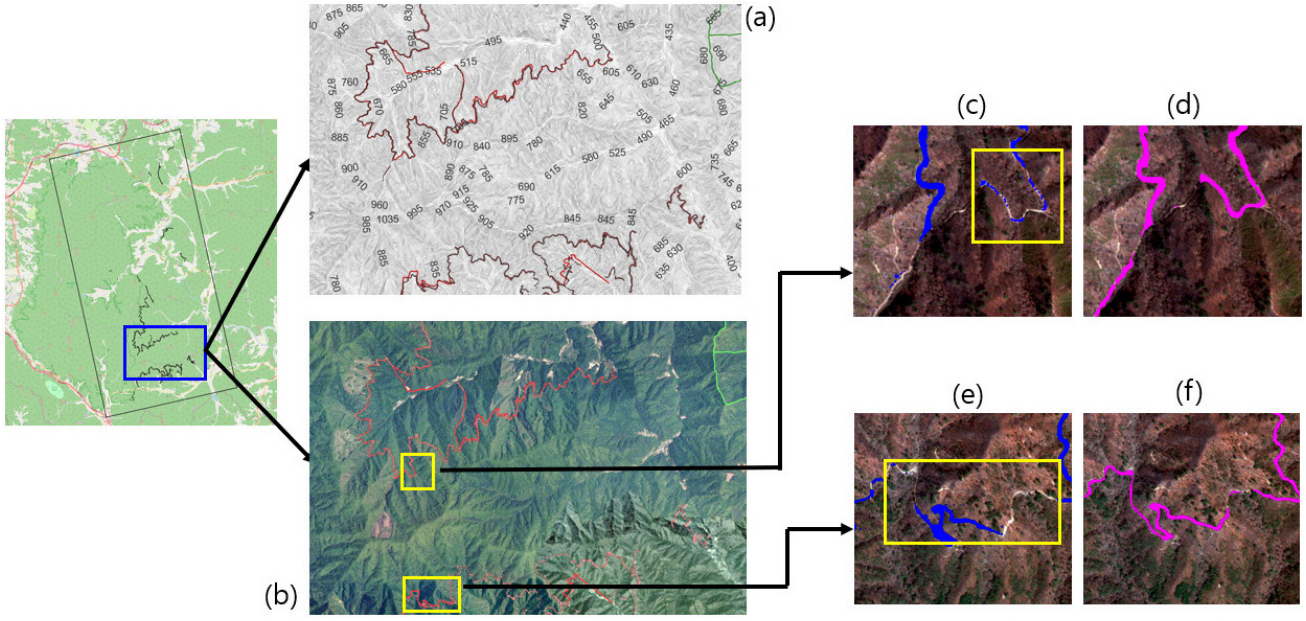

Fig. 7. Extracted results in a part of study area: (a) forest roads on contour line, (b) forest roads (red line) and hiking road (green line) on Landsat image, (c) exaggerated view of forest roads using SAM in a specific area, (d) forest roads using fine-tuned SAM, (e) exaggerated view of forest roads using SAM, and (f) forest roads using fine-tuned SAM.

In comparison with previous studies, this study confirmed that SAM performs well even in specialized domains, such as forest roads, which were not addressed in Osco et al. (2023). It also showed that high accuracy can be achieved without using large-scale training datasets, as suggested by Sultan et al. (2024). The substantial improvement in AWmIoU after decoder fine-tuning suggests that the proposed approach can partially transfer road extraction capability across domains. However, the current framework should be regarded as a first step towards cross-domain forest road mapping, and further work is needed to construct a larger, in-domain CAS500-1 training set and to integrate more robust domain adaptation strategies to better address the GM–CAS500-1 sensor mismatch.

Regarding the scalability of GM-based data collection, the current process relies on manual and semi-automated methods. To enable large-scale expansion, a pipeline for automated mask extraction and quality verification would be necessary. This approach emphasizes the importance of model adaptation and domain-specific data preparation in optimizing performance for complex geospatial tasks.

5. Conclusions

This case study demonstrated the feasibility of applying a fine-tuned SAM to extract forest roads from high-resolution satellite imagery. The main finding is that fine-tuning with a few hundred domain-specific samples substantially improves accuracy. The experimental results demonstrate that fine-tuning the SAM model yields substantial performance gains, with improvements of 25.78% in mIoU and 20.06% in AWmIoU. Furthermore, the fine-tuned model enhances the continuity of extracted objects, thereby providing tangible advantages that validate the effectiveness of domain-specific optimization. Although this study is a case study conducted using a specific region and a particular satellite image, the target area represents a typical forest environment. Therefore, the results can be considered generalizable and applicable to other high-resolution optical satellite images with similar spatial resolution. Overall, it is concluded that incorporating foundation vision models into remote sensing workflows is a viable approach, highlighting the importance of model adaptation and domain-specific data preparation to optimize performance in complex geospatial tasks.

Author Contributions

Conceptualization: Lee K, Kim K; Data curation, Formal analysis: Lee K, Lee H; Funding acquisition: Lee K, Investigation, Methodology: Lee K, Kim K, Lee H; Project administration, Supervision: Lee K; Validation: All authors; Writing–original draft: Lee K, Lee H; Writing–review & editing: All authors.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Funding

This research was financially supported by Hansung University.

Data Availability Statement

The Google Maps satellite images used in this study were obtained through the Google Maps Static API for research and non-commercial academic purposes only. All Google Maps images used for training mask generation were processed in compliance with Google's copyright policy, and no copyrighted visual data were redistributed. This study used OpenStreetMap data (© OpenStreetMap contributors, ODbL 1.0) for reference and validation purposes. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

None.

Supplementary Materials

None.

References

Baker, N. A., and Handmann, U., 2023. Don't waste SAM. In *Proceedings of the 2023 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, Oct. 4–6, pp. 429–434. https://www.esann.org/sites/default/files/proceedings/2023/ES2023-116.pdf

Çalışkan, E., and Sevim, Y., 2022. Forest road extraction from orthophoto images by convolutional neural networks. *Geocarto International, 37*(26), 11671–11685. https://doi.org/10.1080/10106049.2022.2060319

Kelesakis, D., Marthoglou, K., Tokmaktsi, E., Tsiros, E., Karteris, A., Stergiadou, A., et al., 2024. Forest/rural road network detection and condition monitoring based on satellite imagery and deep semantic segmentation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10*, 81–88. https://doi.org/10.5194/isprs-annals-X-4-W4-2024-81-2024

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al., 2023. Segment anything. *arXiv preprint arXiv: 2304.02643*. https://doi.org/10.48550/arXiv.2304.02643

Lee, H., Kim, K., and Lee, K., 2024. Application of Geo-Segment Anything Model (SAM) scheme to water body segmentation: An experiment study using CAS500-1 images. *Korean Journal of Remote Sensing, 40*(4), 343–350. https://doi.org/10.7780/kjrs.2024.40.4.2

Liu, R., Wu, J., Lu, W., Miao, Q., Zhang, H., Liu, X., et al., 2024a. A review of deep learning-based methods for road extraction from high-resolution remote sensing images. *Remote Sensing, 16*(12), 2056. https://doi.org/10.3390/rs16122056

Liu, S., Wang, F., You, H., Jiao, N., Zhou, G., and Zhang, T., 2024b. Context-aggregated and SAM-guided network for ViT-based instance segmentation in remote sensing images. *Remote Sensing, 16*(13), 2472. https://doi.org/10.3390/rs16132472

Osco, L. P., Wu, Q., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., et al., 2023. The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation, 124*, 103540. https://doi.org/10.1016/j.jag.2023.103540

Sultan, R. I., Li, C., Zhu, H., Khanduri, P., Brocanelli, M., and Zhu, D., 2024. GeoSAM: Fine-tuning SAM with multi-modal prompts for mobility infrastructure segmentation. *arXiv preprint arXiv: 2311.11319*. https://doi.org/10.48550/arXiv.2311.11319

Tetteh, G. O., Gocht, A., Erasmi, S., Schwieder, M., and Conrad, C., 2021. Evaluation of Sentinel-1 and Sentinel-2 feature sets for delineating agricultural fields in heterogeneous landscapes. *IEEE Access, 9*, 116702–116719. https://doi.org/10.1109/ACCESS.2021.3105903

Xie, W., Willems, N., Patil, S., Li, Y., and Kumar, M., 2024. SAM fewshot finetuning for anatomical segmentation in medical images. In *Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, Jan. 3–8, pp. 3241–3249. https://doi.org/10.1109/WACV57701.2024.00322